

**ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES**

# **THESE**

pour l'obtention du grade de DOCTEUR DE L'EHESS

Discipline SCIENCES COGNITIVES  
mention Psychologie Cognitive

présentée et soutenue publiquement par

**Anne Faure**

le 13 décembre 2000

TITRE :

**Des sons aux mots,  
comment parle-t-on du timbre musical ?**

---

Directeur de thèse : Stephen McAdams

---

**JURY**

Jean Petitot (président)

Juan Segui (rapporteur)

Charles Tijus (rapporteur)

Danièle Dubois (examineur)

Stephen McAdams (directeur)

**ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES**

# **THESE**

pour l'obtention du grade de DOCTEUR DE L'EHESS

Discipline SCIENCES COGNITIVES  
mention Psychologie Cognitive

présentée et soutenue publiquement par

**Anne Faure**

le 13 décembre 2000

TITRE :

**Des sons aux mots,  
comment parle-t-on du timbre musical ?**

---

Directeur de thèse : Stephen McAdams

---

**JURY**

Jean Petitot (président)

Juan Segui (rapporteur)

Charles Tijus (rapporteur)

Danièle Dubois (examineur)

Stephen McAdams (directeur)



Ceci n'est pas un timbre d'instruments de musique

## Remerciements

Cette thèse est le fruit de nombreuses rencontres qui m'ont chacune enrichies : MERCI !

Je remercie tout d'abord Stephen McAdams pour avoir su garder confiance en cette thèse durant toutes ces années... et pour avoir su à la fois me laisser la liberté de fouiner dans des directions diverses et me redonner confiance en des moments clés. Sa rigueur d'expérimentaliste, en m'obligeant à argumenter, m'a souvent permis d'approfondir ce qui aurait pu rester des "divagations"...

Je voudrais remercier Jean Petitot d'avoir accepté de présider le jury de ma thèse. J'ai apprécié son approche de la catégorisation ainsi que les séminaires de phénoménologie organisés à l'EHESS.

Je remercie Juan Segui d'avoir accepté d'être rapporteur de ce travail. Je le considère comme à l'origine de ce projet puisqu'il est le premier psychologue que j'ai rencontré en sortant de l'ESPCI. Je le remercie de m'avoir présenté le travail de Stephen McAdams, puis accueillie dans son laboratoire. Déjà rapporteur de mon mémoire de DEA, j'ai apprécié le regard bienveillant et dynamisant qu'il portait sur le groupe des thésards du LPE à Paris V.

Je tiens à remercier Charles Tijus d'avoir accepté d'être rapporteur de ce travail, mais aussi de m'avoir invitée à venir utiliser les outils créés dans son laboratoire, et surtout de m'avoir encouragée et aidée à imaginer une arborescence de mes verbalisations. La vraie richesse des descriptions que j'avais récoltées a pu apparaître et être utilisable grâce à cette rencontre. Merci aussi de m'avoir proposé une vraie première expérience d'enseignement grâce à un poste d'ATER en psychologie.

Je voudrais remercier Danièle Dubois qui m'a ouvert sa porte à un moment de doutes qui fut un tournant de ma thèse. Son regard m'a permis de faire un bilan et d'envisager de poursuivre, avec une autre façon de voir les données verbales, notamment à la suite de l'école de sémantique des langues d'Asnelle. J'ai aussi ressenti son encouragement pour terminer ce travail. Merci d'avoir accepté de le commenter et le questionner en participant à ce jury.

J'aimerais remercier aussi tous les étudiants (et aussi les statutaires, ou ceux qui le sont devenus) que j'ai rencontré et avec qui les séances de bavardages ont été très productives. En particulier, Sophie Donnadieu auprès de qui j'ai appris l'essentiel de la psycho que je connais ! Et puis à l'Ircam : Jochen, Gérard, Patrick, Daniel, Olivier, Vincent et Caroline dont j'ai apprécié les encouragements et relectures de Montréal. Et puis aussi Cécile et Bennett (pour refaire le monde, chacun à leur façon !). A Paris 8, je tiens à remercier : Béa et ses anges, Denis et la réalité de la réalité, Emmanuel et ses bosses de chameau, Isabel et ses baby-sitters (!). Je remercie aussi particulièrement Sébastien Poitrenaud qui m'a initiée à l'utilisation de son bébé de pierre (STONE). A Cognivence je voudrais

saluer l'enthousiasme d'un certain "utopisme" actif : Amal, Frédérique, Seb... Je veux aussi remercier nos amis, proches, voisins qui ont donné un coup de pouce à ce travail, en particulier Bruno et Marie (merci pour le CD et sa pochette), Micheline et Bernard (merci pour votre relecture d'éditeur), Sabine, Françoise et Renaud (merci pour votre temps passé aux dernières corrections), et tous ceux qui ont invité nos enfants au moment du "rush final"... en particulier Domi.

Bien sûr, ce travail n'aurait pas été possible sans la contribution de nombreux "sujets" parmi lesquels j'ai compté pas mal de mes amis... mais aussi des inconnus venus visiter l'Ircam et toujours intéressés par les sons... Leur intérêt, la bienveillance et le sérieux avec lesquels ils se sont prêtés au jeu ont été pour moi un aiguillon pour essayer de rendre compte de la richesse de leur travail, notamment en essayant de s'accrocher à ces verbalisations libres...

Et puis bien sûr, incontournable :

**Philippe**

Belle équipe pour mener à bien 2 thèses et 3 enfants en 6 ans ! Merci de ton soutien multiforme (nombreuses vacances de papa en solo ET super technicien de maintenance informatique !) parfois musclé ! (surtout sur les "s" et les A42.II.35, touché-coulé !)

Et puis aussi Chloé (1994), Ariane (1996) et Rémi (2000)... Grâce à vous, j'ai pu notamment assister plusieurs fois au miracle de la naissance (l'invention ?) du langage... Les TPs de sciences cognitives en parallèle avec la théorie ! Les enfants transforment encore plus profondément qu'une thèse. Merci d'avoir prêté un peu votre maman à ce travail sans fin...

## SOMMAIRE

<b>Remerciements</b>	<b>4</b>
<b>Avant-propos</b>	<b>6</b>
<b>Introduction</b>	<b>9</b>
<b>1- Parler des sons : un programme controversé</b>	<b>9</b>
<b>2- Définitions du timbre : un résumé de la problématique</b>	<b>11</b>
<b>3- Les moyens : expérimentation et portée des conclusions</b>	<b>13</b>
<b>Partie I : Questions et inscription théorique</b>	<b>19</b>
<b>Chapitre 1. Être au monde</b>	<b>22</b>
<b>Chapitre 2. Parler</b>	<b>40</b>
<b>Chapitre 3. Des sons</b>	<b>72</b>
<b>Hypothèses</b>	<b>101</b>
<b>Partie II : Expériences effectuées</b>	<b>102</b>
<b>Chapitre 4. Présentation et Méthodologie des trois expériences réalisées</b>	<b>104</b>
<b>Chapitre 5. Corrélats verbaux des dimensions perceptives</b>	<b>141</b>
<b>Chapitre 6. Description des sons</b>	<b>188</b>
<b>Chapitre 7. Pertinence des échelles et sens des portraits</b>	<b>263</b>
<b>Chapitre 8. Association des sons aux portraits</b>	<b>302</b>
<b>Partie III : Discussion</b>	<b>357</b>
<b>I. Les sons pour nous</b>	<b>357</b>
<b>II. Le langage pour décrire nos perceptions</b>	<b>367</b>
<b>III. Notre perception du monde</b>	<b>371</b>
<b>IV. Conclusion</b>	<b>383</b>
<b>V. Perspectives</b>	<b>384</b>
<b>Postface</b>	<b>385</b>
<b>Bibliographie</b>	<b>386</b>
<b>Table des matières</b>	<b>400</b>
<b>Annexes et CD</b>	<b>410</b>

## Avant-Propos

Cette thèse a été pour moi l'occasion d'un voyage passionnant à travers plusieurs disciplines mené par une même question : comment parle-t-on d'une de nos perceptions : le timbre d'instruments de musique, ou plutôt le timbre de sons imitant des sons d'instruments de musique et des sons inventés. Voyage à travers plusieurs disciplines parce que j'avais une formation en physique et chimie quand j'ai commencé à m'intéresser à nos perceptions auditives au sein de l'équipe Perception et Cognition Musicales de l'Ircam, qui est restée une sorte de "pied à terre" au cours du voyage. J'y ai découvert une façon de connaître et d'appréhender les sons, un lien et un aller-retour constant entre des données de psychologie et d'acoustique.

Et puis, parce qu'on s'intéressait à la manière dont on parlait de certains sons, j'ai été confrontée au langage, puis plongée dedans : en particulier le langage oral libre produit spontanément par des gens à qui on fait écouter des sons et à qui on demande de les décrire. L'étude de telles productions relève donc en général de la linguistique. Mais surtout, nous voilà au coeur de la question du rapport entre les mots et les choses, donc du problème de la référence, et c'est un problème qui est étudié à la fois en sémantique et en philosophie depuis fort longtemps. De plus, on ne peut étudier des phrases et des mots sans se poser la question de la polysémie et de la synonymie, donc des liens entre les mots et leur sens. Et avec la question : "peut-on reconnaître un son à partir de sa description ?" on est confronté à la dimension sociale du langage : est-ce un code, une mémoire collective ? Quels sont les rapports entre le langage et la façon dont on catégorise le monde ? Entre le langage et la pensée ?

Comprendre comment on parlait des sons m'a donc demandé de considérer certains problèmes discutés en linguistique et en philosophie. Je ne pourrai pas les éluder. Je dois dire aussi que bien souvent ces éclairages, nouveaux pour moi, m'ont passionnés. Pourtant, et bien que cette thèse ait duré presque 6 ans, je ne suis devenue ni linguiste, ni philosophe et je ne l'étais pas par formation. Ce que je pourrai en dire sera donc sans doute d'une immense naïveté d'un point de vue linguistique et philosophique. Mon regard est en effet encore très parcellaire et sans une perspective qui permettrait de replacer chaque pensée dans un contexte plus général.

A défaut d'approfondir chacun des domaines abordés, cette thèse se propose d'établir un lien entre ces domaines et les questions posées dans chacune de ces disciplines et ainsi montrer qu'elles se ramènent à des questions communes. Ce point de vue global et pluridisciplinaire (plutôt que pointu) fait partie du risque des sciences cognitives, mais c'est aussi ce qui les rend passionnantes.

Moi je joue du piano  
disait l'un  
moi je joue du violon  
disait l'autre  
moi de la harpe moi du banjo  
moi du violoncelle  
moi du biniou... moi de la flûte  
et moi de la crécelle.  
Et les uns et les autres parlaient parlaient  
parlaient de ce qu'ils jouaient.  
On n'entendait pas la musique  
tout le monde parlait  
parlait parlait  
personne ne jouait  
mais dans un coin un homme se taisait :  
"Et de quel instrument jouez-vous Monsieur  
qui vous taisez et qui ne dites rien ?"  
lui demandèrent les musiciens.  
"Moi je joue de l'orgue de Barbarie  
et je joue du couteau aussi"  
dit l'homme qui jusqu'ici  
n'avait absolument rien dit  
et puis il s'avança le couteau à la main  
et il tua tous les musiciens  
et il joua de l'orgue de Barbarie  
et sa musique était si vraie  
et si vivante et si jolie  
que la petite fille du maître de la maison  
sortit de dessous le piano  
ou elle était couchée endormie par ennui  
et elle dit :

"Moi je jouais au cerceau, à la balle au chasseur, je jouais à la marelle, je jouais avec un seau, je jouais avec une pelle, je jouais au papa et à la maman, je jouais à chat perché, je jouais avec mes poupées, je jouais avec une ombrelle, je jouais avec mon petit frère, avec ma petite soeur, je jouais au gendarme, et au voleur

Mais c'est fini fini fini, je veux jouer à l'assassin, je veux jouer de l'orgue de Barbarie."

Et l'homme pris la petite fille par la main  
et ils s'en allèrent dans les villes  
dans les maisons dans les jardins  
et puis ils tuèrent le plus de monde possible  
après quoi ils se marièrent  
et ils eurent beaucoup d'enfants.

Mais

l'aîné apprit le piano  
le second le violon  
le troisième la harpe  
le quatrième la crécelle  
le cinquième le violoncelle  
et puis ils se mirent à parler parler  
parler parler parler  
on n'entendit plus la musique  
et tout fut à recommencer !

Jacques Prévert  
L'orgue de Barbarie (1949)



## **Introduction**

L'objectif de cette thèse est de rendre compte de la manière dont on parle des sons, en particulier du timbre de certains sons, imitant des sons d'instruments de musique réels ou imaginaires. Il s'agit donc, dans un premier temps, de justifier cette démarche, de définir ce qu'est le timbre d'un son et de présenter les questions que nous allons poser et l'ensemble des moyens que nous avons mis en oeuvre pour y répondre.

### **1- Parler des sons : un programme controversé**

L'idée de parler de la musique est très controversée et en général particulièrement épidermique chez les musiciens. Pourtant, depuis que le son est enregistré et que de nouveaux métiers du son et de nouvelles façons d'utiliser le son sont apparus, la nécessité de pouvoir utiliser un vocabulaire *ad hoc* s'est faite sentir. De plus on peut voir dans ce programme un intérêt théorique qui s'insère dans des recherches actuelles en sciences cognitives sur les liens entre nos perceptions et nos concepts à travers le langage.

#### **La musique est de l'ordre de l'ineffable**

Comme le souligne le poème de Prévert (1949) mis en exergue, les musiciens, compositeurs et interprètes, nourrissent toujours une suspicion vis-à-vis de ceux qui parlent de la musique, notamment les musicologues, et les deux milieux sont en général très hermétiquement distincts. Berlioz, qui écrivait dans la Revue et Gazette musicale de Paris (1994 /1841), en est un des rares et brillants contre-exemples. Les musiciens étant souvent très critiques ou du moins méfiants vis à vis des musicologues : "les propos tenus sur la musique, tour à tour vaporeux et prosaïques, passent sans transition des méditations sublimes des inspirés aux laborieuses vaticinations des inspireurs, (et) ne donnent guère confiance dans une approche musicale réelle" (Schaeffer, 1966).

Il s'agit en effet souvent pour les musiciens d'affirmer le caractère ineffable de la musique. On peut, par exemple citer dans cette veine le compositeur Tchaïkovski (1878) : "la symphonie ne doit-elle pas exprimer tout ce pour quoi les mots manquent, mais qui cherche à sortir de l'âme pour être dit ?", ou le chef d'orchestre Scherchen (1953) "une bonne musique et de bons musiciens se comprennent sans intermédiaires (...) parmi les trois moyens d'expression du chef : description par le geste, expression par la mimique et explication par la parole, les deux dernières sont des aides discutables" (p 18 et 32). Mais ce point de vue est souvent repris par les psychologues de la musique : "la musique ne peut être entièrement exprimée par des mots" ou "deviner le programme est une entreprise oiseuse et au fond anti-musicale, puisqu'elle implique qu'on tente de traiter la musique, non comme un art expressif, mais comme un art représentatif ou documentaire" (Teplov, 1966) p 8 et 20)

Il n'est donc pas étonnant que, pendant très longtemps, les musiciens se soient passés de construire un vocabulaire pour parler des sons, décrétant cette entreprise comme inutile voire non souhaitable.

## Un nouveau travail sur le son nécessite de nouveaux outils

Pourtant, à partir du moment où on a commencé à décomposer les sons et qu'on a pu les enregistrer, on a dû commencer à se préoccuper de la qualité sonore des enregistrements. Puis de nouveaux outils de travail du son lui-même sont apparus en même temps que l'idée de matière sonore et ces outils ont été utilisés par les chercheurs, puis par les ingénieurs du son, puis par les compositeurs de toutes les musiques contemporaines, savantes ou populaires... C'est ainsi que les expériences de Hermann von Helmholtz (1877) utilisant des résonateurs, des calculs de sons différentiels et de sons de combinaison, ont donné naissance à d'importantes impulsions touchant l'harmonie. Helga de la Motte-Haber (1994, p 34) explique par exemple que "Edgard Varèse était fasciné par l'idée de la synthèse expérimentale des sons (à l'aide de la sirène), les recherches expérimentales du XIXe siècle ayant eu involontairement des conséquences sur la naissance de la musique électronique. (...) Schönberg aussi fonda le droit à la dissonance grâce à des harmoniques supérieurs éloignés, dont la prétendue perception permit également à Messiaen de construire des accords. Hindemith aussi a fait appel à la physique pour fonder les règles de l'harmonie."

On retrouve cette même importance donnée à la qualité du son en lui-même dans les musiques rock et pop, qui n'auraient pu se produire sans l'amplification. "Dans les musiques pop, chaque groupe cherche à affirmer son identité, à marquer son territoire en trouvant son propre 'son' (...) qui est soigneusement composé, souvent avec invention, comme chez les Pink Floyd - les enregistrements pop se façonnent en studio, et l'ingénieur du son est un membre important du groupe" (Risset, 1994, p 93).

Il n'est donc pas étonnant que Pierre Schaeffer, à la fois scientifique, homme de radio, poète et compositeur, se fasse dès 1966 le défenseur de la nécessité de travailler à l'élaboration d'un vocabulaire du sonore. "*Parler des sons* s'est d'abord présenté comme un événement accessoire. L'embarras du musicien devant un son inusité, dont il ne sait rien dire du tout, n'a d'équivalent que sa terminologie embarrassée pour décrire un son usuel, dont il a trop à dire. Nous avons observé que ce mutisme pouvait être lié à une autre infirmité musicale : la surdité. La découverte, dans tout phénomène musical, de l'intention d'entendre, nous porte à croire que parler des sons n'est pas si accessoire. Il le faut d'abord pour révéler ce qui allait de soi : qu'on entendait les sons traditionnels de telle ou telle façon, en sonorité autant qu'en musicalité; il faut en parler ensuite pour convenir entre divers auditeurs d'une intention d'écoute commune. Ainsi apparaît la notion d'un métalangage, qui est autre chose qu'une pédagogie ou une critique explicative : une nécessité organique, attachée à un art, de posséder une terminologie adéquate, résultant d'une analyse réaliste de ce qui est distinctif ou sensible dans les objets qu'il utilise. Les objets assemblés au niveau supérieur forment bien un langage original, qui peut se passer de commentaires ou s'élucider lui-même. Il n'en va pas de même à deux niveaux plus bas, celui des éléments constituant les objets. Leur description, la détermination de leurs fonctions, à ce niveau en tout cas, réclame un vocabulaire" (Schaeffer, 1966, p 481).

Aujourd'hui, la nécessité de rassembler un vocabulaire pertinent se renforce encore avec l'apparition récente de la notion de "design sonore" et de bases de données sonores, notamment sur le Web. C'est

le cas en particulier de celle mise en place par l'Ircam, le Studio on Line (SOL) (<http://www.ircam.fr/studio-online>). Pour ces bases de données, il faut imaginer des interfaces permettant de choisir un son ayant des caractéristiques voulues, parmi des dizaines de milliers d'échantillons (Misdariis et al., 1998; McAdams & Misdariis, 1999).

C'est ainsi que s'est formé le projet initial de cette thèse. Alors que la démarche psychoacoustique avait permis d'une part de modéliser notre perception du timbre musical sous forme d'espace multidimensionnel (Grey, 1977; McAdams et al., 1995), puis de rechercher des corrélats acoustiques de ces dimensions perceptives (Krimphoff et al., 1994), on envisageait de pouvoir trouver des corrélats verbaux de ces mêmes dimensions. Ainsi pouvait se former l'utopie d'un synthétiseur de la prochaine génération pour lequel les boutons de réglage de synthèse des sons seraient directement transcrits en langue "naturelle" et non plus au moins d'algorithmes physiques qui nécessitent l'intervention d'un ingénieur du son ou assistant de composition entre le compositeur et la machine.

### **Liens entre langage, pensée et monde**

La volonté de comprendre comment nous mettons des mots sur nos perceptions rejoint aussi une préoccupation théorique des sciences cognitives. Il s'agit là de comprendre les liens entre nos percepts et nos concepts à travers le langage. Ainsi, nous comptons sur le fait que l'analyse de la façon dont nous, humains, parlons des sons que nous percevons nous donnera accès à la fois, au sens que nous donnons au monde sonore, à la manière dont nous utilisons le langage pour le décrire et à la façon dont nous structurons notre perception.

Dans la partie théorique de ce travail, nous nous pencherons d'une part sur notre rapport au monde à travers notre activité de perception, d'autre part sur notre façon d'utiliser le langage pour décrire le monde et rendre compte de nos pensées et enfin sur ce que nous connaissons des sons, sur notre manière de les percevoir, de les catégoriser, de les identifier et d'en parler.

## **2- Définitions du timbre : un résumé de la problématique**

Ce travail, dans sa partie expérimentale, ne consiste pas à comprendre comment on parle des sons en général mais va se focaliser sur la façon dont on parle du timbre des sons, au moyen de l'analyse de trois expériences menées sur le timbre d'un ensemble de douze sons très particuliers qui sont décrits ci-dessous. Il est donc nécessaire de tenter de définir ici ce qu'est le timbre. Cette tentative de définition va révéler une partie de la problématique touchant l'étude du timbre, voire de sa résolution.

### **Démarche analytique vs globale**

Il existe au moins deux façons de définir le timbre. Ces deux démarches semblent s'exclure au premier abord bien qu'elles se complètent plutôt.

D'une part, la notion de timbre renvoie à une qualité du son : on parle de *timbre* d'un instrument de musique, on dit d'une voix qu'elle est *timbrée*. L'association américaine de normalisation (American Standards Association, 1960) donne du timbre la définition suivante : "le timbre est l'attribut de la sensation auditive qui permet à l'auditeur de différencier deux sons de même hauteur et de même

intensité et présentés de façon similaire". Le timbre est ici défini comme un résidu, comme ce qu'il n'est pas : ni hauteur, ni dynamique. Ce serait par exemple ce qui permet de différencier un son de violon d'un son de piano joués sur la même note pendant la même durée et avec la même intensité. Pourtant, en tant que qualité du son, il apparaît immédiatement comme "étrange et multiple" (Cadoz, 1991). Son caractère multidimensionnel semble se refléter dans la représentation perceptive que nous en avons, qui a été modélisé par un espace en trois dimensions obtenu après analyse multidimensionnelle des jugements de dissemblance des paires de sons (Plomp, 1970; Grey, 1977; McAdams et al., 1995). Ces études soulignent par exemple la prépondérance du temps d'attaque, de l'importance relative des différents partiels (centre de gravité spectral) et de l'évolution temporelle du spectre des sons (Krimphoff et al., 1994) dans la perception de cette qualité. Schaeffer (1966) résume la situation ainsi : "Le timbre perçu est une synthèse des variations de contenu harmonique et de l'évolution dynamique; en particulier, il est donné dès l'attaque lorsque le reste du son découle directement de cette attaque" (p 231).

Pourtant, le sens premier du timbre pour Schaeffer serait plutôt : "ce à quoi on reconnaît que divers sons proviennent du même instrument". Il semble d'ailleurs que l'origine du mot "timbre" justifie cette attitude : "au départ, il désignait une sorte de tambour comportant une série de cordes tendues et donnant au son une couleur caractéristique" (Schaeffer, 1966, p 55). Il y avait alors pratiquement coïncidence entre le mot "timbre" et la chose elle-même en tant qu'instrument de musique. Dans la même perspective, on peut aussi penser au sens de "timbre" comme marque d'origine apposée à tel ou tel objet pour indiquer sa provenance, et l'on admettra avec Schaeffer que "nous ne puissions guère, pour commencer, faire du timbre musical autre chose qu'un renvoi à l'instrument, une marque de fabrique" (ibid.).

Cette notion de timbre, que nous avons située pour l'instant dans le cadre des sons de type "musicaux" qui ont une hauteur, peut s'étendre aussi à tous les autres sons : "Un cheval possède un *timbre* bien facile à repérer. Remarquons qu'il consiste alors aussi bien en divers hennissements qu'en bruits de sabots. Un amateur de voiture reconnaît, à la fois, le régime moteur et la marque de la voiture" (Schaeffer, 1966, p 336). Dans certaines professions, comme la réparation des moteurs, la détection de bateaux par sonars ou dans le domaine de la résistance des matériaux (canalisations d'eau...), les diagnostics sont faits à l'écoute des modifications du timbre. Ceci renvoie à une des fonctions de base de l'audition qui est d'extraire des signaux sonores des informations sur la source de ces signaux, mais aussi de se situer dans l'espace en localisant les choses autour de nous. "Dans un milieu naturel hostile, il est vital de repérer d'où vient un bruit nouveau (orientation, distance) et d'en identifier la source" (Castellengo, 1994, p 82).

Ces deux types de définitions : une analytique (la qualité sonore ou la couleur du son), l'autre globale (identité de la source sonore) renvoient à deux types de théories de la perception en psychologie qui seront présentées plus loin : la théorie de l'information, démarche analytique de reconstruction d'une représentation à partir d'attributs d'une part et d'autre part la théorie Gibsonnienne de la perception directe qui est une démarche globale sans intermédiaire. Ils renvoient aussi à deux types d'intention

d'écoute, définis par Schaeffer comme "écoute naturelle" (identification) ou "écoute réduite" (qualification) (Castellengo, 1994).

### **Un phénomène complexe...**

Malheureusement, si on essaye de distinguer dans un son : "le timbre, la hauteur, la puissance, la durée, le volume, le caractère, l'intensité, l'attaque" (Scherchen, 1953), ces composantes interagissent parfois.

Ainsi, séparer le timbre de la hauteur est parfois impossible : chez les "cuivres : tout son bouché a tendance à altérer la justesse de l'accord" (Scherchen, 1953, p 142), tout comme la hauteur de l'intensité : "il faut veiller à ne pas accorder trop haut, c'est-à-dire avec trop d'aigus, la caisse claire, sinon son entrée donne aussitôt une impression d'intensification sonore, alors qu'elle doit être à peine perceptible" (p 202). Les flûtistes, comme beaucoup d'instrumentistes à vent, connaissent bien la difficulté de moduler le timbre de leur instrument en gommant les effets de tessiture sans affecter la justesse. Pour Schönberg, la hauteur est d'ailleurs une partie du timbre (Risset, 1994).

Mais il est aussi souvent difficile de séparer timbre orchestral et timbre instrumental, timbre et caractère d'un passage (ainsi que sa dynamique) qui sont décrits avec les mêmes mots : doux, agité, plein d'entrain, violent, sombre et retenu ou brillant et exubérant (Scherchen, 1953).

De plus, le timbre est lié au registre d'un instrument et au jeu de l'interprète, il serait donc plus juste de parler des différents timbres d'un même instrument. Ceci mène au paradoxe souligné par Schaeffer (1966) "qui veut à la fois que les instruments aient *un* timbre, et que chaque objet sonore qu'on en tire ait, pourtant, *son* timbre particulier" (p 233). Il faut en effet considérer que le phénomène appelé "timbre" recouvre à la fois : "une multiplicité d'instruments et, pour chaque instrument, une multiplicité non seulement de notes mais aussi de factures", ce qui permettrait de classer les sons écoutés en fonction de "l'émergence du timbre, l'émergence de l'interprétation, l'émergence des sonorités, l'étendue et la finesse du registre" (ibid.).

Il faudrait enfin considérer la modification du timbre d'un son en fonction de la salle dans laquelle il est écouté, voir du point d'écoute dans cette salle (Castellengo, 1994).

### **3- Les moyens : expérimentation et portée des conclusions**

L'expérimentation décrite dans cette thèse se situe dans le cadre de la psychologie expérimentale. Elle n'abordera pas la totalité des nombreuses composantes du timbre. Nous avons voulu aborder essentiellement les liens entre les sons et les mots qu'on utilise pour les décrire. Notre travail s'est centré sur les trois questions suivantes : Comment parle-t-on des sons ? Comment les juge-t-on sur des dimensions sémantiques ? Comment les reconnaît-on à partir de leur description ?

Nous allons présenter maintenant l'ensemble du protocole expérimental envisagé, les sons utilisés ainsi qu'une discussion critique sur les choix effectués.

### **Trois expériences**

Trois expériences ont été envisagées. Les deux premières visent la manière dont on parle des sons : on fait écouter des sons aux participants et on leur demande à travers plusieurs protocoles (verbalisations libres ou utilisations d'échelles sémantiques) de décrire leur perception. Dans une troisième expérience, on utilise les descriptions recueillies dans les deux premières phases de l'expérimentation pour les présenter à d'autres participants qui devront essayer de reconnaître quels sons sont décrits par les différents "portraits verbaux" proposés.

Ces trois expériences forment un ensemble original qui permet de "boucler la boucle" passant de la perception à la description des sons et à leur reconnaissance à partir du langage.

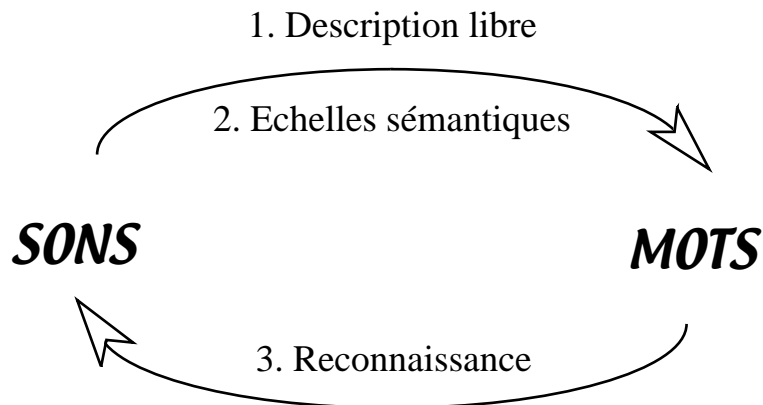


Figure 1 : Plan général de l'expérimentation

Plus précisément, la première expérience consiste à présenter 12 sons par paires aux participants et à leur demander d'une part de donner un jugement numérique de dissemblance perceptive entre les 2 sons de chaque paire, et d'autre part de décrire librement les différences et les ressemblances perçues entre les 2 sons. La deuxième expérience reprend sous forme d'échelles sémantiques 23 des adjectifs les plus utilisés spontanément dans la première expérience pour comparer les qualités des sons 2 à 2. Les participants doivent positionner chacun des 12 sons sur chaque échelle proposée, puis juger de la pertinence de l'échelle pour eux (est-elle adaptée à l'ensemble des sons ? A chacun des sons ?). Enfin, dans la troisième et dernière expérience, 201 descriptions (portraits verbaux), appartenant à des catégories différentes, sont proposées à de nouveaux participants. Ceux-ci devront associer un ou plusieurs sons à chacun des portraits puis donner un jugement sur le sens de ces portraits pour eux . Deux grandes catégories de portraits reflétant la richesse des verbalisations libres ont ainsi été testées :

- un premier groupe décrit une source sonore possible des sons entendus, soit sous forme de nom d'instrument de musique (accordéon, clarinette, flûte...) ou de famille d'instruments (famille des bois, des claviers, des cordes...), soit sous forme d'actions pouvant engendrer un son (on frappe, on froisse, on gratte...), de matériau vibrant (de l'air, de la tôle, de la peau et du métal...) ou de source sonore décrite comme combinaisons d'actions sur une matière (du vent dans un tuyau, on casse du verre, on frappe du bois...)
- un deuxième groupe est composé de qualités sonores des sons. Il reprend les adjectifs testés sous forme d'échelle sémantique dans l'expérience 2 auquel il ajoute un autre ensemble d'adjectifs utilisés spontanément pour caractériser les sons et non les comparer (agressif, chaleureux, cuivré...). Enfin, il

comprend des associations de 2 adjectifs utilisés ensembles spontanément ("aigu et métallique", "court et ramassé", "doux et froid"... ) ou bien de 3 adjectifs construits comme les plus représentatifs de chacun des 12 sons après l'expérience 2, ("doux, sourd et soufflé"...) )

Ce protocole reprend donc les deux manières de définir le timbre qui ont été retrouvées dans les verbalisations libres des participants. Les méthodologies de chacune de ces expériences sont décrites en détail dans le chapitre 4.

### **Douze sons pour représenter le timbre**

L'ensemble de cette expérimentation est basée sur 12 sons (voir CD en Annexe) qui ont été synthétisés par Wessel et al. (1987) sur un TX802 FM Tone Generator Yamaha. Ils ont été choisis parce que leurs caractéristiques tant perceptives (Krumhansl, 1989; McAdams et al., 1995; Donnadieu, 1997) qu'acoustiques (Krimphoff, 1993; Krimphoff et al., 1994) avaient déjà été beaucoup étudiées. Cela nous permettait de bénéficier dès le départ de notre étude de l'ensemble de ces caractérisations. En particulier, nous savions que ces sons variaient selon les différentes dimensions perceptives du timbre et en présentaient un ensemble varié et relativement représentatif.

Les sons ont été synthétisés soit pour imiter des sons d'instruments traditionnels occidentaux, comme la trompette, soit pour créer des hybrides, chimères intermédiaires entre deux instruments comme par exemple le *vibrone*, mélange de vibrapone et de trombone. Ils sont égalisés en hauteur, en sonie et en durée. Ce choix présente bien évidemment des avantages et des inconvénients.

1- Le fait de n'étudier qu'un ensemble réduit de sons permet de ne pas trop alourdir l'expérimentation afin qu'elle puisse se dérouler dans des temps acceptables pour les participants (chaque session ne dépassant pas 1h30, ce qui est déjà long). Le fait de maintenir ce choix pour l'ensemble de l'expérimentation nous permet de parcourir toute la boucle expérimentale décrite figure 1 : à partir du moment où les descriptions avaient été faites sur un ensemble particulier de sons, nous ne pouvions tester la possibilité d'une reconnaissance que sur le même ensemble de sons. Néanmoins, à la lumière de la complexité et de la diversité des phénomènes sonores regroupés sous le mot "timbre", il est bien évident que ces 12 sons très particuliers ne peuvent être à eux seuls représentatifs de ce qu'est LE timbre.

2- Le fait de choisir des sons de synthèse, donc pas "naturels", permet une homogénéité entre tous ces sons malgré le fait que certains sont des imitations de sons d'instruments réels, et d'autres des sons imaginaires. La présence de ces chimères est intéressante pour montrer que la démarche d'écoute est la même, quelle que soit l'intention de synthèse du compositeur et on retrouvera pour ces deux types de sons les deux types d'écoutes correspondant aux deux définitions du timbre : recherche d'une source sonore et descriptions de qualités. Néanmoins, le caractère artificiel prononcé de tous ces sons a été souvent perçu par les auditeurs comme leur première qualité, au risque de voiler les autres composantes de leur timbre.

3- Le fait de choisir d'égaliser les sons en hauteur (fréquence fondamentale identique), en durée perçue et en sonie (intensité perçue) nous permet, dans le cadre de la définition analytique du timbre de circonscrire dans ces sons seulement ce qui concerne leur timbre et donc de pouvoir donner des conclusions se rapportant à cette seule qualité. Le fait que le timbre perçu puisse entrer en interaction avec les notions de jeu, de tessiture, de hauteur, nous a encouragé dans ce sens. Néanmoins, nous ne pouvons voiler le côté rébarbatif de l'expérimentation engendré par cette répétition d'une même note ainsi que le fait que le timbre de chacun de ces "instruments" soit réduit à celui d'un seul de ses sons. De plus, le fait de devoir présenter une même note jouée par des instruments de tessitures différentes induit le fait que chacun de ces sons n'est pas toujours choisi dans le registre le plus "naturel" de chacun de ces "instruments". Par exemple, le cor anglais est en général utilisé dans un registre plus grave que la trompette. Ici, les deux sons imitant ces deux instruments (CRA et TPT, voir chapitre 4) sont présentés sur la même note.

### **Expérimentation écologique ?**

Si nous avons pris le soin de décrire les avantages et inconvénients liés au matériel expérimental utilisé dans ce travail, c'est parce que les protocoles de psychologie expérimentale (Dubois, 1994; Dubois, 1996) et l'expérimentation dans le domaine sonore en particulier sont actuellement soumis à critique (Castellengo, 1994). En effet, alors que les premiers résultats relevant de la psychologie expérimentale ont été obtenus en laboratoire dans des conditions de contrôle des différents paramètres en jeu qui ont souvent induit un décalage avec le contexte habituel d'observation des comportements étudiés, de plus en plus d'études de terrain ont montré en comparaison l'influence du contexte d'expérimentation sur les résultats obtenus (Rogoff, 1984). En opposition à un mode d'expérimentation très contrôlé, qui part du principe que c'est en analysant un problème complexe et en testant séparément les différents composants de ce problème qu'on construit des connaissances scientifiques, il s'est construit un nouveau courant d'expérimentation dit "écologique" qui essaye au contraire de conserver toute la complexité des situations naturelles et de les analyser globalement en contexte.

Il est pour autant bien entendu que même si dans certaines expérimentations, "l'utilisation d'un matériau sonore naturel et contextualisé" (Guyot et al., 1998) est soulignée, il est rare qu'elles s'interdisent d'effectuer l'écoute en laboratoire, sous casque ou dans un caisson traité phoniquement, donc relevant toujours d'une forme d'isolement et présentant donc toujours un caractère étrange et en tout cas hors contexte d'écoute "normale" de ces sons. Une étude a d'ailleurs été effectuée sur la comparaison de jugements sur l'intensité de bruits de voitures effectués "en contexte" sur le bord de la route puis sur les mêmes sons enregistrés et écoutés cette fois-ci dans le contexte du laboratoire (Schulte-Fortkamp & Weber, 1990). Les différences observées seraient sans doute encore augmentées si l'on considérait les différentes intentions d'écoute de bruits de voiture possibles : quand on conduit, quand on est passager, quand on cherche à réparer, quand on habite au bord d'une route très passante, ce qui peut multiplier à l'infini les contextes possibles, tous "écologiques".



Mais la notion même d'environnement "écologique" de l'homme est problématique et renvoi à l'éternel débat nature / culture, puisque l'une des caractéristiques principales de l'homme est justement qu'il modifie en permanence son environnement : dans nos sociétés, on habite dans des maisons, on utilise des voitures, l'eau coule du robinet, on porte des chaussures..., et tout ça nous paraît "écologique". Ainsi certaines différences, centrales dans certains champs d'expérimentation (par exemple objets "naturel" versus "manufacturés" en neuropsychologie), peuvent être questionnées.

Et cette perpétuelle modification de son environnement par l'homme est particulièrement à souligner quand il s'agit du son et de la musique. En effet, tous les instruments de musique sont des "objets manufacturés". De nouveaux instruments ont été inventés à chaque époque avec, parmi les plus récents le piano, le saxophone, avant des instruments électroniques. Nous avons déjà souligné comment de nouveaux outils apportés par l'analyse des sons ont été utilisés par les compositeurs. Il faut souligner encore l'importance de l'écoute de sons via des chaînes hi-fi, dans nos sociétés occidentales, ce qui modifie encore notre environnement sonore "écologique". Ainsi, comme le souligne Risset (1994, p 92) : "Aujourd'hui on peut se procurer nombre d'enregistrements de musiques extra-européennes. La disponibilité sans précédent de musiques d'époques et de provenances diverses brouille les perspectives historiques, spécialement chez les jeunes auditeurs qui sont nés avec cette situation. De même, les options de prise de son influencent la perspective sonore. Michel Decoust (cité par Risset, 1994) remarque ainsi qu'aujourd'hui on écoute surtout de la musique électroacoustique émanant de haut-parleurs : même s'il s'agit d'une symphonie de Beethoven, on n'entend pas ce qu'on entendrait dans une salle de concert. Que chacun de nous s'interroge sur la proportion de fois où il a entendu un orchestre en direct en situation de concert ou reproduit à travers une chaîne hi-fi.

De plus, s'il nous paraît important de garder à l'esprit l'importance du contexte expérimental tant au niveau des protocoles que du matériel utilisé afin de rester conscient des limites et de la portée des résultats obtenus, il nous paraît impossible de concevoir une expérimentation sans que le regard de l'expérimentateur influe sur le comportement observé. C'est d'ailleurs ce que rappelle Dubois (1998) : il faut "se garder de la facilité de penser que les phénomènes cognitifs puissent être 'transparents' aux modes d'interrogations" (p 10). Ce résultat a même été mis en évidence en mécanique quantique : le fait même de mesurer quelque chose induit une perturbation de son fonctionnement. Et aussi loin qu'on puisse essayer de questionner le monde de la manière la plus "écologiquement valide", on ne pourra pas enlever cette perturbation due au fait même d'observer. Ou bien il faudrait se tourner vers des méthodologies relevant de l'éthologie, menant par exemple à l'observation d'enfants en situation de jeu dans une crèche derrière une glace sans tain. (Encore faudrait-il pouvoir prouver que des enfants en situation de jeu dans une crèche présentent des comportements représentatifs d'autre chose que du fait de jouer dans une crèche...). Et les problèmes apparaissent quand il s'agit de dépouiller ce genre de données...

Il n'existe pas d'expérimentation idéale, sinon on ne pratiquerait plus que celle là. Face à cette réalité, seules deux attitudes paraissent possibles : renoncer à toute expérimentation ou en accepter les

contraintes et les garder à l'esprit comme une certaine limitation de la portée de nos conclusions. On pourrait ainsi rejoindre Krishnamurti dans *Les limites de la pensée* : "Nous agissons tout au long de notre vie dans le champ du réel et nous y constatons des distorsions : la vérité, c'est la perception de ces distorsions" (Krishnamurti & Bohm, 1975, p 30).

Nous allons donc, dans la partie théorique de ce travail, considérer les différents champs d'investigation que l'action de description de nos perceptions sonores impliquent, à savoir :

- que cela nous confronte à notre relation au monde qui nous entoure et dont nous faisons partie (chapitre 1 : être au monde);
- que cela nous mène à utiliser le langage comme outil de description et comme outil de communication de nos impressions (chapitre 2 : Parler);
- et que cela nous informe sur le monde sonore (chapitre 3 : Des sons).

Puis, dans la partie expérimentale, nous décrivons l'ensemble des expériences effectuées (chapitre 4), avant d'analyser les données dans le cadre des quatre chapitres suivants, concernant :

- la recherche de corrélats verbaux de dimensions perceptives (chapitre 5);
- un panorama de la manière dont nous décrivons les sons (chapitre 6);
- une analyse du sens et de la pertinence que les auditeurs attribuent au vocabulaire utilisé (chapitre 7);
- et la façon dont les participants associent des sons à des descriptions proposées (chapitre 8).

Enfin, nous discuterons de ce que les données recueillies ici apportent aux questions rencontrées, sur les sons, le langage et le monde tel que nous le percevons.

## Partie I : Questions et inscription théorique

Chercher à comprendre comment nous décrivons nos perceptions rejoint la question théorique des liens existant entre nos percepts et nos concepts à travers le langage. En analysant la façon dont nous parlons des sons que nous percevons, nous espérons avoir accès à la fois au sens que nous donnons au monde sonore, à la manière dont nous utilisons le langage pour le décrire et à la façon dont nous structurons notre perception. Or, depuis l'antiquité, on a coutume de représenter les liens entre le langage, la pensée et le monde (ou les mots, leur sens et les choses, ou encore le signifiant, le signifié et le référent) sous la forme d'une triade appelée triade sémiotique (Figure 2). En effet, "dans un célèbre passage du *D'interprétation*, Aristote trace un triangle : les paroles sont liées d'un coté aux concepts (ou aux passions de l'âme) et de l'autre aux choses." (Eco, 1999, p 410). Ainsi, pour le sens commun, quand on utilise le mot formé des lettres C H A I S E , on fait ainsi référence à un objet du monde qui sert pour s'asseoir et qui a en général quelques propriétés constantes (des pieds, un dossier, pas d'accoudoirs, peut se déplacer, pas trop lourd...) à travers le concept de "chaise".

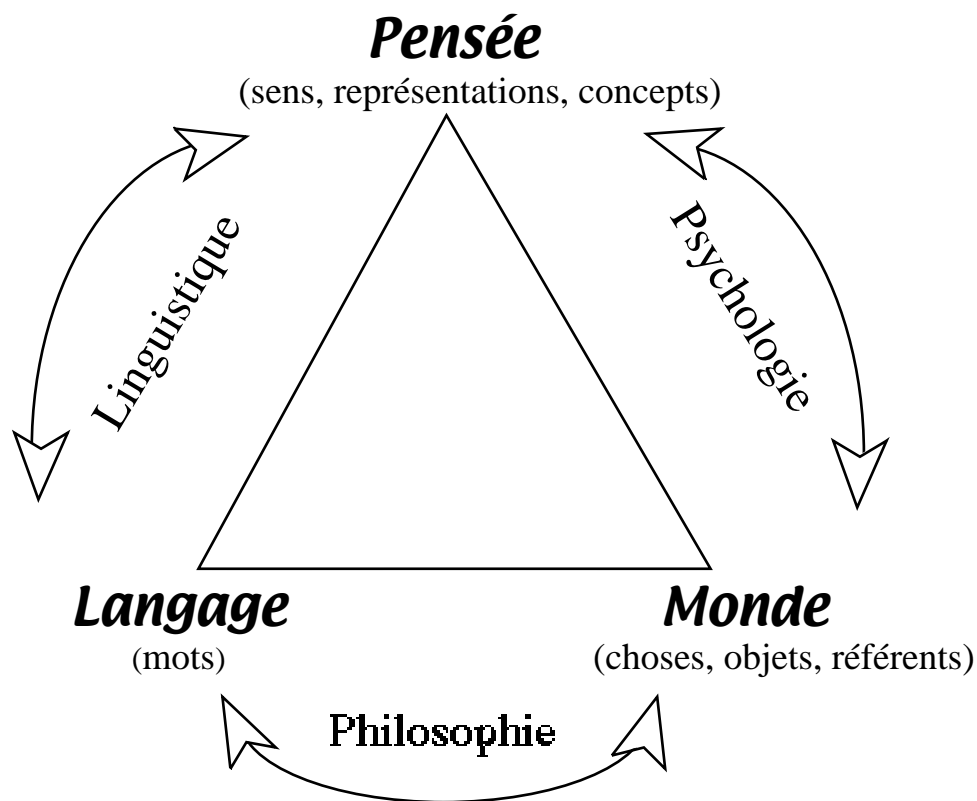


Figure 2 : Triade sémiotique revisitée

En simplifiant, on peut considérer que le lien entre les mots et leur sens relève de la linguistique, celui entre concepts et référents relève de la psychologie cognitive, les rapports entre les mots et les choses relevant de la philosophie formelle (Rastier, 1990).

Ainsi, même si le langage a beaucoup d'autres fonctions que de décrire le monde, par exemple : "éveiller une passion, provoquer une action ou en détourner, mettre l'esprit dans une disposition particulière" (Berkeley, 1991), quand il est utilisé pour décrire les choses perçues, se joue le lien entre le réel (s'il existe) et nos concepts (pensées ou représentations du monde ?). Dans notre perspective de comprendre comment on parle du timbre des sons, c'est-à-dire comment nous décrivons une de nos perceptions auditives, il nous semble donc important, avant tout, de comprendre quel est notre relation au monde, comment nous le percevons à travers les questions : Existe-t-il en dehors de nous ? Quel sorte de monde existe pour nous ? Doit-on le catégoriser ? Doit-on le représenter ? Puis nous questionnerons ce qu'est le langage, depuis l'utilisation de signes jusqu'à la construction ou l'émergence du sens d'une phrase : Quels sont les rapports entre le langage et la pensée ? Entre le langage et le monde ? Enfin nous aborderons le domaine du sonore afin de faire un état des lieux des études effectuées pour définir, catégoriser, identifier et parler des sons.

Nous verrons que le langage, de part sa structure catégorielle, se positionnera comme outil de "passage", ou de "concrétisation" d'une part des intuitions en pensées et d'autre part des interactions avec le monde en perceptions.

Si, comme le définit le programme Cognitique, l'enjeu théorique des sciences de la cognition est de "comprendre les rapports entre le fonctionnement de l'*esprit* et celui de la *matière* (cerveau, corps, ordinateur), à travers l'étude de la perception, la mémoire, le langage, le raisonnement, l'action, la planification et la prise de décision, les émotions, la conscience, etc.", il rejoint un programme mystique récurrent dans nos sociétés occidentales marquées par la dualité : comprendre et réaliser le lien entre la matière et l'esprit.



L'homme, entre matière et esprit, avec les sens et la parole comme "interfaces".

## **Chapitre 1. Être au monde**

Toute théorie de la perception repose sur une conception du monde et de nos rapports avec lui, même si cette conception est implicite seulement. Pour rendre ces a priori explicites, il nous faut donc questionner ce qu'est le monde pour nous. En effet, autant le langage que la perception font référence au monde. D'une part, la linguistique est traversée par l'idée de conditions de vérité qui découlerait de l'adéquation ou non des phrases aux états du monde, et l'idée du sens des mots pose la question de la référence, qui est toujours une référence prise dans le monde. D'autre part nous essayons de relier les mots que nous utilisons non pas aux sons tels qu'ils pourraient exister dans l'absolu mais tels que nous les percevons. Se pose la question des liens entre ce que nous percevons du monde et ce qu'il est ou pourrait être par lui-même, en dehors de nous. Dans ce chapitre, nous allons donc essayer de rendre explicite les différents types de positions sur les questions de perception, représentation et catégorisation du monde dans lequel nous vivons.

### **I. Le monde existe-t-il en dehors de nous ?**

Selon Eco (1999), le problème de l'Être est contre nature, parce qu'il ne se pose pas dans le sens commun. En effet, si Leibnitz, s'est demandé "pourquoi y a-t-il quelque chose plutôt que rien", l'idée que le monde existe semble une évidence. Pourtant, il faut bien considérer que nous n'avons pas d'autre expérience de l'existence du monde qu'à travers nos perceptions, qui seront toujours une forme d'interprétation, et à travers ce que nous pouvons en concevoir, voir en dire. On peut ainsi comprendre l'idée que "l'ontologie d'Aristote, (...) a des racines verbales : L'être, en tant que pensable, se présente à nous, depuis le début, comme un effet de langage" (Eco, 1999, p 27). On pourrait objecter que la substance est indépendante de ce que nous en disons. Mais jusqu'à quel point ? Il n'empêche que l'idée du monde, même si elle n'est qu'illusion, est opératoire dans le sens qu'elle nous permet d'agir et qu'elle est collectivement partagée par l'ensemble des humains, et sans doute en partie par les animaux. De plus, notre environnement, dans la manière dont il se présente à nous, semble suggérer des possibles et des "lignes de résistances" qui contraignent ce qu'est "la réalité".

#### I.1. Le monde existe

Dès qu'on se pose la question de ce qu'est la perception, il va de soit qu'on considère qu'elle est perception "de quelque chose", la réalité. Ainsi, la plupart des courants "réalisme, rationalisme, empirisme, phénoménisme se sont pensés dans leur conception du rapport entre réalité et perception" (Jimenez, 1997, p 11). La perception nous montre une réalité indépendante, pour les réalistes. La perception est fautive, mais on peut atteindre la réalité grâce à la raison pour les rationalistes ou grâce à

l'expérience pour les empiristes. En effet, "nous sommes à la recherche d'une substance, de quelque chose de solidement présent sous les apparences. Et cette chose, nous la cherchons dans le réel. Depuis la nuit des temps, nous cherchons, derrière les apparences mouvantes, une réalité solide et permanente qui puisse expliquer ces apparences et nous les rendre compréhensibles." (Krishnamurti & Bohm, 1975, p 23).

### **Evidence**

D'après Eco (1999), si l'être n'est pas un problème du sens commun, c'est qu'il est "la condition même du sens commun". Il cite ainsi Saint Thomas, dans son De Veritate : "Qu'il y ait quelque chose, c'est la première chose que notre intellect conçoit, comme la plus connue et la plus évidente. Tout le reste vient ensuite. Ce qui veut dire que nous ne pourrions penser si nous ne partions pas du principe (implicite) que nous sommes en train de penser quelque chose". Les animaux ont aussi cette évidence de l'être, et l'enfant l'aurait dès la naissance... et sans doute même avant ! "Aristote n'a jamais douté de la réalité du monde tel qu'il apparaît dans l'expérience quotidienne" (Eco, 1999, p 27). Cette évidence découlant de la perception, mène par exemple Dan Sperber à affirmer dans Les idées qui viennent (Droit & Sperber, 1999, p 83) : "Je soutiens que dans notre activité cognitive ordinaire, et dans l'ordinaire de la recherche scientifique, nous fonctionnons tous comme des réalistes naïfs : nous tenons pour évident qu'il peut et doit y avoir une correspondance entre nos représentations mentales et l'état de choses qu'elles représentent. Ainsi, puisque nous nous représentons le monde en couleur, nous tenons spontanément pour évident qu'il y a des couleurs dans le monde, indépendamment de notre perception." De telles évidences peuvent pourtant mener le chercheur sur des "fausses pistes".

### **Dualisme**

Le fait de considérer que la réalité existe en dehors de nous induit de manière logique une position dualiste. Il est important de réaliser la place de cette conception dans le champ des sciences cognitives. Si on doit à Descartes, l'idée que l'âme et le corps sont deux "substances" ou deux "choses" radicalement distinctes par leur nature même : "l'âme est de la pensée, le corps est de l'étendue" (Droit, 1998), cette idée est implicite dans la notion de "représentation" qui est centrale en particulier dans le courant cognitiviste. En effet, "cette idée d'un monde divisé en deux, fut en outre favorisée par le modèle de la *tabula rasa* des empiristes anglais, qui suppose tout d'abord un esprit entièrement vide qui, par la perception, se remplit de représentations sur le monde" (De la Motte-Haber, 1994, p 40). La nécessité d'avoir recours à une représentation interne du monde apparaît en effet dès qu'on imagine le monde comme extérieur et l'activité de perception comme une recherche d'informations. De plus, le dualisme cartésien, qui voit dans le langage le signe de la pensée, et qui affirme "que la seule indication certaine qu'un autre corps possède un esprit humain au lieu d'être un simple automate, c'est son aptitude à utiliser le langage de façon normale", fut considéré par Chomsky (1990) comme précurseur de sa conception mentaliste. Conception qui est encore particulièrement fertile dans le champs des sciences cognitives.

### **Monde réel, mondes possibles**

Il convient de rappeler que dans la littérature linguistique, la référence au monde se fait parfois de manière identique, que le monde en question soit considéré comme un monde réel ou un monde possible : "Voilà ce que nous entendrons par le mot être : quelque chose" (Eco, 1999, p 18). Cette "chose" sera donc considérée de la même façon qu'elle soit du domaine du réel ou du possible, indépendamment de toute question d'existence. La licorne a donc ici le même statut que les canaris qui abondent dans la littérature psychologique. On peut d'ailleurs constater que nous considérons souvent de la même manière les choses réelles et imaginaires et qu'un enfant de 4 ans vous corrigera si la corne que vous avez dessiné à sa licorne ne correspond pas à l'image "typique" qu'il s'en est fait. Nous verrons ainsi que les chimères auditives que nous avons présentées à nos participants seront considérées et perçues de la même façon que les sons qui se voulaient plus proches d'une "réalité". En effet, "La réalité est une forme d'apparence qui peut être exacte, inexacte ou tout à fait illusoire. L'habitude nous incite trop fréquemment à considérer que la réalité se confond avec *ce qui est*." (Krishnamurti & Bohm, 1975, p 22).

#### I.2. Tout n'est qu'illusion

Même si l'idée du monde réel apparaît donc comme une évidence, on peut se demander, comme David Bohm dans Les limites de la pensée (1975, p 23), si "la réalité dans son ensemble n'est peut-être dotée d'aucune substance. Elle n'a pas d'existence autonome, peut-être n'est-elle qu'un simple champ, et peut-être que derrière cette réalité se cache la vérité." En effet, comme le souligne Nyckees (1997) : "pour un sujet 'jeté dans le monde' (...), ce ne sont pas les objets qui sont premiers, mais nécessairement ses interactions avec eux, en sorte que ce qui doit primer pour une sémantique fondée sur la compréhension humaine ne réside pas tant dans les qualités sensibles, 'concrètes', de tels objets (enregistrés passivement par quelque conscience désincarnée ?) que dans les propriétés mêmes de nos interactions avec eux."

#### **La matière est "forme-pensée"**

De toute façon, nous n'avons de connaissance du monde qu'à travers nos perceptions de celui-ci. Cette autre "évidence" a été soulignée déjà par Platon dans le mythe de la caverne, mais aussi par Montaigne (cité par Droit, 1999, p 82) : "cessons de rêver : jamais nous ne pourrions voir hors de notre regard. (...) Du monde comme il est nous ne saurons jamais rien." Ceci rejoint la position bouddhiste, reprise par Krishnamurti (1999), comme quoi le monde matériel ne serait peuplé que de "formes-pensées". La réalité est alors conçue comme une projection de la pensée (de l'objet de nos pensées ou de nos réflexions). En effet, un objet naturel disposé dans son contexte prend une signification par l'intervention d'un sujet cognitif humain ou animal. "C'est le sujet humain qui leur donne un sens. Il n'y a pas de représentation du monde sans participation du sujet ni de représentation du sujet sans une conception du monde" (Basso, 2000). Cette idée que la simple présence d'un être vivant transforme le monde physique, est aussi soulignée par Merleau-Ponty (1945), qui rappelle que c'est le regard de l'homme qui fait apparaître ici des "nourritures" ailleurs une "cachette" et donne aux



"stimuli" un sens qu'ils n'avaient pas intrinsèquement. Le sens que nous donnons aux choses est d'ailleurs fortement lié à ce que nous pouvons imaginer avoir comme action avec ou sur ces "choses" (pouvoir les manger, permettre de se cacher ou de s'asseoir). Les choses n'apparaissent comme telles que indissolublement liées à notre manière d'être et de percevoir le monde. Par exemple, nous n'avons pas coutume de regarder une fleur comme une nourriture, ce qui doit être le cas des abeilles. Le monde dans lequel vit chaque espèce animale n'est donc sans doute pas le même, ni le même que le notre. On rejoint donc l'idée que "pour les phénoménistes, la *réalité* est construite par la perception et n'a pas nécessairement d'existence indépendante" (Jimenez, 1997). Le mouvement, la locomotion du sujet percevant fait "sortir" l'objet du monde, le crée en quelque sorte et crée donc aussi le monde. On comprend ainsi la contribution du mouvement à la mise en place d'un monde et d'objets (Petit, 1998). Ainsi, une des intuitions fondatrices d'un traitement phénoménologique serait le fait que les choses en dehors de nous ne sont pas déjà là "toutes faites avant nous". Ce qui nous fait revenir à l'origine du monde perçu : il faut se défaire de l'aspect de préalable. Il ne s'agit plus de "re-présenter" après coup mais d'assister à l'origine du monde perçu (Petit, 1998). Ceci sous-entend aussi l'importance de l'irréalité dans le réel. Les phénomènes psychiques ayant une dimension subjective comportent au moins autant d'irréel (convention, habitudes, fiction..) que de "réel" en tant qu'impact extérieur (ayant un corrélat) sur une surface sensible. "L'objet se donne. Il y a une participation active du sujet percevant et une offre du monde qui s'expose. La perception est notre ouverture sur le monde" (Petit, 1998).

#### **Vacuité ou néant**

Si le monde n'est qu'une illusion, ou une "forme-pensée", est-ce pour autant qu'il n'y a rien que le néant et que le monde n'existe pas ? C'est à cette question que se sont attelées les logiciens bouddhistes avec l'idée de "vacuité". Ils ont en effet un système qu'ils appellent "vérité ultime", celle que l'on découvre dès lors qu'on a mis à l'écart, de façon systématique les théories opposées. "Si vous demandez à ces logiciens : *quelle est votre thèse ?*, ils diront : *nous n'en avons pas*. La vérité ultime, c'est tout simplement la vacuité, ce qui n'est pas le vide inerte, le fait qu'il y ait dans le monde le néant, c'est l'espace ouvert, frayé par la mise à l'écart, à la limite, de l'affirmation comme de la négation. Evidemment, de cette vérité ultime qu'est la vacuité, il n'y a rien à dire" (Droit & Sperber, 1999, p 47). Cette position semble rejoindre celle des sceptiques qu'on pourrait résumer par : il convient de se méfier de ce qui se donne pour vrai, et de s'en défier toujours, si on veut s'approcher de la sagesse. "Sceptique, en l'occurrence, ne désigne pas une volonté de douter de tout, de manière continue, systématique et excessive (...), ce que soutient le sceptique antique est à la fois plus simple et plus subtil. (...) La réalité n'est pas un leurre, un mirage, une illusion, mais il n'existe 'pour l'instant' aucune voie d'accès conduisant à une connaissance assurée de son essence et de son fonctionnement" (Droit & Sperber, 1999, p 145).

#### **Monisme ?**

Si nous n'avons d'idées sur le monde qu'à travers nos perceptions, la façon dont "le monde" est agencé doit surtout refléter notre façon de percevoir, donc nos concepts et notre pensée. Par exemple,

si nous voyons des "objets" autour de nous, c'est sans doute parce que notre cerveau en a besoin pour agir. Ainsi, on peut considérer que le monde et nous ne sont qu'une seule et même chose. Il devient donc possible d'étudier l'esprit en observant la façon dont il voit "le monde". "Si l'esprit est le fondement de la matière ou du moins son pendant, on doit inévitablement l'observer grâce à des phénomènes psychiques concrets" (De la Motte-Haber, 1994, p 33). Fechner, le "père" de la psychophysique, était d'ailleurs un moniste convaincu, qui considérait le corps et l'âme comme les deux faces d'une même chose, et qui se donnait pour tâche de prouver cette relation. "Le corps, indicateur de l'âme, rend(ant) possible la mesure du spirituel" (De la Motte-Haber, 1994, p 35).

### I.3. Une illusion vitale

Néanmoins, que la réalité existe ou qu'elle ne soit qu'une projection de notre pensée, nous ne pouvons y échapper : "Je vis dans la réalité - qui n'est autre que tout ce qu'a élaboré la pensée, toute l'activité de la pensée lorsqu'elle se penche ou réfléchit sur un objet - faussé ou rationnel. Je vis depuis toujours dans ce cadre" (Krishnamurti & Bohm, 1975/1999) . Il s'agit de comprendre pourquoi nous avons besoin d'une telle réalité, en quoi elle est construite dans l'intersubjectivité et comment elle permet l'action.

#### **Nécessité**

L'ontologie est une nécessité : il faut croire que le monde existe pour survivre. "Cela veut dire seulement que certaines illusions sont plus intéressantes que d'autres. (...) De même que Spinoza affirme que nous ne désirons pas une chose parce qu'elle est bonne mais nous la jugeons bonne parce que nous la désirons, de même on pourrait dire que c'est parce que nous désirons la vérité que nous allons la transformer en quelque chose de bon, sans pour autant savoir pourquoi des savants, des écoles, des sociétés se mettent à désirer ce type de vérité" (Droit & Sperber, 1999, p 84).

#### **Une illusion partagée**

L'idée de réalisme va donc être remplacée par une autre forme de réalisme dit "faible" et défini par l'intersubjectivité (Jimenez, 1997). Ces connaissances partagées sont des conventions sociales, donc des préjugés. Mais, après avoir reconnu le fait qu'une vérité objective, extérieure, n'existe pas a priori, tout se remet en place—selon un dispositif où "tout à changé" et "rien n'a changé"—, par la reconnaissance d'une vérité conventionnelle qui est celle des relations humaines et des mots du langage. La prédicabilité des objets, l'usage pratique des dénominations s'inscrivent dans cet espace de conventions (Droit & Sperber, 1999). Ce lien entre vérité et convention est aussi souligné par Eco (1999, p 48) : "La vérité est précisément une multitude mouvante de métaphores, de métonymies, d'anthropomorphismes élaborés poétiquement et qui sont ensuite figés en savoirs. Nous nous habituons ainsi à mentir selon une convention établie..." Ainsi, pour Eco, la seule raison valable que nous aurions de penser que les Américains sont vraiment allés sur la lune, c'est que cela n'a pas été contredit par les Russes ! Connaissant tous les trucages possibles de l'image et du son, cette croyance

que nous avons ne repose en effet que sur le témoignage de trois personnes et de quelques instruments de mesure...

### **L'action reste possible**

Le savoir de la vacuité ne paralyse pourtant pas l'activité pratique, même si on y adhère plus. "Je sais qu'il n'y a pas d'objet possédant une continuité, une nature propre, une essence à laquelle correspondrait le mot que je vais utiliser, mais je parle quand même. Tout dans le réel est discontinu : le prétendu sujet, ses perceptions, la réalité du monde. Il n'empêche que je vais utiliser un seul et même mot pour désigner cette tasse et 'faire comme si' elle possédait une unité essentielle. La prédicabilité du monde coexiste avec le fait qu'il soit entièrement vide. (...) C'est parce qu'il n'y a ni l'objet ni le néant que devient possible, voire nécessaire, l'usage de la convention linguistique (Droit & Sperber, 1999, p 48). On peut essayer de comprendre la formule canonique : "Il n'y a ni tasse ni non-tasse, et c'est pourquoi l'on parle de tasse" de la manière suivante. La tasse en soi n'existe pas, parce que la notion de tasse est liée à une utilisation par l'homme (par exemple comme contenant individuel de liquides chauds buvables). Pourtant, dans la mesure où je suis en train de boire au moyen de cette "tasse", je ne peux pas dire qu'il n'y a rien. Mais l'idée de tasse n'est peut-être liée qu'au mot qui me sert à identifier et communiquer cette expérience qui peut être partagée par un autre humain.

On retrouve l'idée de l'impossibilité de connaître le monde de façon objective exprimée par Montaigne (cité par Droit, 1998) : "Du monde comme il est nous ne saurons jamais rien". Mais ce n'est pas une raison pour renoncer à en construire, par plaisir comme par besoin, des modèles explicatifs élégants et plausibles, des interprétations astucieuses et neuves. Insurmontable, notre ignorance n'est pas triste. "Nous devons apprendre à persister sereinement dans l'inquiétude d'être sans chemin" (Droit & Sperber, 1999, p 143). Ce qui est sans solution n'engendre pas nécessairement tourments et angoisses ainsi que le résume une phrase du Dalaï lama dans le film Sept ans au Tibet : "Si le problème a une solution, pas la peine de s'angoisser, s'il n'a pas de solution, s'inquiéter ne sert à rien !"

### **Lignes de résistances**

Pourtant, pour reprendre l'exemple de la tasse, l'univers de nos interactions possibles avec cette "chose" n'est pas infini. Globalement, on s'en sert essentiellement pour boire des liquides chauds. On pourrait alors se demander, comme Eco (1999) où sont les contraintes posées par le monde, car il y en a quand même. Toutes les vérités ne se valent pas. On ne peut pas dire n'importe quoi, même si "il y a toujours et encore ceux qui veulent démontrer que la terre est carrée, (...) que les statues pleurent, que certaines personnes sont capables de tordre des fourchettes à travers la télévision ou que le singe descend de l'homme" (Eco, 1999, p 52) ! On retrouve avec l'exemple de la tasse, un mode d'explication proche de la notion d'affordance développée par Gibson (1950) qui est l'idée qu'un objet "contient" l'ensemble des activités possibles avec cet objet et que c'est ce qui lui confère son sens. Par exemple, l'utilisation qu'on fait d'un tournevis pour visser une vis est imposée par l'objet lui-même, mais on pourrait très bien s'en servir pour se gratter l'oreille si ce n'était un outil pointu et

couplant. On ne peut donc pas découper le monde comme on veut : "il y a des lignes de résistance". L'exemple donné par Eco (1999, p 55) est éloquent : "C'est comme pour le boeuf ou le veau : il est découpé de façons différentes dans les différentes cultures et c'est pourquoi le nom de certains plats n'est pas toujours facilement traduisible d'une langue dans une autre. Mais il nous serait bien difficile d'imaginer un morceau qui offre à la fois l'extrémité du museau et l'extrémité de la queue de l'animal." Ainsi, les limites posées par le monde seraient dues au fait que l'imaginaire humain nous permet d'envisager plein de "mondes possibles" qui ne "marchent pas". "C'est parce que notre esprit nous offre des représentations imaginaires de mondes possibles que nous pouvons demander aux choses d'être ce qu'elles ne sont pas. Et ce n'est que lorsqu'elles persistent à être ce qu'elles sont que nous pensons qu'elles nous répondent d'un 'non' et nous opposent une limite." (Eco, 1999, p 58) Par exemple, pour un chien, une porte fermée peut prendre le sens de "porte" (dans le sens de frontière entre deux lieux qui peut être ouverte ou fermée) que parce qu'il peut imaginer qu'elle soit ouverte tout en butant dessus. On comprend ainsi l'idée que "la limite est dans notre désir, dans notre être - tendu - vers une liberté absolue" (Eco, 1999, p 58).

### **Voie moyenne**

Il faut donc plaider "pour une voie moyenne entre l'objectivité, postulant un monde donné de traits à représenter, et le solipsisme, niant toute relation avec le reste du monde" (Varela, 1989, p 223), ce qu'il appelle aussi la position de la poule : le monde extérieur existe, le système perceptif doit le capter, et la position de l'oeuf : le système cognitif crée son propre monde et toute son apparente solidité repose sur les lois internes de l'organisme. Ceci est peut-être possible du point de vue de l'"autonomie", pour qui "le système et son monde émergent en même temps". Avec pour conséquence que "notre monde et nos actions sont inséparables", idée que l'on retrouve dans les philosophies orientales "Ce que vous êtes dans la vie quotidienne est de la plus haute importance (...) Car c'est de cela, de ce que vous êtes que vous créez le monde" (Krishnamurti, 1967, p 78). Ainsi, "les conceptions sur la connaissance et le vivant véhiculent aussi une signification politique et sociale" (Varela, 1989, p 224).

## **II. Catégorisation et représentation interne du monde ?**

En développant de plus en plus d'instruments de mesure du réel, l'homme a souvent constaté que ces instruments présentaient des continuums là où nous "voyons" des objets. L'idée de voir des frontières, lieux de discontinuités, dans ce continuum n'est pas du tout évidente à faire faire par un robot, une caméra, un analyseur vocal... On peut se poser la question de savoir si notre façon de "capter" les informations du monde présente la même difficulté. Ce point de vue nécessite de considérer la perception comme une réception de données venant de l'environnement. Nous essaierons de présenter ici les réponses aux questions sur la nécessité de segmenter, de catégoriser, de représenter, le monde qui nous entoure et dans lequel nous vivons, en fonction des différentes théories de la perception : théorie de l'information, Gestalt, perception directe et phénoménologie.

Enfin, nous essayerons de comprendre en quoi le langage peut être considéré comme un outil privilégié pour catégoriser.

## II.1. De la nécessité de catégoriser le monde ?

Le monde est-il continu ou rempli d'objets ? Il est difficile d'y répondre dans la mesure où il semble qu'on puisse en avoir deux types de "représentations" selon qu'on consulte des appareils de mesure construits par l'homme ou qu'on fait confiance à notre propre expérience vécue. Ainsi par exemple, quand nous écoutons une phrase prononcée par quelqu'un, nous identifions spontanément un certain nombre de mots alors que c'est difficile actuellement pour une machine. De même, des recherches sont en cours visant à extraire des objets d'images recueillies par une caméra. Il semble que l'opération de catégorisation soit une forme d'économie cognitive liée une fois de plus aux activités humaines.

### **Continuum et segmentation**

Dans des expériences de modélisation de notre cognition par des robots, notamment de notre capacité à dénommer des objets, le robot doit d'abord trouver un moyen de segmenter le continuum fourni par ses capteurs (Steels, 1998). En effet, pour un robot, le monde réel pose deux problèmes : les objets ne sont pas "clairs" et ils ne viennent pas avec des mots. Le monde réel est en effet continu et non pas segmenté. Il ne faudrait néanmoins pas tomber dans le piège de croire que parce qu'on modélise un comportement d'une certaine façon, ce comportement se fait réellement de cette façon chez l'humain. Il est en effet tout à fait discutable de penser que pour un humain, les objets ne sont pas "clairs" et ne viennent pas avec des mots. Nous verrons que l'homme naît déjà dans un monde de langage rempli de sens. Reste que pour des robots équipés d'une caméra avec deux degrés de liberté, les données entrant en continu, il s'agit de les segmenter et de les catégoriser. Si une chose "change" sur un fond immobile (par exemple à cause d'un mouvement), elle "saute aux yeux". En filtrant les parties de l'image en fonction de leur luminosité, on crée des segments, équivalents des "objets", puis on passe par une phase de catégorisation. Le robot essaye alors d'utiliser les différences qu'il connaît pour distinguer un objet d'un autre dans le contexte. En cas d'échec, l'agent utilise un trait supplémentaire pour couper le monde en deux parties, une avec, l'autre sans le trait supplémentaire (Steels, 1998).

On peut néanmoins se poser la question de savoir si pour nous, "le monde est un continuum et non une série d'états discrets (s'il était donc segmentable et non segmenté)" (Eco, 1999, p 41). Dans le cas de la vision, un premier mode d'explication s'est inspiré de la métaphore de la photographie (O'Regan, 1992). Dans cette perspective, la perception servait à réceptionner des informations à la manière d'une pellicule photographique pour construire une image fixe du monde extérieur. On s'achemine actuellement vers une conception proche du point de vue phénoménologique, qui considère la vision comme un flux continu d'esquisses, provenant des multiples points de vue dûs aux saccades oculaires. On pourra donc se demander à nouveau comment nous voyons des objets dans ce flux (Tijus, 1994), de même que nous voyons des couleurs alors qu'elles n'existent

manifestement pas dans le "monde réel" qui présente par contre des modifications de luminosité, d'intensité, de contraste, de longueur d'ondes absorbées par les différents corps présents.

### **Voir le même à travers le différent**

Pour expliciter la notion phénoménologique "d'esquisse", on peut prendre pour exemple la perception d'un dé à jouer. On ne voit jamais toutes ses faces à la fois. A chaque perception, nous en avons une nouvelle esquisse (c'est à dire un vécu perceptif) qui pointe vers une autre esquisse. Chaque tranche de "vécu-perception" comporte un renvoi vers la tranche ultérieure. La perception est donc dynamique et subjective (Petit, 1998). Considérant la multitude de points de vue engendrés par les saccades oculaires, qui sont fondamentales dans notre perception visuelle (si l'oeil s'arrête, il ne "voit" très rapidement plus rien), il faut bien considérer comme central, le fait que "nous percevons le même à travers le différent" (Petit, 1998). Pour reprendre l'exemple du dé : "qu'il soit immobile, qu'il pivote, qu'il soit vu de face ou avec des faces différentes, on perçoit le même hexaèdre. L'acte de perception relie ensemble toutes les perceptions en un seul solide, ce qui aboutit à l'idée d'unité de l'objet" (Petit, 1998). On peut alors poser la question, comme Husserl (1930), de la genèse de cette unité de l'objet chez l'enfant : d'où vient l'idée de l'objet permanent ? A partir de quel moment le bébé a-t-il une notion de la constance de l'objet ?

### **Economie cognitive**

La notion de catégorisation a été introduite en psychologie cognitive à partir de considérations sur les capacités perceptives : on intègre dans une même classe des expériences différentes comme mode d'économie cognitive. En tout cas, il semble que si on a la possibilité de "garder" les multitudes d'images particulières, ce qui semble avoir été le cas du mnémoniste étudié par Luria (1970) dont la mémoire semblait infinie, on est handicapé dans les fonctions de reconnaissance liées à la catégorisation. En effet, le mnémoniste, alors qu'il pouvait, 20 ans après, rappeler sans problème les chiffres d'un tableau qu'on lui avait demandé d'apprendre pour les besoins d'une expérience, avait des difficultés à reconnaître les visages (peut-être même à les "voir") : "Ils sont tellement inconstants, disait-il, ils varient selon la disposition d'esprit du moment de la rencontre; ils changent constamment de couleur, se brouillent, et il devient difficile de se les rappeler" (Luria, 1970).

Pour pouvoir considérer des objets, il semble donc nécessaire de gommer le multiple des points de vue toujours différents. Ceci représenterait un premier niveau de catégorisation. Il semble en effet, qu'il y ait au moins trois niveaux de catégorisations. "Le premier niveau pour établir l'unicité de l'objet. Un second niveau pour l'inclure dans un groupement à partir de caractéristiques physiques, ses différentes parties par exemple, qui correspond au niveau de base (Rosch et al., 1976), et un troisième niveau superordonné au niveau de base, même si ces niveaux ne sont peut-être pas indépendants" (Tijus, 1996). La constitution des autres niveaux de catégorisation semble relever aussi d'un principe d'économie cognitive. C'est ainsi par exemple qu'on peut voir "cette poule-ci", reconnaître que c'est "une" poule, pour enfin la considérer comme un membre possible de la basse-cour, des animaux domestiques, ou des oiseaux...

### **Catégories et activité humaine**

On peut remarquer que les poules sont plus facilement considérées comme des volailles que comme des oiseaux. Est-ce dû au fait qu'on les voit rarement voler ou à l'utilisation principale qu'on en fait ? En effet, "nos catégories sont toujours inséparables de l'activité du sujet connaissant, c'est-à-dire, pour ce qui nous concerne, de l'activité humaine. (...) On pourra en conclure que, pour comprendre nos catégories, il importe de les rapporter systématiquement à l'un ou l'autre des multiples domaines de l'activité humaine. Encore convient-il de se rappeler que nos enquêtes ne pourront guère nous livrer d'accès qu'aux catégories constituées dans le langage" (Nyckees, 1997, p 109). On retrouve ce point de vue chez Eco (1999) pour qui "tout énoncé sur ce qui est et sur ce qui pourrait être implique un choix, une perspective, un point de vue. (...) Nos descriptions sont liées à ce que nous sommes (matériellement et culturellement)". De plus, contrairement à l'idée que nous avons tous à peu près les mêmes catégories taxinomiques, celles du dictionnaire, il semble que les catégories aient des "extensions variables" qui se fixent dans le cadre d'une tâche et d'un contexte (Tijus, 1996). Par exemple, si on peut acheter une lampe dans un magasin de meubles, la lampe ne sera pas incluse dans la catégorie "meuble" dans l'expression "on va pousser les meubles", mais le sera dans "je vais déménager mes meubles".

### II.2. De la nécessité des représentations mentales ?

Dès qu'on considère que le monde existe en dehors de nous de manière séparée, nous devons pour survivre nous en faire une représentation interne et la perception est considérée comme le moyen de construire ces représentations. En effet, "à l'idée de représentation est sous-jacente - même si cela n'est pas explicite - l'idée de coupure épistémologique entre une réalité extérieure et un sujet cognitif" (Basso, 2000). On peut donc ici retomber dans la discussion précédente de savoir où pourrait bien se situer cette coupure dans la mesure où le sujet cognitif est lui-même partie intégrante de l'univers qu'il prétend objectivement observer. L'idée et la nécessité de représentation mentale traverse donc, elle aussi, les différentes théories de la perception.

### **Théorie de l'information**

Il faut considérer la théorie de l'information, comme dominant actuellement le champs des sciences cognitives. Il s'agit essentiellement de considérer que le système nerveux recueille des "informations" en provenance de son environnement, qu'il les "traite", pour aboutir à une "représentation" du monde extérieur à l'intérieur du cerveau de l'homme ou de l'animal (Varela et al., 1993). La figure 3 en présente une version "légèrement" satirique mais néanmoins assez réaliste.

Il s'agit en effet d'une vision analytique et séquentielle de la perception comme prise d'informations sur l'environnement qui induit l'idée de réception passive, comme une impression sur une cire vierge, les "traitements" étant effectués en aval. Cette conception peut paraître simpliste dans la mesure où elle ne conçoit pas de traitements en parallèles, ni simultanément. Elle s'harmonise bien avec une conception modulaire du cerveau (Fodor 1983), dans la mesure où chaque étape du traitement de l'information pourrait être vue comme un sous-module indépendant des autres.

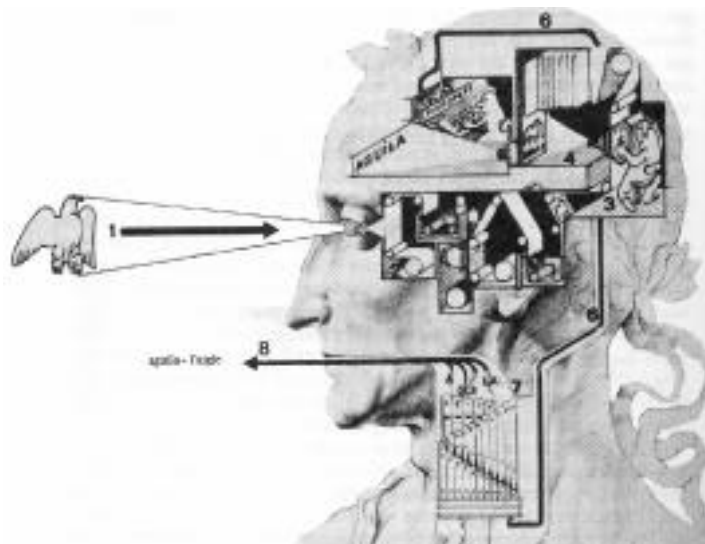


Figure 3 : Théorie de l'information (extrait de Varela, 1989)

Nous verrons que nous n'avons pas échappé à cette façon de concevoir la perception dans notre recherche sur comment on parle des sons. Nous sommes en effet parti de la conception d'échange d'informations entre représentations perceptives et sémantiques (voir figure 4).

son -> représentation perceptive -> représentation sémantique -> lexicque -> verbalisation
--

Figure 4 : Parler des sons d'après la théorie de l'information

Nous verrons dans notre discussion comment cette représentation du problème a au moins l'avantage de permettre de questionner nos résultats et comment ce qu'ils apportent la remet en cause. En ce sens, nous suivrons l'idée de Sperber (1999) pour qui "ces représentations que les organismes humains ont d'eux-mêmes sont à expliquer plutôt qu'à prendre comme des évidences sur lesquelles on pourrait fonder une explication".

En effet, on ne peut passer sous silence que, par exemple dans le cas de la vision, le premier "site" où arrivent l'information en provenance de la rétine, le Corps Genouillé Latéral (CGL), reçoit 80% de ses "informations" du cortex. Ce qui fait qu'une bonne partie de ce que "voit" le CGL provient précisément de l'endroit même où sa propre activité est en train de se projeter (Varela, 1989). Il est d'autre part inutile d'imaginer nécessaire de construire une image du monde à l'intérieur de soi, alors que justement le monde est disponible pour que nous puissions l'interroger en cas de doute (O'Regan, 1992).

On peut alors s'interroger sur la genèse de cette conception. A l'origine de l'idée selon laquelle le système nerveux détient tout un stock de représentations, d'origines génétiques ou ontogénétiques (innées ou acquises), "on peut trouver l'idée que lorsqu'on étudie le comportement animal, on est frappé par le fait que dans nos descriptions, tout se passe comme si les interactions passées semblaient



déterminer la conduite actuelle, comme si ces interactions s'inscrivaient comme modification du système nerveux, et agissaient par relation causale sur le comportement." (Varéla, 1989, p 146). Pourtant, "la perception du goéland ne correspond pas à l'idée d'une pure réception d'informations, elle apparaît finement structurée et activement structurante. (...) L'idée d'une perception qui soit simplement *réceptive*, passive, ne permet pas de rendre compte de tels phénomènes" (Cyrulnik, 1995, p 29). Ceci rejoint la thèse exprimée par Varela (1989, p 154) de la manière suivante : "La perception sensorielle ne peut pas être comprise comme un processus n'ayant que des entrées. La perception et l'action ne peuvent pas être séparées parce que la perception exprime la clôture du système nerveux." En ce sens, "ce sont les gens et non les yeux qui voient. Les caméras et les yeux sont aveugles" (Hanson, 1958). C'est-à-dire que "le bruit en provenance de l'environnement *devient* objet, (...) parce que la perception équivaut à construire des invariants par un couplage sensori-moteur." La perception serait vue comme "un processus de compensation qu'effectue le système nerveux au cours d'une interaction." Ainsi, par exemple, "l'organisme n'extrait pas la distance des caractéristiques de l'environnement ; au contraire, par un processus de compensation des perturbations, il *engendre* les distances perçues, comme un mode de comportement compatible avec l'environnement" (Varela, 1989, p 166).

#### **Le tout est plus que la somme des parties**

On peut remarquer les analogies entre la théorie de l'information, constructiviste, et l'empirisme du 18ème siècle. Cette vision analytique et séquentielle a déjà été questionnée au début du siècle. "Vers 1900, l'idée que la capacité de connaître l'homme pourrait être expliquée par la description des structures de la conscience, faite à partir de sensations isolées, est une idée qui a perdu de sa vigueur" (De la Motte-Haber, 1994).

C'est ainsi que va se dessiner l'idée d'émergence de propriétés complexes à partir d'un ensemble de propriétés plus simples, reprenant l'idée aristotélicienne que le tout est plus que la somme des parties. Cette théorie, qui va prendre le nom allemand de Gestalt, est d'ailleurs partie d'études sur la transposition des mélodies (Ehrenfels, 1890). En effet, si on transpose une mélodie, c'est-à-dire si on décale toutes ses notes d'un certain nombre de notes, on peut reconnaître cette mélodie comme la même, bien qu'aucune des notes présentes ne soient les mêmes. D'autres études ont suivi, notamment l'observation de mouvement apparent (Wertheimer, 1923) engendré par des spots lumineux fixes, qui est à l'origine même du principe du cinéma, et son équivalent musical : Kurth (1931) décrivait "une impression de mouvement apparaissant lorsque les sons d'un accord de neuvième sont joués de façon arpégée." Cette école a élaboré plus d'une centaine de lois de la perception : simplicité, groupement, prégnance, tendance à la bonne forme, qui peuvent être considérées comme des analogues des notions de catégories, pertinence ou saillance.

#### **Perception directe**

Avec la théorie de la perception directe de Gibson (1950), c'est cette fois-ci la notion de représentation qui est remise en cause. Le monde contient plus d'informations qu'il n'en faut pour le comprendre, il n'est donc pas nécessaire d'en reconstruire une image à l'intérieur de soi. La perception est une mise

en correspondance directe entre les structures de la sensation et les réponses adaptées. "Puisque le réseau optique ambiant apporte sans aucune ambiguïté la structure du monde environnant, le sujet qui perçoit n'a nullement besoin de 'traiter l'information' ou de 'construire une signification'. Cette information est déjà dans la structure de la lumière, la signification est déjà dans la réaction du sujet. Puisque cette structure est directement mise en correspondance avec cette réaction, le sujet n'a aucun besoin de se représenter perceptivement le monde" (Jimenez, 1997, p 17). Pour Gibson, la signification est irrépressible pour les êtres humains, "la perception signifiante apparaît même lorsque la tâche ne l'impose pas". Par exemple, "devant des peintures de cadavres en décomposition, le peintre Bartolomé Estéban Murillo disait qu'il fallait se boucher le nez pour pouvoir regarder ces deux tableaux. Tout se passe comme si leur perception visuelle s'accompagnait de l'avènement irrépressible de la signification de ce qu'ils représentent et des propriétés, entre autres olfactives, de cette signification" (Jimenez, 1997, p 20).

On trouve ainsi l'idée d'*affordance*, c'est-à-dire que les objets contiennent en eux-mêmes déjà toutes les actions possibles sur eux (à quoi ça sert, ce que cela nous procure, etc.) et que cela constitue leur sens. Ainsi par exemple, une chaise contient l'idée qu'on peut s'asseoir dessus mais qu'on ne peut pas s'en servir pour couper quelque chose.

#### **Recherche d'indices**

A l'inverse de l'idée d'une perception comme réception passive d'information, on trouve l'idée que la perception serait au contraire une recherche d'indices dirigée par nos attentes. Ainsi, la notion d'"intention", dans le sens de visée, comportement anticipateur, est définie comme suit par la phénoménologie : "Les intentions sont des projets sur le monde, vides au départ, et qui se trouvent remplis par le monde. Nous avons des intentions (peu différentes du sens de 'mauvaise intention'). Nous nous orientons vers un objet avec l'intention de le traiter comme un objet unique par exemple" (Petit, 1998) . Cette conception que la perception est dirigée par des attentes, notamment de situations probables, est étayée dans le domaine de la vision par la mise au jour de toute une série de présupposés ou biais, de la perception visuelle. Par exemple, a priori la source lumineuse est fixe, à l'infini ou à 30 cm devant l'objet, l'observateur est dessus, il regarde l'objet dans une position générique et non particulière (Mamassian, 1996).

Ceci rejoint l'idée que la perception est essentiellement une simulation du cerveau, on ne prend des informations dans le monde qui nous entoure que de temps en temps, pour confirmer des hypothèses qui sont fortes. "L'important, ce n'est donc pas la liste des capteurs mais les questions que le cerveau curieux pose au monde à partir d'hypothèses qu'il élabore et de tâches qu'il se propose d'accomplir." (Berthoz, 1997, p 197).

#### **Dynamique, flux**

La perspective phénoménologique, reprend l'idée que le sujet percevant n'est pas un récepteur passif de stimulations externes mais qu'il est centre d'action et d'attribution de sens. La perception, c'est donner un sens. En effet, les choses n'étant pas déjà là, toutes faites, avant nous, il ne s'agit plus de 're-présenter' après coup mais d'assister à l'origine du monde perçu. La perception d'une chose n'est

pas un *double* de cette chose dans l'esprit. La perception est constituante des choses : il n'y a pas des choses déjà données (Petit, 1998). L'interprétation est première.

La notion de *flux* est alors centrale, car notre expérience est vue comme un flux. Même si je ne bouge pas, je bouge toujours un peu, à travers par exemple des mouvements oculomoteurs. La différence entre réception et activité n'a plus lieu d'être. La sensation est un flux d'esquisses qui doit être maintenu jusqu'à la phase ultérieure, c'est-à-dire jusqu'à ce qu'on ait la perception d'un objet permanent. Il intègre le mouvement dans la constitution de l'objet perçu. Par exemple, les mouvements oculaires sont constituants pour l'objet visuel, le mouvement est essentiel au "voir" : si on immobilise la rétine, on ne voit plus ! "L'activité perceptive, comme ressaisie de la constance de la chose, lutte contre ce principe de flux" (Petit, 1998).

### II.3. Le langage est créateur

Il s'agit ici de s'intéresser à l'influence du langage sur notre perception du monde. En effet, le fait de nommer quelque chose sous-entend toujours une opération de catégorisation. Merleau-Ponty affirme en effet (1945, p 205) que "nommer un objet, c'est s'arracher à ce qu'il a d'individuel et d'unique pour voir en lui le représentant d'une essence ou d'une catégorie". Si le langage ne peut se concevoir sans catégorisation, la question est de savoir si, réciproquement, on peut envisager une perception pure, indépendante du langage ? Au contraire doit-on considérer que c'est le langage qui participe à la création de nos concepts à travers lesquels nous percevons le monde ?

#### **Perception pure ?**

Si la perception du monde passe par un processus de catégorisation, on peut se poser la question de l'indépendance ou du lien qu'entretiennent ces catégories avec le "mot" qui les désigne dans le langage. En espérant ne pas tomber dans de l'anthropomorphisme, on peut considérer que les animaux possèdent des catégories perceptives permettant notamment la dichotomie proie/prédateur, mais sans doute beaucoup plus fines. Pour Nyckees (1997), "ces catégories sont différentes des catégories sémantiques car on n'a pas d'accès direct à ces catégories perceptives quand nous parlons. Dans cette perspective, les catégories sémantiques seraient reconstruites aux fins du langage dans une logique d'interaction. Ces catégories perceptives 'pures' précéderaient les catégories linguistiques de façon historique mais une fois que le langage est là, plus rien n'est tout à fait pareil, il ne peut plus y avoir de perceptif pur." En effet, pour Nyckees, "il y a un arbitrage collectif dès qu'il y a perception". Depuis que le langage existe, mes perceptions sont contaminées par les catégories dans lesquelles mes prédécesseurs et mes éducateurs catégorisent le monde. On naît dans un monde qui est déjà baigné dans le langage. Ainsi, on peut concevoir le langage comme "une mémoire à long terme pour l'individu" (Edelman, 1992) ou "pour le groupe" (Nyckees, 1997).

Si nos perceptions sont contaminées par le langage, on peut comprendre cette remarque de Cyrulnik (1995) sur l'intérêt d'introduire des caméras dans la méthode éthologique : "A l'oeil nu, on ne voit que ce que l'on pense. Dès l'instant où l'on introduit la caméra, on voit des choses que l'on avait sous les yeux et qu'on n'apercevait pas." De même, des études faites sur les rapports de perception

d'Ovnis montrent des liens très intéressants entre ce que les gens disent avoir vu et les préoccupations de la société au moment donné (Jimenez, 1997). Notamment, il semble exister des "schémas cognitifs" d'OVNI largement partagés par les occidentaux, dont la "précision est déterminée par l'intérêt qu'on porte au phénomène et par la lecture de livres et le suivi d'émissions concernant le sujet." On "voit" donc en présence de notre culture, de notre société.

Pourtant, il y a peut-être danger d'écraser le réel sur le réel perçu alors qu'il faudrait souligner la différence entre la perception et la possibilité de mise en mots. En effet, toutes les catégories ne sont pas lexicalisées, par exemple, la catégorie des "objets qui servent à écrire" (Tijus, 1996). Le problème des liens entre nos concepts et nos percepts peut s'énoncer de la manière suivante : "est-ce que je tire la définition de l'évidence que m'offre la sensation (...) ou bien est-ce la préconnaissance de la définition qui m'autorise à abstraire l'essence ?" (Eco, 1999). Ceci rejoint le questionnement de Krishnamurti, qui interroge à maintes reprises "peut-on regarder sans mots ?". "Avec quels yeux regardons nous ? Les seuls yeux des mots ? Les seuls yeux de la mémoire ? (...) Savez-vous ce que signifie : regarder un arbre et n'avoir aucune pensée, aucun souvenir interférant avec votre observation, vos sensations, votre sensibilité, votre état nerveux dans ce moment d'attention, de sorte qu'il y ait seulement l'arbre et non vous qui regardez cet arbre ?" (Krishnamurti, 1994, p 34).

### **Nominalisme**

L'idée d'arriver à considérer comme identiques des vécus différents, c'est-à-dire de former des catégories, peut être mise en rapport avec la question de la réalité des universaux qui a été centrale au moyen âge (Nyckees, 1997). Cela renvoie au débat en philosophie du langage entre *Nominalisme*, *Réalisme*, *Conceptualisme* et *Conceptualisme linguistique*. Le réalisme des universaux défend l'idée que les idées générales (le concept d'"homme" ou celui de "cheval", par exemple) ont une existence objective indépendante des hommes et de leur langage. Il présuppose que notre langage soit lié à la réalité objective, ce qui sous-entend un principe divin qui va faire la jonction entre les deux. Le nominalisme s'est voulu une réaction anti-réaliste. Pour Guillaume d'Ockham : "il s'agit de distinguer nettement entre le langage et les choses. La philosophie doit travailler à ne plus confondre les signes qu'emploie notre pensée avec des objets réels, à ne plus prendre nos manières de dire pour des modalités de l'être" (cité par Droit, 1998, p 80). On use en effet de manière absolument différente d'un terme apparemment semblable, quand on dit que l'homme traverse la rue, que l'homme est un animal raisonnable, ou que l'homme est un mot de cinq lettres. Les idées générales n'ont pour Ockham aucune unité ni aucune existence réelle. L'"humanité" ou la "chevalité" ne sont pas des essences subsistant par elles-mêmes, ou dans l'entendement divin, dont il serait pertinent de se demander comment elles se réalisent dans tel ou tel individu. Purement relatifs à notre esprit et à notre langage, ces signes ne renvoient à rien de réel. "Pierre Alfiéri a montré comment toute la pensée du philosophe s'organisait autour de trois concepts : singulier, série, signe. (...) Le monde selon Ockham est avant tout discontinu, atomisé, constitué de choses singulières, c'est-à-dire uniques - cette pierre, cet arbre, cet homme. (...) Chaque chose est une pure singularité dont nous faisons l'expérience. Le problème est alors de trouver ce qui, dans notre expérience, permet de constituer des séries - les pierres, les arbres, les hommes - donnant leur unité mentale à ces rassemblements d'unités

éparses, chaque série étant désignée par un signe - les termes 'pierres', 'arbre', 'homme' - il s'agira de savoir quelles combinaisons de ces signes peuvent se référer adéquatement à telle ou telle réalité singulière" (Droit, 1998, p 81). On peut néanmoins distinguer trois formes de nominalisme (Nyckees, 1996) : le nominalisme au sens strict, pour lequel les idées générales n'ont pas d'autres existences que linguistique, le conceptualisme de Locke, pour lequel les idées générales correspondent à des conceptions mentales, et le conceptualisme linguistique, pour lequel les idées générales sont formées par l'esprit, mais l'esprit humain est aidé par les signes linguistiques (Condillac, 1749 ; Edelman, 1992).

#### **Le langage crée les concepts**

Tyvaert (2000) reprend l'idée de Condillac en montrant comment le mécanisme de la référence permet de concevoir une apparition naturelle des concepts comme provoquée par l'usage des langues. "L'exercice langagier qui suppose la langue, apparaît comme le moteur réalisant progressivement l'émergence des pensées. C'est dans l'observation du fait langagier, dans l'échange discursif qui le réalise, que petit à petit se définissent pour les sujets humains leurs pensées. (...) Perception du monde et perception du langage (qui relèvent originellement des mêmes capacités de l'espèce) provoqueraient petit à petit et de plus en plus rapidement la constitution des connaissances, sur la base du traitement psychique des sensations et en fonction d'une accélération provoquée par l'exercice du langage" (Tyvaert, 2000). La genèse de ceci serait due au fait que "l'enfant entend les adultes qui sont autour de lui énoncer des phrases qui sont alors, en première analyse, comparables à des mots étiquetant des faits sensibles. Il semble alors tout à fait inutile d'introduire des entités intermédiaires de type conceptuel qui seraient ni de l'ordre du monde, ni de l'ordre des mots : on ne voit pas pourquoi on devrait le faire puisqu'il est possible de rendre compte directement de l'association de la phrase au phénomène" (ibid.). Alors que les verbes renverraient directement à des situations vécues dans le monde, "les noms renvoient eux à des concepts (renvoi qui s'instaure naturellement au cours de l'acquisition de la capacité langagière), et ces concepts ne sont appréhendés par l'entendement que parce qu'ils sont nommés" (ibid.).

#### **Le langage comme outil de discrétisation du monde**

Pour Eco (1999), "si le monde était un continuum et non une série d'états discrets (...) la segmentation pourrait alors être l'oeuvre de l'Esprit lui-même qui, par limitation propre, ne pourrait penser le continuum qu'en le segmentant (...) afin de rendre le monde homologue à la nature discrète de son système de symboles". Ainsi, si le réel existe mais de manière non déterminée (Basso, 2000), c'est la détermination : un acte conscient fait par un sujet cognitif qui "crée" l'objet déterminé. Pour cela, le sujet cognitif a besoin d'"un langage opérant ce découpage, cette différenciation du réel". "Puisqu'il n'est pas possible d'apporter dans la discussion les choses elles-mêmes, mais qu'au lieu des choses nous devons nous servir de leur noms comme de symboles, nous supposons que ce qui se passe dans les noms se passe aussi dans les choses (...) Or, entre noms et choses il n'y a pas ressemblance complète : les noms sont en nombre limité, ainsi que la pluralité des définitions, tandis que les choses sont infinies en nombre. Nous n'avons que peu de noms et peu de définitions pour une

infinité de choses singulières. Le recours à l'universel n'est donc pas une forme de pensée mais une infirmité du discours. Le tragique vient de ce que l'homme parle toujours de façon générale de choses qui sont toujours singulières. Le langage nomme en recouvrant l'évidence indépassable de l'existence individuelle." (...) "La tragédie de l'être, ce n'est pas qu'il soit seulement un effet du langage. C'est que le langage ne réussit même pas à le définir" (Eco, 1999, p 29).

### III. Résumé

Nous sommes donc partis de l'idée, proche du sens commun, que le monde existe en dehors de nous, pour nier cette assertion, reconnaissant que le monde tel qu'on le connaît n'existe qu'à l'intérieur de nous et est une sorte de reflet de notre façon de penser. Puis nous avons convenu que la prédicabilité de nos actions lors de nos interactions avec le monde lui confère une certaine objectivité et rend possible la construction de ce monde intersubjectif. Nous garderons à l'esprit, dans la partie expérimentale de cette thèse, l'idée que ce qu'on perçoit est toujours teinté de notre façon de vivre (comme individu et comme société). En effet, le sens des objets est lié à ce que nous en faisons. On comprend ainsi comment les verbalisations censées décrire des sons extérieurs à nous vont plutôt exprimer nos liens avec eux : façon de les produire, conditions d'écoute, affects et images déclenchées par leur audition...

Nous avons ensuite considéré l'activité de catégorisation qui semble intervenir à plusieurs niveaux lors de la perception. C'est tout d'abord ce qui permet de "voir" l'unité de l'objet dans le flux continu des stimuli, de "gommer" la spécificité de chaque point de vue et, en considérant comme semblables des situations différentes, à la fois d'effectuer une "économie cognitive" et de permettre la reconnaissance d'un objet. La catégorisation est aussi le processus qui permet de considérer un objet comme représentant d'une classe de "choses" du monde qui partagent le même type de "fonctions" ou avec lesquelles nous engageons le même type d'interactions.

Puis nous avons abordé les idées d'émergence et d'affordance, proposées comme alternative à des conceptions respectivement analytiques et utilisant l'idée de représentation, qui peuvent être vues comme le processus qui attribue du sens à nos perceptions. Dans ce contexte, la perception ne sera pas vue comme un moyen de capter des signaux de l'environnement, mais comme un processus actif de construction du sens (donc du monde), notamment à travers la recherche d'indices motivée par nos attentes dans un contexte donné.

Enfin, nous avons fait un retour sur la question de la catégorisation en essayant de comprendre son lien avec le langage. Dans la mesure où la perception est attribution de sens et recherche d'indices, ce processus est influencé par l'ensemble de nos connaissances (processus *top-down*), en particulier par le langage. Réciproquement l'activité de description en langue est liée à la catégorisation. Nous garderons à l'esprit dans l'analyse des verbalisations le fait qu'utiliser un mot pour décrire une perception, c'est toujours faire entrer une perception particulière et personnelle dans une catégorie communicable socialement.

## Alimenter la conversation

Avez-vous remarqué qu'à table les mets que l'on vous sert vous mettent les mots à la bouche ?

J'en ai fait l'observation un jour que je dînais seul.

A la table voisine... il y avait deux convives qui mangeaient des steaks hachés...

Et tout en mangeant,  
ils alimentaient la conversation.

Au début du repas, tandis que l'un parlait,  
l'autre mangeait... et inversement !

L'alternance était respectée.

Et puis... les mets appelants les mots,  
et les mots les mets...

Ils se sont mis à parler et à manger en même temps :

"Ce steak n'est pas assez haché disait l'un"

"Il est trop haché pour mon goût disait l'autre !",

Les mots qui voulaient sortir

Se sont heurtés aux mets qui voulaient entrer...

Ils ont commencé à mâcher leurs mots  
et à articuler leurs mets !

très vite, la conversation a tourné au vinaigre.

A la fin, chacun ayant ravalé ses mots  
et bu ses propres paroles,

il n'y eu plus que des éclats de "voie" digestive  
et des "mots" d'estomac !

Ils ont fini par ventriloquer...

Et c'est à qui aurait le dernier rôle !

Puis l'en d'eux s'est penché vers moi.

Il m'a dit :

"Monsieur, on n'écrit pas la bouche pleine !"

Depuis, je ne cesse de ruminer mes écrits !

Je sais... vous pensez :

"Il a écrit un sketch alimentaire,  
un sketch haché !"

Et alors ?

Il faut bien que tout le monde mange !

Raymond Devos  
A plus d'un titre (1989)

## Chapitre 2. Parler

L'idée de comprendre le lien entre nos perceptions et les mots que nous utilisons pour les décrire est au centre de notre travail. Nous allons donc aborder le problème du sens des mots et des expressions du langage, à travers les phénomènes de compositionnalité, de polysémie et de création de sens ou de mots nouveaux.

Deux paradoxes apparaissent d'emblée avec le langage.

- D'une part l'imbrication entre ses dimensions individuelles et collectives : "la langue" existe mais elle n'est en même temps qu'un ensemble de locutions d'individus particuliers à des moments particuliers. "L'idée générale est que les phénomènes sociaux sont l'effet cumulé des actions individuelles, elles-mêmes guidées par des choix rationnels. (...) L'explication de ces macrophénomènes est à chercher, à chaque fois, dans l'enchaînement de millions de microprocessus. (...) Des chaînes causales, tout comme les épidémies qui propagent les maladies, propagent, infléchissent, stabilisent des idées, des valeurs, des normes à travers des sociétés entières. Elles déterminent ainsi les macrophénomènes de la vie culturelle et sociale" (Sperber, dans Droit & Sperber, 1999, p 104).

- D'autre part, on peut à la fois dire qu'on veut, inventer schtroumpfs ou chmilblic, ET en même temps on ne peut pas dire ce qu'on veut : si on veut être compris, il faut utiliser un code que d'autres peuvent comprendre donc utiliser les mots dans un cadre proche ou généralisable. D'où les critiques faites au langage sur sa capacité à rendre compte de ce qui est personnel, le mot-étiquette étant forcément conventionnel, il ne peut dire que des généralités (Bergson, 1938).

Pourtant, la langue évolue. Même des inventions à court terme, comme le verlan, peuvent être comprises par un grand nombre de personnes. Et dans une certaine mesure, chaque individu peut influencer la langue. La médiatisation actuelle accélère peut-être encore ce phénomène, comme voie de communication à grande vitesse. Ainsi par exemple, c'est Boris Vian qui en 1958 a imposé le terme de "tube" pour désigner un morceau de musique très connu pour remplacer le terme de "saucisson" qu'il trouvait malsonnant (Vian, 1991). Pourtant impossible de dire si nous utiliserions encore ce terme s'il ne l'avait pas fait.

Enfin, toute philosophie du langage se trouve confrontée à deux questions : A quoi nous référons-nous lorsque nous parlons ?, mais aussi : Qu'est-ce qui nous fait parler ? (Eco, 1999).

C'est ainsi que nous allons tout d'abord questionner la fonction référentielle du langage pour aborder ensuite sa fonction communicative. Ceci nous permettra de nous re-questionner sur les liens entre le langage et la pensée avant d'essayer de comprendre comment ces différentes composantes du langage entrent en jeu lors de son acquisition par l'enfant.

### I. Lien entre le langage et le monde



"Au commencement était le Verbe  
et le verbe était avec Dieu  
et le verbe était Dieu.  
Tout fut par lui  
et sans lui rien ne fut".

(Jean, I, 1-3)

Si on se pose la question des liens entre le langage et le monde, un grand nombre de traditions rapportent, comme dans le passage de la Bible cité ci-dessus, que c'est le verbe qui a créé le monde. Cette conception créatrice de la parole ("Le Seigneur dit : que la lumière soit, et la lumière fut") se retrouve en effet chez les Egyptiens, pour qui "Path, seigneur de Memphis, créa le monde en articulant ses pensées." (De la Motte-Haber, 1994) et chez les Hindous. En effet les Vedas, "affirment que chaque réalité de ce monde a été générée par un son. Le nom de Brahma signifie d'ailleurs "parole créatrice" ou "verbe créateur"" (Payeur, 1995). Or, si nous avons été créés à l'image et ressemblance de Dieu, nous avons, nous aussi cette capacité de créer le monde par le langage. On retrouve d'ailleurs cette conception dans la Cabale Chrétienne (Eco, 1994) pour qui l'hébreu, langue divine et sacrée par excellence a un caractère magique qui lui permet non pas de communiquer des choses mais d'agir sur les choses. Il semblerait aussi que ce soit le mode premier de conception de l'enfant : "l'objet n'est connu que lorsqu'il est nommé, le nom est l'essence de l'objet et réside en lui au même titre que sa couleur et que sa forme" (Piaget, cité par Merleau-Ponty, 1945), qui renchérit : "Pour la pensée pré-scientifique, nommer l'objet c'est le faire exister et le modifier : Dieu crée les êtres en parlant d'eux". Mais on peut aussi comprendre cette idée de parole créatrice sans forcément tomber dans une conception magique. De manière plus prosaïque en effet, pour reprendre les idées développées dans le chapitre précédent, si on arrive à réaliser que ce qu'on perçoit est une forme de projection de notre réalité intérieure sur des "variations de densité" du monde, et si on conçoit le langage comme notre outil privilégié de catégorisation consciente, alors ce serait bien au moyen du langage qu'on créerait le monde...

(D'ailleurs, si on considère que la chute du paradis correspond à une rupture avec une vision du monde intérieure et reliée avec Dieu (non différenciée de Dieu) symbolisée par le fait d'ouvrir les yeux sur le monde comme s'il était extérieur à nous (voir qu'on est nu), il suffirait de fermer ces yeux extérieurs, ou plutôt d'ouvrir les yeux intérieurs, ou peut-être encore tout simplement de reconnaître que ce que nous croyons extérieur est un reflet de notre intérieur... il suffirait donc de regarder le monde de l'intérieur pour voir... le paradis ! (Ceci correspondrait d'ailleurs avec la parole du Christ : "le Royaume, c'est le dedans de vous" !). Pour regarder le paradis, c'est-à-dire réaliser qu'on est toujours dedans mais seulement incapables de le voir, il faudrait arriver à concevoir "où" il est puis à le "ressentir", c'est-à-dire arriver à voir notre action de voir ?)

### I.1. Sémantique : recherche du sens

Si on cherche à mettre une ligne de démarcation entre les phénomènes relevant de la sémantique et ceux relevant de la pragmatique, la sémantique se préoccuperait de la signification comme d'une relation directe entre les mots et les choses (partant donc de la fonction de *représentation* du langage).

La pragmatique considérerait la signification plutôt à partir de la fonction *expressive* du langage en étudiant "ce qui est dit" (Récanati, 1999). Nous allons donc tout d'abord considérer comment des conventions définissent le sens des éléments simples et comment des règles du langage peuvent être utilisées pour en déduire le sens de suites d'éléments simples.

### **Signe**

"Un discours oral, ce sont des vibrations de l'air. Un texte écrit, ce sont des marques sur le papier. Ce qui rend ces matérialités identifiables comme des discours ou des textes, c'est qu'elles sont interprétables et qu'elles ont été voulues comme telles par leurs producteurs" (Droit & Sperber, 1999, p 97). On peut ainsi définir la notion générale de signe à l'aide d'une relation de *renvoi* entre deux choses : la chose qui sert de signe, et celle qu'elle signifie. Arnauld et Nicole (1970) décrivent ainsi le double aspect du signe : "Quand on ne regarde un certain objet que comme représentant un autre, l'idée qu'on en a est une idée de signe, et ce premier objet s'appelle signe. C'est ainsi qu'on regarde d'ordinaire les cartes et les tableaux. Ainsi le signe renferme deux idées : l'une de la chose qu'il représente, l'autre de la chose représentée ; et sa nature consiste à exciter la seconde par la première." C'est ainsi par exemple, dans les expérimentations visant à apprendre le langage aux singes, qu'on peut déterminer à quel moment un singe utilise un signe : "le jeton de plastique devient un mot quand les caractéristiques qui lui sont attribuées sont celles de l'objet désigné" (Lestel, 1995). L'idée de relation de signe apparaîtrait par la co-occurrence, ce qui renvoie déjà à l'idée de référence avec une chose du monde : "Un son ne deviendrait jamais pour un enfant le signe verbal d'une chose s'il n'avait été entendu, pour le moins, une fois au moment où cette chose est perçue" (Condillac, 1749).

### **Naturel / conventionnel ?**

En fonction de la relation qu'entretient le signe avec son contenu, on peut le considérer comme naturel ou conventionnel. Les signes conventionnels sont arbitraires alors que les signes naturels semblent véhiculer une information portant sur ce dont ils sont les signes (Ludwig, 1997). Ainsi, la fumée est le signe du feu, et le nombre de cercles que l'on peut observer sur la souche d'un arbre est le signe de son âge. Les Grecs considéraient que leurs mots étaient forgés sur la nature par imitation directe des choses (Eco, 1994), ce qui pouvait leur donner un caractère "naturel". Pourtant, Saussure (1955) souligne comme premier principe l'arbitraire du signe. "Ainsi l'idée de 'soeur' n'est liée par aucun rapport intérieur avec la suite de sons s-ö-r qui lui sert de signifiant." Ce choix du signe pour désigner une chose résulte donc d'une convention. Mais, pour Merleau-Ponty (1945), les conventions sont un mode de relation tardif entre les hommes, elles supposent une communication préalable, et il faut replacer le langage dans ce courant communicatif. "Si nous ne considérons que le sens conceptuel et terminal des mots, il est vrai que la forme verbale - exception faite des désinences - semble arbitraire. Il n'en serait plus ainsi si nous faisons entrer en compte le sens émotionnel du mot, ce que nous avons appelé plus haut son sens gestuel, qui est essentiel par exemple dans la poésie" (p 218). Ainsi, pour Merleau-Ponty, "la prédominance des voyelles dans une langue, des consonnes dans une autre, les systèmes de construction et de syntaxe ne représenteraient pas autant de conventions arbitraires

pour exprimer la même pensée, mais plusieurs manières pour le corps humain de célébrer le monde et finalement de le vivre" (ibid.). Contrairement à la position des Grecs rapportée par Eco, les sons utilisés pour former chaque mot seraient liés à la chose représentée non pas dans le but de l'imiter dans une forme de ressemblance objective (à la manière des onomatopées), mais parce qu'ils en exprimeraient *l'essence émotionnelle*. De là viendrait que le sens "plein" d'une langue n'est jamais traduisible dans une autre. Nous pouvons parler plusieurs langues, mais l'une d'elles reste toujours celle dans laquelle nous vivons. Pour assimiler complètement une langue, il faudrait assumer le monde qu'elle exprime et l'on n'appartient jamais à deux mondes à la fois. Même pour des bilingues, on observe toujours une asymétrie dans la rapidité d'accès au mot concernant le même concept dans chacune des deux langues (Kroll & Stewart, 1994), en rapport avec le fait qu'une des deux langues a été apprise en premier. Il semble de plus que les bilingues ne possèdent pas exactement le même vocabulaire dans les deux langues mais seulement celui correspondant à l'expérience vécue dans chacune d'elle.

### Référence

Les mots ont du sens parce qu'on peut les associer à des aspects de la réalité ou de mondes possibles. Eco, dans sa Leçon inaugurale au collège de France (1994) fait remonter l'histoire de la référence à la nuit des temps puisqu'un processus de ce type est décrit dans le chapitre 11 de la Genèse : "Le Seigneur Dieu modela encore du sol toutes les bêtes sauvages et tous les oiseaux du ciel, et il les amena à l'homme pour voir comment celui-ci les appellerait : chacun devait porter le nom que l'homme lui aurai donné. L'homme donna des noms à tous les bestiaux, aux oiseaux du ciel et à toutes les bêtes sauvages". Pourtant, force est de constater que le référent échappe aux hommes. Une taxinomie n'est nette que si elle est vide de sens. Dès qu'elle renvoie au monde extérieur, malentendus et imprécisions apparaissent (Kibbedi-Varga, 1996). Eco imagine néanmoins, dans Kant et l'ornithorynque (1999, p 40) quatre modèles possibles de la référence vue comme la correspondance faite par un Esprit entre les objets d'un monde imaginaire et des signes (qui peuvent eux-aussi faire partie du monde).

#### 1- Un objet <-> un signe

Le premier et le plus simple des modèles possibles est l'idée d'une représentation spéculaire du monde avec autant d'objets que de signes, la référence consistant en une bijection mot <=> chose. Dans ce cas, l'esprit n'interprète pas le monde, il le représente comme un miroir. Pourtant, si le monde contient trois objets et trois signes, il y a déjà six langues possibles correspondant à 6 points de vue différents. Ceci rejoint la remarque de Pierre Cadio (1998) : "le réalisme en matière de référence est contradictoire avec ce que c'est qu'une langue. (...) Il y aurait alors une explosion combinatoire des signes." On peut aussi se poser la question de l'intérêt d'avoir ainsi chaque chose en double : un signe par objet n'apporte rien.

#### 2- Moins de signes que d'objets du monde

Apparaît alors l'homonymie : beaucoup de choses différentes ont le même nom. Par exemple, le mot "homme" sert à nommer la multiplicité des individus... L'esprit pourrait alors essayer de décrire le monde en procédant morceau par morceau, comme s'il regardait par le trou d'une serrure.

### 3- Beaucoup plus de signes que d'objets

Ceci donnerait la possibilité de rendre compte de l'objet mais aussi de l'heure et du jour auquel le monde est représenté, de l'état interne de l'esprit à ce moment précis, des intentions et des fins... On constate alors un excès de pensée par rapport à la simplicité du monde, menant à une abondance de synonymes. On remarque néanmoins que les synonymes ne recouvrent jamais strictement la même réalité. Ils ne sont pensés comme "synonymes" qu'après catégorisation dans la même "boîte" de plusieurs points de vues d'une même chose.

### 4- Infinité de part et d'autre...

Cette dernière hypothèse, permettant de comprendre à la fois les phénomènes d'homonymie et de synonymie se rapprocherait de la langue telle que nous la pratiquons. On peut en effet considérer qu'il y a une infinité d'objets du monde et une grande quantité de mots, avec la possibilité d'en créer de nouveaux à l'infini.

### **Compositionnalité**

Les langues parlées partagent avec les langues logiques la propriété de compositionnalité : c'est-à-dire que le sens d'une phrase n'est pas appris *a priori* mais peut se calculer à partir du sens des différents constituants ou mots et de la construction syntaxique de cette phrase. Cette propriété (ajoutée à la propriété de récursivité des règles de composition) permet de construire une infinité de phrases et de calculer leur sens à partir d'un nombre limité de mots et de règles de combinaison de ces mots. Pendant longtemps on a cru que le langage était un code et que la connaissance des règles de compositionnalité était suffisante pour attribuer à la phrase le sens du message qu'on voulait communiquer. La communication se résumant à des opérations de codage / décodage, les ordinateurs comprendraient rapidement les langues parlées, ce qui faciliterait la communication avec eux et rendrait possible des traductions automatiques par exemple. Mais il s'est avéré que les ordinateurs, mêmes munis des règles de la langue française et du vocabulaire nécessaire étaient incapables de déterminer le sens d'une phrase du type :

*"Le professeur envoya le cancre chez le censeur : il lançait des boulettes"*

Il est en effet impossible de déterminer qui, du professeur, du cancre ou du proviseur est le sujet de la deuxième partie de cette phrase uniquement à partir de règles conventionnelles de syntaxe. Cette tâche nécessite une connaissance minimale de ce qu'il est possible de se passer dans un lycée français au XXème siècle. En fait, le référent est déterminé en terme de *saillance* : c'est la personne la plus plausible dans un contexte donné, en fonction des croyances qu'on a : par exemple, qu'un proviseur ne lance habituellement pas de boulettes, du moins dans l'exercice de ses fonctions. Il est intéressant de remarquer que si on fait varier les croyances, on fait varier la plausibilité des différentes hypothèses ! On peut noter aussi que cette explication en terme de saillance renvoie en fait à la mémoire et à la perception. Est saillante la situation la plus fréquemment rencontrée (expérimentée ou rapportée) dans un contexte semblable.

De même, si on considère les deux phrases suivantes qui ne diffèrent que par l'inversion des deux locutions "travailleur" et "peu discipliné":

*"Jean est peu discipliné mais travailleur"*

*"Jean est travailleur mais peu discipliné"*

Ces deux expressions ont un sens tout à fait différent bien qu'elles utilisent les mêmes significations. Or, si des règles de compositionnalité suffisaient pour déduire le sens d'une phrase, on pourrait s'attendre qu'il y ait commutativité, comme dans l'addition  $1+3=3+1$ . Pourtant, la première phrase aura le sens d'une louange, la seconde d'un reproche, mais cela peut être inversé en fonction du locuteur et du contexte (Basso, 2000). De plus, changer un seul mot d'une phrase peut amener à modifier le sens des autres mots, ainsi que le montrent les deux phrases suivantes (Victorri, 1999) :

*"un livre sur la table de chevet"*

*"un livre sur la table de logarithmes"*

Le mot "table" présentant respectivement dans ces deux phrases les sens de meuble et de liste de chiffres, "sur" indiquant respectivement une relation spatiale ou notionnelle.

De plus, le langage est rarement utilisé à des fins purement objectives, si on considère l'étendue des discours possibles : langage allusif, humour, sous-entendus... dont le sketch de Devos (1989) mis en exergue offre un panaché. A l'extrême, une même phrase peut prendre plusieurs sens distincts en fonction du contexte considéré. C'est notamment le cas, dans le sketch de Devos par exemple, de *"Depuis, je ne cesse de ruminer mes écrits"*, pour lequel le contexte permet d'imaginer cette action au sens "propre" de manger, alors que l'acceptation classique de cette expression est habituellement comprise au "sens figuré" de retravailler, repenser (ou re-mâcher ?) sans cesse. C'est aussi le cas de l'exemple connu de *"la petite brise la glace"* (Récanati, 1992) dans laquelle, en fonction du contexte, le mot "brise" prend le sens de "petit vent" ou est une forme du verbe briser. "La glace" prend alors aussi deux sens différents mais qui sont liés au premier choix. Deux paraphrases possibles seraient alors : "le petit vent léger refroidit cette femme" (ou lui glace le sang), ou "une petite fille est en train de briser une vitre", (ou un miroir, ou la surface glacée d'un lac, ou déclenche un extincteur...). Mais aussi : "une petite fille est en train de mettre à l'aise cette assemblée"...

Alors que dans le cadre de la compositionnalité, il n'existe aucune différence qualitative entre un tout et ses parties, le sens de la phrase émerge ici sans être directement calculable à partir des sens de ses parties (Basso, 2000). Ceci rejoint l'idée gestaltiste que le tout est plus que la somme de ses parties. On comprend alors la remarque de Hofstadter (1988) "la question de l'émergence du signifié à partir de composants sans signification, combinés suivant certains schémas, devient alors déroutante."

## I.2. Polysémie

Les quelques exemples précédents permettent d'imaginer l'ampleur du phénomène de la polysémie. C'est en fait un phénomène massif en langue : plus un mot est courant, plus il est utilisé tôt dans l'enfance, plus il est polysémique. Pourtant, dans des conditions naturelles d'interlocution, la compréhension est immédiate. Il semble donc qu'un énoncé ne constitue pas un problème à résoudre pour l'interlocuteur, mais qu'il consiste au contraire en une série d'instructions. Ces instructions le conduisent à construire de manière directe le sens de cet énoncé, sans qu'il ait à mettre en oeuvre des stratégies de parcours dans un espace de recherche, de retour arrière, et autres ingrédients typiques de la résolution de problème (Victorri & Fuchs, 1996). De plus, des langues nouvelles construites de

toutes pièces comme l'espéranto deviennent rapidement polysémiques. Compte tenu de l'évolution des langues, si la polysémie était une entrave à notre capacité de compréhension, ce phénomène aurait vraisemblablement disparu. Il s'agit donc de comprendre que la polysémie n'est pas un phénomène secondaire, mais qu'elle doit vraisemblablement avoir une utilité.

### **Emiettement des sens**

La conception classique de la polysémie, qui correspond à la construction et l'usage des dictionnaires, consiste à considérer que chaque mot peut avoir plusieurs sens distincts, voir indépendants, sans qu'il y ait de lien entre ces différents sens, et donc de les lister. Par exemple, le mot "souris" aurait trois sens distincts : 1- Petit rongeur du genre rat ; 2- Petit muscle qui tient au manche du gigot ; 3- Petit élément d'un ordinateur servant à déplacer un curseur sur l'écran (seules les deux premières définitions sont actuellement dans le Larousse). Il n'empêche qu'on peut se demander si ces trois sens sont bien distincts ou si il n'y aurait pas un point commun entre ces trois acceptions (une ressemblance de forme ?), qui permettrait par exemple à "un homme politique au plus haut sommet de l'état", peu familier du vocabulaire informatique, d'avoir tout de même retenu l'essentiel en demandant face à un ordinateur : "où est passé le mulot ?"

Pour certains verbes, le nombre de valeurs recensées par les "gros" dictionnaires peut ainsi atteindre la centaine. "Les valeurs et emplois d'un verbe sont régis par des facteurs profondément hétérogènes, d'ordre sémantique, syntaxique, lexical mais aussi aspectuo-temporel. Ainsi, par exemple, pour ce qui est de la syntaxe, un même verbe peut être transitif, intransitif, réflexif : *le temps passe / Paul passe devant la maison / passe moi le sel / il a passé la frontière / les vacances se passent bien (...)* mais aussi : *passer quelqu'un par les armes / par la fenêtre / passer sur un point / passer un détail sous silence / passer pour un imbécile / passer en justice, etc.*" (Saunier et al., 1994). Confronté à cette diversité de valeurs, on peut choisir de ramener la description d'un "lexème verbal" à un inventaire (plus ou moins raisonné) des emplois. "Les approches consistant à spécifier les rôles thématiques du verbe en relation avec une ou plusieurs configurations syntaxiques sont en fait des entreprises de normalisation sémantique du verbe (ce qui entraîne, souvent, une multiplication des valeurs idiomatiques 'figées')" (ibid.).

### **Noyau de sens**

A l'inverse, d'autres approches s'attachent avant tout à sauver l'unité du lexème par la mise en évidence d'un noyau sémantique minimal compatible avec différentes réalisations contextuelles, ou par la mise en évidence d'une valeur prototypique et des chemins qui la relient aux autres valeurs. Ainsi, Cadio (1997) montre qu'il est possible de rendre compte simplement de la totalité des emplois d'un mot, à condition de poser en des termes nouveaux la question du rapport existant entre les noms et les objets, et en premier lieu à condition de s'interroger sur le(s) rapport(s) que l'on a avec eux. "Il nous semble (en effet) indispensable de comprendre qu'il n'y a pas, pour la langue, de monde sans l'homme et que les mots assument d'emblée un monde *avec* l'homme" (ibid.). Apparaît ainsi la Sémantique Cognitive, pour laquelle il n'y a pas de signification en dehors de la compréhension humaine, en réaction contre la Sémantique Formelle, qui considère que "les signes et les phrases du

langage humain peuvent renvoyer directement à la réalité indépendamment de l'expérience humaine" (Nyckees, 1997). Le langage n'est plus considéré comme une compétence spécifique relativement autonome. Il y a donc des liens beaucoup plus forts avec des modèles issus de la psychologie.

Ainsi, Cadio & Nemo (1997) vont concevoir deux types de propriétés permettant de décrire un objet : ses propriétés intrinsèques (P.I.) qui lui seraient propres "indépendamment du regard humain" (est-ce possible ?) et ses propriétés extrinsèques (P.E.) dues au type de rapport que l'on entretient avec cet objet et qui rejoint sans doute la notion d'affordance de Gibson (1950). Le sens d'un nom relèverait alors de ses P.E. Ainsi par exemple, "le mot '*lit*' dans une expression comme '*faire le lit du fascisme*' signifie quelque chose comme 'permettre au fascisme de s'étendre'". Il y a entre l'objet "lit" et le fait de s'étendre un rapport étroit, le fait de s'étendre étant une forme spécifique du contact avec l'objet "lit" (et donc une P.E.) et non une propriété intrinsèque de l'objet "lit" (et donc une P.I.). Cette P.E. permettrait aussi de comprendre que *le lit de la rivière* peut désigner l'ensemble des zones où elle peut s'étendre, ou que quand on décrit quelqu'un *sur son lit de mort*, "il ne s'agit pas d'un lit particulier (avec des P.I. spécifiques) mais bien le moment où une personne est étendue (...) avant de mourir." (ibid.)

On pourrait comprendre ainsi l'idée que si un énoncé évoque quelque chose, cette évocation est faite en puisant dans notre expérience phénoménologique. "Le mot est considéré alors comme un outil qui convoque cette expérience phénoménologique" (Victorri, 1999). La signification lexicale n'est alors plus conçue en terme de description (P.I.). Le mot renvoie directement aux types de relations que le locuteur entretient avec les référents variés que ce mot lui permet de construire, et que ces référents entretiennent avec leur environnement (Cadio & Nemo, 1997). Pour reprendre l'exemple de nos trois souris, indépendamment de leur constitution physique spécifiques, elles auraient une taille similaire qui permettrait par exemple de les prendre dans le creux de la main.

#### **Schéma de sens et construction dynamique du sens**

Entre "émiettement" des sens à la manière des dictionnaires et "invariance abstraite" qui serait une forme de résumé de tous les emplois possibles du même mot, il semble exister "une troisième voie où la mise en évidence d'un pôle d'invariance ne se fait pas au détriment de la diversité des emplois : le pôle d'invariance étant avant tout une façon de penser la variation" (Saunier et al., 1994). Ceci revient à considérer que l'identité du mot est inséparable de son interaction avec son "co-texte", défini comme l'ensemble du contexte textuel (phrases) dans lequel il est présenté. "L'identité du verbe est alors définie par une 'forme schématique' (assimilable à un scénario abstrait) inscrivant le verbe dans un double rapport au co-texte : d'une part, le verbe informe le co-texte, d'autre part, certains éléments du co-texte donnent corps à cette forme abstraite." (ibid.) Ainsi par exemple, la forme schématique de "passer" serait : "*passer* dit qu'une discontinuité liée à une continuité première est reformulée comme continuité" et permettrait de comprendre l'exemple *Paul passe la frontière* comme : "dans la trajectoire de Paul, la frontière, qui est *apriori* une discontinuité, n'est pas une discontinuité" (ibid.) De même, la forme schématique de "sur" serait : "*sur* convoque E1 et E2 tel que E2 soit un fermé pouvant servir de localisation à E1" (Victorri, 1999). *Sur*, évoquerait ainsi une relation de localisation de E1 par la surface de E2 en relation avec E1. On pourrait ainsi comprendre que "sur le mur" ne désigne pas la

même partie du mur dans les deux phrases suivantes : *une poule sur un mur* et *une affiche sur un mur*, mais respectivement "au dessus" et "contre". Mais cette conception demande une nouvelle fois de faire appel à la notion de saillance perceptive pour trouver la surface en relation avec E1 dans chaque cas. Alors que la plupart des unités lexicales et grammaticales sont porteuses d'une grande diversité de sens suivant les contextes, on pourra alors comprendre comment elles peuvent jouer ce rôle d'instructions univoques (non ambiguës) dans un énoncé donné et des conditions d'énonciation précises "si l'on conçoit le comportement de chaque unité en termes de processus de convocation - évocation, qui permet de décrire l'interaction de cette unité avec son co-texte et son contexte comme un processus dynamique au cours duquel son sens se précise en même temps que se construit le sens global de l'énoncé" (Victorri & Fuchs, 1996). L'exemple déjà cité de *la petite brise la glace* montre bien que le sens de chacun des mots d'une phrase dépend du sens des autres. Le mot est alors conçu comme un outil de construction d'une "scène verbale", qui serait une construction intersubjective et qui peut éventuellement se référer à des objets en situation dans le monde. Ceci rejoindrait l'idée de Merleau-Ponty (1945) que "le mot disparaît au fur et à mesure qu'il construit son sens".

#### **Sens nouveaux et évolution du sens**

Si on considère de plus l'évolution du sens des mots et l'apparition de mots nouveaux, l'affirmation de Merleau-Ponty (1945) qu'"on ne peut nous parler qu'un langage que nous comprenons déjà" semble contraignante : elle semble réduire la possibilité d'apprendre quoi que ce soit de nouveau à travers le langage. Pourtant, il existe toujours la possibilité d'émergence de propriétés ou d'événements échappant à toute règle prédictive. "Cette incommensurabilité du possible apparaît dans l'usage du langage quotidien sous forme de métonymie, d'homophonie, dans l'allusion, dans le jeu de mots notamment. Mais aussi dans la possibilité d'évolution d'une langue où de nouveaux mots sont créés, où des mots connus prennent à un certain moment d'autres significations et induisent d'autres sens, sans pour autant qu'une règle ait déterminé à l'avance ces nouveaux usages" (Basso, 2000). Il y a donc bien une sorte de tension entre le sens comme convention et la possibilité de son évolution. En effet, s'il existe nécessairement une sorte de norme ou de convention partagée réglant l'usage des mots, norme sans laquelle on voit mal comment les usagers s'accorderaient sur leur emploi et se comprendraient entre eux, "les signes appris sont indissociables d'une certaine mise en forme de l'expérience et de la pensée. Ils sont un principe de structuration de l'expérience bien plus que des 'étiquettes' à mémoriser" (Nyckees, 1997). Nyckees défend alors l'idée d'une évolution du sens au cours de l'histoire comme étant une sorte d'enregistrement dans la langue des mutations d'expériences collectives. Ainsi par exemple, le mot *arriver* est passé d'un sens technique utilisé dans la marine pour désigner les manoeuvres pour approcher de la rive à un sens beaucoup plus général mais qui conserve l'idée de se rapprocher d'un but. "Systématisant cette idée maîtresse, je poserai comme hypothèse de travail que toutes les évolutions sémantiques résultent en dernière analyse de l'écart structurel - non pas constant, mais toujours possible - entre la 'signification du locuteur' et la 'signification du récepteur' (ibid.). Nyckees définit alors l'idée de *contrainte de vérissemblance* comme suit : "dans les occurrences où se produit le changement de sens, les énoncés n'enfreignent pas les conditions de vérité, si l'on entend par là que rien ne permet de décider, dans les situations où



ils sont prononcés, si les énoncés recourent déjà à la signification nouvelle du mot concerné ou s'ils continuent à recourir à son ancienne signification." Par exemple, dans le cas d'"arriver", on peut imaginer une situation dans laquelle deux personnes sont dans un bateau et l'un dit à l'autre qu'ils vont bientôt arriver. Il est possible que l'un des deux comprenne cela dans le sens de "faire la manoeuvre d'accostage" alors que l'autre est déjà dans l'acception d'"annonce de la fin du voyage", les deux sens étant compatibles dans cette situation.

L'usage que chacun de nous fait des mots de notre langue correspond donc à la fois à un usage conventionnel et à un usage spécifique qui enrichit son sens de possibilités nouvelles mais à chaque fois accessible à l'autre dans le contexte. Ceci permet de comprendre la remarque de Merleau-Ponty (1945) comme quoi "à vrai dire, le sens d'un ouvrage littéraire est moins fait par le sens commun des mots qu'il ne contribue à le modifier."

Merleau-Ponty (1945) résout l'apparente contradiction entre le fait qu'on ne parle qu'avec des mots déjà connus et que pourtant des sens nouveaux peuvent apparaître en considérant que "chaque mot d'un texte difficile éveille en nous des pensées qui nous appartenaient auparavant, mais ces significations se nouent parfois en une pensée nouvelle qui les remanie toutes. (...) L'intention significative nouvelle ne se connaît elle-même qu'en se recouvrant de significations déjà disponibles, résultats d'actes d'expressions antérieurs" et en concluant que "tout langage en somme, s'enseigne lui-même et importe son sens dans l'esprit de l'auditeur."

### I.3. Résumé

Nous avons donc tout d'abord essayé de comprendre comment quelque chose pouvait devenir le signe d'une autre chose (de manière naturelle ou conventionnelle). Puis nous avons envisagé la construction du sens d'une phrase selon des règles de compositionnalité, pour en constater les limites. Enfin nous avons présenté différentes explications du phénomène de polysémie : comme émiettement des sens à la manière d'un dictionnaire ou au contraire en essayant de reconnaître un noyau de sens ou une forme schématique qui permettrait de rassembler ou de construire tous les sens possibles d'une même locution.

Ce voyage dans les explications de la construction ou de l'émergence du sens des mots et des phrases nous a paru nécessaire dans la mesure où notre projet initial était de trouver des "étiquettes" verbales aux dimensions perceptives du timbre. Ceci sous-entendait en effet une conception référentielle du langage. De plus, à partir du moment où on a voulu rassembler des verbalisations de même sens pour construire une arborescence, nous avons été confrontés à ce "calcul" du sens des locutions des participants et donc aux problèmes de synonymie, polysémie et compositionnalité décrits ci-dessus. En effet, en première instance, on a tendance à regrouper ensemble les locutions contenant des mots identiques... ce qui conduit parfois à des surprises. Enfin, ces considérations sur le langage vont nous permettre de comprendre comment un vocabulaire vague et polysémique peut adapter son sens à une situation particulière et précise.

## II. Langage comme moyen de communication

Si le langage permet de représenter la réalité, de véhiculer des informations sur le monde, on peut supposer que c'est essentiellement afin de permettre aux hommes de communiquer leurs pensées. "Il s'agit alors d'envisager le langage dans son utilisation et non simplement dans son essence : c'est étudier les différents usages que les hommes peuvent faire des mots" (Ludwig, 1997).

Nous avons déjà introduit la pragmatique comme s'intéressant à "ce qu'on veut dire" (Récanati, 1979) au delà même des mots que l'on utilise. Ainsi, par exemple, si Pierre et Paul se trouvent dans une grande pièce, dont la fenêtre est ouverte, que Pierre est près de la fenêtre alors que Paul est à l'autre bout de la pièce, le fait que Paul s'exclame en regardant la fenêtre : *Il fait froid* aura comme signification non ambiguë "peux-tu fermer la fenêtre". La signification conventionnelle des mots ne suffit donc pas toujours à déterminer ce qu'un locuteur veut dire en les employant. Ce n'est donc plus la phrase, mais l'énonciation de la phrase, ancrée dans un contexte qu'il faut considérer comme l'unité fondamentale douée de sens : la signification est associée à l'acte qui consiste à énoncer la phrase. En d'autres termes, si on peut dire des mots qu'ils *veulent dire* quelque chose, on peut aussi le dire des locuteurs et ainsi parler de la *signification du locuteur* à côté de la signification conventionnelle des expressions linguistiques (Grice, 1957).

### II.1. Communiquer

Dans la plupart des espèces animales, les individus communiquent entre eux, ne serait-ce qu'à-propos de la question cruciale de la reproduction. Pourtant la communication prend d'autres dimensions chez l'homme en n'étant plus directement liée à la notion de survie : de l'individu, du groupe ou de l'espèce. On peut alors se poser la question de la nécessité particulière de la communication chez l'homme.

#### **Pourquoi ?**

Le fait que les êtres humains cherchent inlassablement à fournir des informations à leurs congénères peut apparaître comme un paradoxe (Dessales 2000). "Le comportement langagier semble faire exception à la théorie darwinienne, qui prévoit que les organismes se préoccupent avant tout de leur propre survie." Dessales remonte alors aux origines du langage pour en venir à l'idée que l'émergence de notre manière de communiquer est liée au mode d'organisation particulier des groupes humains. "Les premières formes de langage seraient apparues, chez les hominidés, comme un moyen pour les individus de se choisir afin de former des coalitions. Ainsi, loin de résulter d'une tendance évolutive générale, l'apparition du langage serait une conséquence de l'organisation sociale singulière de notre espèce" (ibid.).

#### **Comment ?**

Sperber (1986, 1999) décrit comment une idée peut être communiquée d'un individu à un autre de la manière suivante : "Les idées (au sens psychologique) qui se forment dans la tête d'un individu n'en sortent jamais : elles naissent, vivent et meurent à l'intérieur d'une boîte crânienne. Ce qui sort, ce sont, par exemple, les bruits d'une parole, laquelle parole peut 'exprimer' une idée. Si la communication verbale (orale ou écrite) ne déplace pas littéralement les idées d'un cerveau à l'autre, elle donne naissance à une version de l'idée de celui qui parle dans la tête de celui qui écoute." On comprend dans cette perspective comment ce processus a été modélisé par un codage des idées en langue puis une interprétation de ce code par l'auditeur.

### **Une nouvelle illusion ?**

On peut alors se demander avec Merleau-Ponty (1945) si l'expérience de la communication ne serait pas une illusion. En effet, à première vue, on peut croire que la parole entendue ne peut rien apporter à celui qui l'écoute puisque c'est lui qui donne leur sens aux mots, aux phrases, et la combinaison des mots et des phrases. Tout ce qui est entendu par l'auditeur ne peut pas lui être étranger puisqu'il ne pourrait le comprendre si cela ne rencontrait pas chez lui le pouvoir de le réaliser spontanément. "Ici comme partout, il apparaît d'abord vrai que la conscience ne peut trouver dans son expérience que ce qu'elle y a mis elle-même. (...) Une conscience construit - pour X - cette machine de langage qui donnera à une autre conscience l'occasion d'effectuer les mêmes pensées, mais rien ne passe réellement de l'une à l'autre." (Merleau-Ponty, 1945). Pourtant, il est indéniable qu'on apprend des choses nouvelles en permanence. Elles ne sont donc construites qu'avec des idées que nous possédons déjà, par une sorte de processus de réorganisation : mettre côte à côte des concepts anciens peut faire émerger une nouvelle idée, inconnue jusqu'alors. Ce qui serait communiqué serait donc une nouvelle façon de lier des idées communes aux deux interlocuteurs.

## II.2. Parler pour l'autre

Il y a beaucoup de phénomènes de langage qu'on ne peut pas comprendre sans considérer leur but de "communication", c'est-à-dire leur aspect collectif (Steels, 1998). Pour parler, il faut au moins être deux. Cette idée est reprise par Lestel (1995) comme critique de l'idée d'essayer d'apprendre à parler à des singes : "Parler une langue ne consiste pas seulement à en connaître la grammaire et le vocabulaire, mais également à en maîtriser les usages. L'utilisation d'une langue, autrement dit, n'a de sens qu'à l'intérieur d'une communauté." Les singes à qui on apprend le langage humain ne font pas pour autant partie de notre communauté, même si leurs capacités physiques ressemblent suffisamment aux nôtres pour que nous partagions certaines "affordances" des objets. Ils ne peuvent utiliser ce langage que pour communiquer avec des humains (ou des singes sortis de leurs communautés propres, comme eux-mêmes) car la communication entre singes se passe de notre langage humain. On peut aussi s'interroger sur le fait de communiquer au moyen d'un langage construit sur une autre réalité que la leur. Quel sens peut avoir le mot "voiture" ou "chaussures" pour un singe qui ne peut ni conduire la première, ni enfermer ses "pieds" dans les secondes ? (mais un singe as-t-il des pieds, lui qu'on appelle aussi "quatre-mains" ?).

### **Parler pour soi / pour l'autre**

Quand on parle, en général c'est pour quelqu'un d'autre. Des expériences ont d'ailleurs montré qu'on ne formulait pas les choses de la même façon pour soi et pour l'autre. Ainsi par exemple, s'il s'agit de construire des descriptions de manière à pouvoir reconstituer une série de cartes définies par des combinaisons de sept formes et cinq couleurs différentes (Loewenthal, 1967), on observe que le langage varie avec la volonté de communication, selon que la description est destinée à soi-même directement après, soi-même dans quelques semaines ou quelqu'un d'autre. Par exemple, il n'y a développement d'un code explicite que quand quelqu'un d'autre doit l'utiliser (ou soi-même beaucoup plus tard), sinon les participants utilisent une sorte de "parole intérieure" ou code personnel. Dans ce cas en effet, ils n'ont pas besoin d'expliquer le sens de leurs notations. On peut donc considérer différentes façons de penser à propos des mêmes choses. Le langage comme communication s'inscrit donc résolument dans l'intersubjectivité, ce qui rejoint l'idée que les significations sont le produit d'une culture (Nyckees, 1997).

### **Modélisation**

Apparaissent depuis peu des modélisations informatiques ou robotiques des interactions entre locuteurs qui mènent à la création et à l'utilisation d'un langage commun (Steels, 1998; Kaplan, 1999; Popescu-Belis, 1999; Steels & Kaplan, 1999). En effet, "l'étude des dynamiques collectives conduisant à l'apparition d'un lexique dans une population d'agents artificiels peut fournir des hypothèses pertinentes pour comprendre les phénomènes en jeu dans l'évolution des langues naturelles et dans l'origine du langage humain" (Kaplan, 2000). Ceci s'écarte donc beaucoup du courant connexionniste où on s'occupe de modéliser un cerveau seul (on donne des exemples et on étudie comment l'abstraction peut arriver). Le langage est considéré ici comme un système complexe et adaptatif : beaucoup d'éléments sont en interaction locale non linéaire avec le monde, il est question d'auto-organisation, le comportement des éléments peut changer, le sens change tout le temps, de nouveaux concepts apparaissent. La question sous-jacente est : "Comment les mots de ce lexique peuvent-ils correspondre à des sens partagés alors que les différents agents ne perçoivent ni ne catégorisent le monde de la même façon ?" (ibid.). La méthode utilisée consiste à modéliser les interactions entre agents selon la théorie des jeux, à partir d'une succession de modèles de complexité croissante. Dans les plus simples, les agents artificiels se réduisent essentiellement à une mémoire associative et les objets de l'environnement à un ensemble de symboles. Dans les plus complexes, les agents sont des robots dotés de capacités visuelles qui interagissent à propos de leur perception du monde réel. A chaque étape, les dynamiques collectives qui permettent l'auto-organisation du lexique sont identifiées. "Nous montrons comment ce lexique peut être transmis de génération en génération, comment il se modifie, comment il s'affine et devient toujours plus adapté pour décrire l'environnement réel ou virtuel auquel les agents sont confrontés. Nous montrons comment la structure de ce lexique se régularise et se simplifie pour devenir plus facile à apprendre, plus facile à transmettre." (ibid.)

### II.3. Se faire comprendre

Maintenant que la communication est considérée comme une entreprise collective, il faut reconnaître qu'agir à plusieurs suppose d'accepter une forme de coordination entre partenaires. Dans une telle entreprise, il faut donc d'une part accepter certaines règles, et d'autre part être capable de tenir compte des pensées et des intentions de ses partenaires, afin de pouvoir anticiper sur leurs actions. En effet, ce n'est pas tant le sens des mots qui varie d'un contexte à l'autre, mais plutôt les implications du fait de les prononcer. Par exemple, dire "Quel honnête homme !" lors de l'enterrement d'un savant réputé pour son intégrité, ou lors du procès d'un escroc notoire ne signifie manifestement pas la même chose (Ludwig, 1997). Grice (1979) a été le premier à définir de manière explicite les règles observées dans la conversation. Nous présenterons aussi le principe de l'interprétation pragmatique tel qu'il est décrit par François Récanati (1999) avant d'en voir les liens avec l'idée de théorie de l'esprit.

#### **Maximes de Grice**

Il semble en effet assez évident que lorsqu'on discute avec quelqu'un, il faut respecter certaines règles afin que la conversation ne débouche pas sur un échec complet. Selon Grice (1979), partir du principe qu'il existe de telles règles permet de comprendre comment nous interprétons, en général, les actes de communication, afin de saisir non seulement le contenu explicite, mais également les messages seconds qu'ils véhiculent. Grice propose ainsi de formuler un principe général, appelé "principe de coopération", qu'on s'attendra à voir respecter par tous les participants. "En supposant qu'un principe général de ce genre est acceptable, peut-être peut-on en distinguer quatre catégories (...) que j'appellerai : Quantité, Qualité, Relation et Modalité. La catégorie de QUANTITE concerne la quantité d'information qui doit être fournie, et on peut y attacher les règles suivantes : 1- Que votre contribution contienne autant d'informations qu'il est requis. (...) 2- Que votre contribution ne contienne pas plus d'information qu'il n'est requis (...). A la catégorie de QUALITE on peut rattacher la règle primordiale : Que votre contribution soit véridique et deux règles spécifiques : 1- N'affirmez pas ce que vous croyez être faux. 2- N'affirmez pas ce pour quoi vous manquez de preuves. A la catégorie RELATION, je ne rattache qu'une seule règle : Soyez pertinents. (...) Enfin, à la catégorie de MODALITE, qui ne concerne pas, contrairement aux précédentes, ce qui est dit, mais plutôt comment on doit dire ce que l'on dit, je rattache la règle essentielle : Soyez clairs, ainsi que les maximes : 1- Evitez de vous exprimer avec obscurité. 2- Evitez d'être ambigus. 3- Soyez brefs. 4- Soyez méthodiques."

Grice a énoncé ces règles en supposant que le but recherché était une efficacité maximale de l'échange d'information. Cette définition est bien sûr étroite, et il faudrait généraliser ce schéma de règles de façon à pouvoir tenir compte de buts aussi généraux que le désir d'influencer ou de mener les autres. Il ne faudrait pas pour autant penser que Grice se fait une image trop figée du fonctionnement de la conversation. Les maximes conversationnelles sont aussi là pour être intentionnellement transgressées, ce qui est le cas par exemple de l'interprétation ironique de "Quel honnête homme !" proposée ci-dessus. "C'est alors la connaissance de cette violation par l'interlocuteur qui lui permet de

recouvrer le contenu signifié" (Ludwig, 1997). Il faut alors reconnaître que cette "connaissance" apparaît ici aussi comme saillante dans le contexte correspondant.

### **L'interprétation pragmatique**

Cette description de la communication est en un sens proche du sens commun. "En tant que locuteurs, notre intention est que nos auditeurs reconnaissent notre intention de les informer d'un état de choses. En tant qu'auditeur, nous essayons de reconnaître ce dont le locuteur a l'intention de nous informer" (Sperber & Wilson, 1987). Pour l'auditeur, il s'agirait donc d'effectuer une interprétation "pragmatique" de l'énoncé prononcé par le locuteur. Cette interprétation reposerait sur la "capacité bien connue des humains à attribuer des intentions à leurs semblables" (ibid.). En effet, nous sommes naturellement équipés pour interpréter les actions humaines, c'est-à-dire leur donner du sens. Il s'agit d'une théorie naïve, implicite, qui considère que les gens agissent en fonction de raisons ou d'intentions : mélange complexe d'états mentaux, croyances, désirs. Quand on voit un humain faire quelque chose, on cherche "pourquoi" (la raison de cette action), c'est-à-dire à identifier ses intentions, désirs, croyances...

L'interprétation pragmatique d'un énoncé aurait trois principales caractéristiques (Récanati, 1999) : celles d'être holistique, rationnelle et empathique. Elle est holistique dans la mesure où elle n'est pas faite de manière déductible et correspond toujours à la meilleure hypothèse possible dans le contexte global, et peut donc toujours être remise en cause. N'importe quelle donnée nouvelle peut se révéler pertinente : on ne peut donc pas circonscrire le contexte pertinent qui reste donc global.

On peut illustrer ceci avec un exemple analysé par Sperber et Wilson (1987) :

Pierre : *Veux-tu du café ?* Marie : *Ca me maintiendrait éveillée.*

Pierre, pour pouvoir interpréter la réponse de Marie doit considérer qu'elle est pertinente malgré le fait que cela n'apparaisse pas de façon immédiate. Sperber et Wilson en concluent que Pierre doit supposer que l'intention de Marie "est qu'il infère de sa réponse qu'elle ne veut pas rester éveillée, et que donc elle ne veut pas de café." On peut néanmoins imaginer que Pierre et Marie sont chargés de surveiller une situation pendant toute la nuit. Dans ce contexte il serait donc évident que Marie veut rester éveillée et qu'elle veut donc du café.

L'interprétation est rationnelle dans le sens où on présume toujours la rationalité du locuteur. Cette rationalité rejoint en fait la troisième caractéristique qui est l'empathie. En effet, on considère un agent comme rationnel s'il est "comme nous" et c'est ainsi qu'on donne un sens aux actions d'agents non humains, comme des machines, des animaux en leur prêtant ces qualités. Le caractère empathique consiste à pouvoir imaginer se mettre à la place du locuteur. On interprète donc ses actions à la lumière de ce que nous-même ferions. Cette attitude est peut-être à la racine de l'anthropomorphisme.

### **Théorie de l'esprit**

Ceci nous permet de faire le lien avec l'idée de "Théorie de l'esprit" qui nous semble reposer sur un principe d'empathie. On rejoindrait alors la réponse de Bergson (1919) à la question : Comment pouvez-vous savoir que je suis un être conscient ? "Entre vous et moi il y a une ressemblance extérieure évidente ; et de cette ressemblance extérieure vous concluez, par analogie, à une similitude

interne." C'est ainsi que notre esprit explique le comportement des autres personnes par ses propres croyances et ses propres désirs parce que le comportement des autres est en fait causé par leurs croyances et leurs désirs (Pinker, 2000). (Ainsi, quand on prête à l'autre de la malveillance, ce n'est possible que si on peut soi-même s'imaginer malveillant dans la même situation !) Ainsi, le langage ne serait possible que pour quelqu'un possédant une "théorie de l'esprit", et si l'autisme est bien une défaillance de la capacités à élaborer une "théorie de l'esprit" (Baron-Cohen, 1995), on peut comprendre pourquoi cette défaillance interdit l'utilisation du langage.

#### **Tout est possible en contexte ?**

La souplesse du langage naturel, utilisé dans la communication, apparaît donc extraordinaire. Selon les contextes, nous pouvons communiquer une variété presque infinie de messages à l'aide des mêmes mots. Ainsi, l'énoncé : "Le dragon a encore terrorisé une innocente victime" peut décrire les méfaits d'un animal légendaire ou d'une voisine acariâtre. Et, avec un peu d'imagination, on pourrait trouver des contextes dans lesquels il communique un autre message encore (Ludwig, 1997).

On peut ainsi se demander s'il suffit de trouver le bon contexte pour que toute phrase puisse tout vouloir dire, ou au contraire, comme Eco (1999) s'il existe des contraintes qui "posent des limites à notre liberté de parole". En effet, les conceptions d'Eco se fondent sur une notion *contractuelle* (donc liée à une notion de contrat entre les individus d'une même communauté) aussi bien de nos schémas cognitifs que du signifié et de la référence. Le fait que le réel puisse poser des limites permet de "modérer une vision éminemment culturelle des processus sémiotiques en fonction du fait que, quel que soit le poids de nos systèmes culturels, il y a quelque chose dans le *continuum* de l'expérience qui pose des limites à notre interprétation" (ibid.). Ceci rejoint donc la discussion abordée au chapitre 1, que nous pourrions résumer avec l'exemple de la découpe du boeuf qui, bien que présentant une infinité de possibles, ne permet pas d'imaginer un morceau qui offre à la fois l'extrémité du museau et l'extrémité de la queue de l'animal.

#### II.4. Résumé

Nous nous sommes d'abord posés la question de savoir ce qui passait du locuteur à l'auditeur lors de la communication d'une idée. Puis nous avons rapporté des modélisations de ce processus et les règles auxquelles il obéit, pour enfin se focaliser sur le processus d'interprétation des énoncés langagiers.

Cette description des théories considérant le langage comme un outil de communication nous sera précieuse pour interpréter les résultats de notre troisième expérience. En effet, tester si on peut retrouver, à partir d'une description, le son dont le locuteur a parlé, c'est une façon de tester la capacité de communiquer nos perceptions auditives au moyen du langage.

De plus, ces considérations nous permettent de positionner le statut scientifique de notre étude. En effet, l'analyse de verbalisations ne peut faire l'économie de l'interprétation (Sperber, 1999). Le point crucial est alors de montrer que l'interprétation d'une locution peut avoir un caractère "objectif" dans un contexte précis du fait que des auditeurs différents en feraient la même.

### **III. Lien entre le langage et la pensée**

Combien de fois avons-nous eu l'impression d'être trahis par l'insuffisance ou l'inadéquation du langage ? Faut-il déduire de cette difficulté que le langage est impropre aux besoins de notre communication ? La pensée a-t-elle quelque chose à perdre dans le langage, ou au contraire à gagner ? Peut-elle exister sans le langage ? Nous allons donc poser ces questions chez l'homme puis chez l'animal.

#### III.1. Le langage signe ou accomplissement de la pensée

Pour Merleau-Ponty (1945), "si la pensée était indépendante des mots, la traduction d'une langue à une autre serait aisée ! (...) Le monde linguistique et intersubjectif ne nous étonne plus, nous ne le distinguons plus du monde même, et c'est à l'intérieur d'un monde déjà parlé et parlant que nous réfléchissons". La question de savoir si le langage était nécessaire à la pensée, si la pensée pouvait être reflétée correctement par le langage a pourtant été souvent discutée. Nous essayerons de rendre compte ici des différents arguments.

##### **Est-ce que le langage peut décrire la pensée ?**

Est-ce un hasard ? Les poètes, qui, pour Eco (1999), réinventent en permanence le langage et "font de l'ambiguïté substantielle du langage la matière même de leur travail", sont souvent ceux qui ont dénoncé les limites du langage. On peut ainsi écouter Gibran (1923) : "les choses furent dites avec des mots. Mais en son coeur restait un discours muet, car il ne pouvait exprimer son plus profond secret." Ou "vous parlez lorsque vous cessez d'être en paix avec vos pensées; (...) Et dans une grande part de vos propos, la pensée gît à demi assassinée. Car la pensée est un oiseau d'espace qui dans la cage des mots peut déployer ses ailes, mais non s'envoler." On retrouve le même type d'interrogations chez Ponge (cité par Kibbedi-Varga, 1996) pour qui la description verbale exhaustive est impossible : les mots ne suffisent jamais. Un portrait verbal a toujours tendance à généraliser alors qu'un portrait peint individualise, mais nécessite un titre. Aucun des deux ne restitue l'individu (Kibbedi-Varga, 1996). Bergson (1938) reprend l'idée que non seulement le langage ne peut rendre compte de la pensée mais qu'il la limite. En effet, le mot, par son caractère conventionnel, "oublie les différences et ne permet que la fixation des généralités". En renvoyant à une classe d'objets plutôt qu'à un objet spécifique de cette classe, le langage a tendance à égaliser les contours. Ceci nous ferait manquer la mobilité qui est la marque de la vraie réalité, et nous habituerait à ne plus la penser. Ainsi, quand nous éprouvons un état d'une inhabituelle intensité, nous ressentons cette inadéquation du langage : "il n'y a pas de mots pour dire ce que je ressens". Cette idée d'un au-delà des mots, l'idée qu'il y a de l'ineffable, que la part la plus précieuse, la plus intime de notre pensée se dénaturerait si on tentait de l'exprimer avec des mots, revient à postuler que la pensée repose par essence sur quelque chose d'antérieur au langage et à l'intelligence et qui est de l'ordre de l'intuition (Cavallier, 1997).



### **Pas de pensée sans mot**

Au contraire, Hegel (1807) écrit : "C'est dans les mots que nous pensons. Nous n'avons conscience de nos pensées déterminées et réelles que lorsque nous leur donnons la forme objective, que nous les différencions de notre intériorité. (...) l'ineffable, c'est la pensée obscure, (...) qui ne deviendra claire que lorsqu'elle trouvera le mot." Cette idée est reprise par Saussure (1955) pour qui : "Sans le secours des signes, nous serions incapables de distinguer deux idées de façon claire et durable". On trouve même une application de cette idée à la description du sonore chez Schaeffer (1966) : "Comme il n'y a guère de pensée sans formulation verbale, il y a peu de chance pour que l'auditeur pense quelque chose de ce son sans recourir implicitement, aux mots et aux idées."

Pour Descartes (1637), le fait même de parler est le signe de la pensée. "En fait, Descartes affirmait que la seule indication certaine qu'un corps possède un esprit humain au lieu d'être un simple automate, c'est son aptitude à utiliser le langage de façon normale" (Chomsky, 1990) ce qui permet notamment de différencier l'homme de l'animal. Fodor (1975) a même imaginé que la pensée prenait la forme d'un langage interne auquel on a donné le nom de "mentalais".

### **Le langage accomplit la pensée**

Merleau-Ponty (1945) présente un point de vue opposé à celui de Descartes : "D'abord, la parole n'est pas le "signe" de la pensée, si l'on entend par là un phénomène qui en annonce un autre comme la fumée annonce le feu. (...) En réalité elles sont enveloppées l'une dans l'autre, le sens est pris dans la parole et la parole est l'existence extérieure du sens. (...) La pensée n'est rien d'"intérieur", elle n'existe pas hors du monde et hors des mots." Pour sa démonstration, il prend l'exemple suivant : "Quand je fixe un objet dans la pénombre et que je dis : c'est une brosse, il n'y a pas dans mon esprit un concept de la brosse sous lequel je subsumerais l'objet et qui d'autre part se trouverait lié par une association fréquente avec le mot brosse, mais le mot porte le sens, et, en l'imposant à l'objet j'ai conscience d'atteindre l'objet" (ibid.). Ainsi, la dénomination des objets ne viendrait pas après la reconnaissance, elle serait la reconnaissance même, ce que Merleau-Ponty résume ainsi : "La parole n'accompagne pas une pensée déjà faite, mais l'accomplit."

### **Pensée pure ?**

Ainsi, "la pensée pure se réduit à un certain vide de la conscience, à un vœux instantané" (Merleau-Ponty, 1945). Ceci rejoint d'ailleurs l'idée que la pratique de la méditation permet d'arrêter la pensée : "Ayez une pensée pendant un dixième de seconde, et vous aurez un objet. Quant on n'a pas une seule pensée, les objets tombent dans l'oubli, l'esprit s'éteint de lui-même et il n'y a plus rien à poursuivre" (Krishnamurti cité par Bianu, 1996). On peut ainsi considérer que la pratique de la méditation, en visant l'arrêt de la pensée, est la fin du langage. "En son fond, le bouddhisme apparaît comme une pratique destinée à arrêter le langage, à suspendre le brouhaha qui constitue la personne" (Bianu, 1996).

## III.2. Et les animaux ?

Grice (1957) affirme que les intentions de communication, du domaine de la pensée, seraient antérieures au langage : "C'est parce que nous *voulons dire* certaines choses avec les signes linguistiques que nous employons, parce que notre comportement linguistique est tout entier dirigé par des intentions de communication, que les termes de notre langage sont dotés de signification." Or, on ne peut nier la présence d'intentions de communication dans le monde animal. Peut-on dire pour autant qu'ils "parlent" ? qu'ils "pensent" ?

#### **Faut-il des mots pour penser ?**

Dès le premier siècle avant l'ère chrétienne, on trouve, chez Lucrèce, l'idée de pensée animale : "Les troupeaux sans parole et les bêtes sauvages par des cris différents expriment la crainte, la douleur et le plaisir qu'ils sentent". Cette idée de continuité entre l'homme et les animaux est reprise par Montaigne (cité par Cyrulnik, 1995) : "Pourquoi considérer que les bêtes n'aient point de pensée ?" Le fait que les animaux puissent éprouver des sentiments proches des nôtres, observables à travers leurs mimiques est ici aussi pris comme preuve de l'existence d'une pensée animale. "Nous devons conclure de pareils affects pareilles facultés" (ibid.). Il est intéressant que l'argumentation repose sur l'idée, qui sera reprise par Darwin (1874), qu'on peut trouver des analogies dans l'expression des émotions chez l'homme et les animaux. En effet, "Darwin accordait aux animaux - séparés de l'homme uniquement par une gradation - les mêmes passions et émotions ainsi que les même capacités intellectuelles d'association d'idées" (Cyrulnik, 1995). Pourtant, on peut se demander si ce n'est pas rejeter l'anthropocentrisme au prix de sombrer dans l'anthropomorphisme ! Plutôt que de parler des mimiques des animaux en général, il serait plus juste de considérer celles de certains animaux, comme les mammifères (voir peut-être seulement les animaux familiers ou domestiques). On peut en effet se demander si nous sommes capables d'interpréter les mimiques d'une mouche. Ceci pourrait d'ailleurs montrer une limite de notre capacité d'empathie ou de projection de nos propres sentiments sur des animaux trop différents de nous...

On a coutume de considérer que Descartes s'est opposé à l'idée d'attribuer des pensées aux animaux. En effet il affirme (1637) que "les paroles et autres signes faits à propos" sont les seules "actions extérieures" qui témoignent de l'existence dans nos corps d'une "âme qui a des pensées". Pourtant, c'est bien en observant des "signes faits à propos", que l'éthologie animale a montré que "dès qu'il perçoit, l'animal confère du sens aux choses qui constituent son monde. Sur l'univers physique, il prélève un matériau à partir duquel il construit ses 'objets' propres" (Cyrulnik, 1995, p 27). Le monde des animaux n'est donc dénué ni de sens, ni d'objets, ni d'intentions de communication, ni de communication d'ailleurs. Le point de discussion n'est pas en effet de décréter que les animaux ne sont pas capables de communiquer, mais plutôt de convenir qu'ils n'ont pas accès à l'abstraction, et qu'ils ne communiquent qu'à propos de leurs besoins vitaux. C'est d'ailleurs le propos de Descartes (1637) : "Il ne s'est jamais trouvé aucune bête si parfaite qu'elle ait usé de signes pour faire entendre à d'autres animaux quelque chose qui n'eût point de rapport à ses passions", les bêtes quant à elles "agissent naturellement et par ressorts, ainsi qu'une horloge", on pourrait retrouver des considérations proches dans l'éthologie moderne, à condition d'éviter d'assimiler les animaux à des automates.

### **Dangers de l'anthropomorphisme**

En effet, Lestel affirme dans Paroles de singes (1995) "les singes n'ont rien à dire ! Le chimpanzé ne dit rien qui n'ait son référentiel précis dans le monde extérieur. Il ne parle que de choses matérielles, dans des situations concrètes. Ironie du sort, cette inaptitude à jouer vraiment avec le langage, à faire de la poésie ou de la métaphysique, à s'interroger sur le sens de l'existence et à interagir avec le langage pour lui-même est précisément ce qui rend FAUSSE cette aptitude au langage." Et Cyrulnik (1995) renchérit : "Les animaux peuvent donc communiquer" mais, "aucun être vivant non homme ne peut transmettre une information en référence à un événement totalement absent." Par exemple, "aucun singe ne peut faire les gestes qui critiquent la pensée d'un autre singe, même si ces deux singes manipulent mieux que moi le langage des sourds américains".

De plus, si on peut considérer la fonction de désignation des objets (pointer du doigt) comme une forme de référentiation pré-langagière, force est de constater que "le chien ne comprend pas le geste de désignation". C'est-à-dire que pour un chien, ce geste ne fait pas signe vers quelque chose d'autre : "Prenez un chien, et indiquez-lui de l'index où aller chercher le morceau de viande qu'il convoite : il vient droit sur votre index pour y poser sa truffe" (Cyrulnik, 1995). On retrouve ici le proverbe chinois : "quand le doigt montre la lune, l'imbécile regarde le doigt". Même si on ne peut pas passer sous silence les contre-exemples des chiens de bergers ou chiens de chasse qu'on parvient à dresser à se diriger dans la direction pointée par son maître, on retrouve bien l'idée de Descartes que le langage et la pensée de l'animal s'enracinent dans le concret.

### **Si les lions pouvaient parler**

En reprenant les arguments permettant de considérer le langage comme fondamentalement inscrit dans une culture, on peut comprendre par exemple les difficultés rencontrées par les Japonais en visite à Paris (Nahapétian & Senk, 2000). En effet, les codes de communication français sont très différents des leurs. Par exemple, le fait de regarder quelqu'un dans les yeux, signe de franchise pour un Français est au contraire de la vulgarité pour un Japonais (Cyrulnik, 1995). Et, comme le montre Merleau-Ponty (1945) : si "la mimique de colère ou celle d'amour n'est pas la même chez un japonais et chez un occidental, (...) la différence des mimiques recouvre une différence des émotions elles-mêmes." Alors que certains soulignent les ressemblances entre les mimiques de différentes espèces animales, ce qui nous permettrait d'interpréter celles des animaux, d'autres remarquent les différences présentes au sein de l'espèce humaine en fonction de la culture de l'individu...

On peut donc affirmer comme Wittgenstein : "Si un lion pouvait parler, nous ne le comprendrions pas" (cité par Cyrulnik, 1995).

### **Evidence d'une pensée sans langage**

Ainsi, si "l'approche éthologique de la parole nous permet de soutenir qu'il existe une pensée sans langage" (Cyrulnik, 1995), elle ne pourrait exister que chez l'homme, et ce même si "mon chien communique plutôt bien ses intentions et tout le monde le comprend quand il veut sortir, manger, jouer, se faire caresser, ou même 'se faire pardonner'" (ibid.).

On a pu en effet réaliser une expérimentation "naturelle" chez des malades qui perdaient pendant quelques heures l'usage du langage, le temps que l'artère sylvienne, bouchée par accident, redevienne perméable. "Quand on lui demande s'il pouvait encore penser pendant ces quelques heures sans parole, le malade (...) répond régulièrement qu'il pensait comme dans un film sans parole, qu'il comprenait ce qui se passait autour de lui, mais qu'il ne pouvait y répondre que par signes gestuels. (...) ce qui différencie l'homme des non-hommes, ce n'est pas tellement la parole, qui peut être considérée sous sa forme matérielle comme un objet sonore appartenant alors à tout être vivant, c'est surtout sa folle aptitude à sémiotiser" (ibid.). Cyrulnik (1995) conclut en forme de provocation : "Et même les étrangers sont peut-être des hommes, alors qu'ils ne comprennent même pas notre langue !"

### III.3. Résumé

La description des pensées d'un homme momentanément privé de sa capacité de langage (à cause d'un infarctus) nous donne une clé pour comprendre les liens entre pensée et langage et nous permet de reprendre l'ensemble des positions, *a priori* contradictoires, présentées ci-dessus.

En effet, la communication animale nécessite un langage mais pas forcément les concepts du langage humain. Les animaux ne semblent en effet pas communiquer sur autre chose que du concret lié à leur survie. Chez les humains par contre, une pensée sans mot est "ineffable", donc non communicable (même à soi-même ?) et reste de l'ordre de l'intuition. Il faut donc en quelque sorte qu'elle s'incarne en mots pour exister pleinement comme "action" de notre monde. Ceci nous semble prendre en compte l'idée d'un langage pour soi-même qui permet de rendre une idée pleinement consciente et peut se rapprocher de l'idée de mentalais de Fodor (1975). De plus, cela nous permet de rendre les points de vue de Bergson et Hegel non antagonistes mais complémentaires : on pourrait avoir des intuitions ineffables mais ce serait leur mise en mots qui les transformeraient en "pensées".

Ces considérations sur la possibilité ou non de transcrire nos pensées à l'aide du langage sont centrales dans notre question du lien entre les sons et la façon dont on en parle. En effet, quand on demande à quelqu'un de décrire ses perceptions, on sous-entend une transcription possible en mots...

## **IV. La vérité sort de la bouche des enfants**

Pourquoi parler d'enfant ? D'un certain point de vue, il n'y a pas eu d'enfant dans cette thèse, qui n'étudie que des verbalisations d'adultes ; d'un autre, il y en a eu trois ! Chloé, Ariane et Rémi, nés tous trois au cours de ce travail, et qui m'ont beaucoup appris, notamment sur le langage. Mais surtout, comme "l'enfant est le père de l'homme", selon la formule de Freud, on peut espérer répondre aux questions posées par le langage adulte en observant comment les enfants s'approprient le langage. Cette idée a d'ailleurs permis la mise en place d'expérimentations éthiquement contestables : Frédéric II de Prusse ayant acheté des enfants pour les élever dans des conditions où il était interdit de leur adresser la parole. Mais l'idée d'observer les enfants dans une perspective proche

de l'éthologie humaine est une voie qui "fut ouverte par Darwin lui-même avec la publication en 1877 de "A Biographical Sketch of an Infant", ouvrage constitué du matériau accumulé depuis 1837 en vue d'une 'histoire naturelle des nourrissons' dont son premier né, William, fut le premier objet" (Cyrułnik, 1995). Darwin s'y intéresse tout particulièrement à l'apparition et au développement du langage, aux manifestations préverbaux du bébé. Nous allons donc essayer de comprendre comment l'observation chez l'enfant de la naissance du langage, (de son apprentissage ? de son appropriation ? de sa création ?) peut nous permettre de comprendre les mécanismes de référence, de communication et de lien avec la pensée, exposés dans les paragraphes précédents.

#### IV.1. Un être de langage

Les psychanalystes, et en particulier Françoise Dolto (1995), insistent pour considérer que l'être humain est avant tout un être de langage et que chez l'enfant "tout est langage", c'est-à-dire que tout acte, toute pensée, tout sentiment, tout objet "veut dire" quelque chose. Ainsi, la conception d'un enfant s'inscrit déjà dans un projet "parlé" de ses parents. On commence à mesurer à quel point chaque étape de sa croissance est "baignée" dans le langage. C'est le cas pratiquement au sens propre pour ce qui est de sa vie intra-utérine, puisque le filtre passe-bas que représente le ventre de sa mère permet d'"entendre" la prosodie du langage (sous forme de vibrations sur la peau puis via les oreilles dès 6/7 mois de grossesse) et de percevoir les émotions de la mère à ce moment. Des enregistrements effectués à l'intérieur du ventre d'une mère permettent par exemple de reconnaître le déroulement d'un journal à la télévision, uniquement grâce à la prosodie, même si on ne peut comprendre aucun mot (Raiman et al., 1994). Et surtout, dès la naissance de l'enfants, ses "productions" seront considérées comme ayant un sens par ses parents.

##### **Prêter du sens aux productions**

Les enfants entrent dans le langage parce qu'on attribue du sens à ce qu'ils prononcent : c'est une procédure d'interaction, présente sans doute dès la naissance. C'est ainsi que Myriam Szejer (1994) décrit le premier cri d'un enfant : "Si le médecin y entend la confirmation que la mutation se fait bien, les parents n'entendent-ils pas, eux, d'abord le son de cette voix nouvelle ? une forme d'appel ? La mère la plupart du temps aura spontanément tendance à entendre dans ce cri l'expression d'un sens, et éventuellement s'en inquiétera : 'il a mal ? ça ne va pas ?' (...) Parce que pour un humain, le cri d'un autre humain a toujours du sens, et que sa préoccupation première consiste à tenter de le déchiffrer. (...) Car désormais, par ce cri, le dialogue entre eux se trouve engagé, et dans chaque cri à sa suite, la mère essaiera d'entendre ce que lui dit son enfant ; qu'il a faim ou soif, chaud ou mal, sommeil, etc. Parce qu'imaginer, pour une femme, qu'un cri émanant de son enfant n'a pas de sens est impensable, et que c'est précisément cette faculté de donner du sens qui caractérise l'humain".

Ainsi, Lestel (1995) critique l'objectivité des "parents"-chercheurs-instructeurs du singe Washoe, qui tentent de lui apprendre le langage en l'accueillant chez eux dès son plus jeune âge, en remarquant que "les instructeurs ont alors tendance à interpréter des signes ambigus de Washoe comme des signes qui lui ont été enseignés auparavant. Un phénomène de renforcement à un stimulus peut ainsi se

produire." En fait ces chercheurs se comportent exactement comme des parents qui sur-attribuent du sens aux productions de leurs rejetons.

### **Regard de l'autre**

Il en est d'ailleurs de même pour ce qui est de considérer que l'enfant "sourit" dans les premiers jours de sa vie. "Le bébé dort, il a les yeux fermés. Tout à coup apparaît ce que nous disons être un sourire. (...) Sur le tracé encéphalographique, nous voyons que le bébé vient d'entrer dans la phase dite du 'sommeil paradoxal'. (...) Vous imaginez difficilement une mère s'extasier : 'il vient de sécréter un neuropeptide !' La maman (...) interprète la mimique faciale comme un sourire. Les conséquences de ce contresens initial s'avèrent considérables, car la voici qui, émue et réjouie, va manifester 'à son tour' sa tendresse (...) qui va en retour influencer le rythme biologique de développement de son bébé !" (Cyrulnik, 1995). Des études ont montré que cette attribution de sens aux productions des bébés étaient déjà complètement inscrites dans les "conventions" et a priori d'une société. Ainsi la mimique d'un nourrisson en train de pleurer est interprétée comme "il a peur" si on dit avant au public que le bébé est une fille ou "il est en colère" si on prévient que c'est un garçon (Cyrulnik, 1989). De même, des observations ont montré que le fait de marcher debout était un comportement social initié et encouragé par l'enthousiasme de l'entourage (Cyrulnik, 1995). Les enfants sauvages, tout comme les "enfants-placards" (élevés dans des conditions de privation affective) marchent à quatre pattes. Cyrulnik définit alors une "Pensée sans parole" comme "une représentation organisée à partir des perceptions de mon propre corps et des émotions issues de mon entourage".

Si on relie ces considérations aux principes de l'interprétation pragmatique d'une phrase et à l'anthropomorphisme qui nous fait interpréter les comportements des animaux, on voit revenir l'idée d'empathie et de projection de nos propres pensées. Considérer a priori que l'autre fonctionne comme nous apparaît comme une des bases du fonctionnement de nos sociétés humaines, mais aussi de leur construction.

On peut alors rejoindre Cyrulnik (1995), quand il dit que : "S'il est indéniable que l'accès au langage crée un univers spécifiquement humain, il n'en est pas moins vrai que, avant de parler, l'univers humain se prépare par une foule d'aventures affectives, (...) prérequis anatomiques et affectifs de la parole".

### IV.2. Pointer du doigt

Le fait de pointer du doigt est un phénomène que nous devons entendre comme "pré-sémiotique, ou protosémiotique (au sens où il constitue le signal qui amorce le processus sémiotique en l'instituant" (Eco, 1999). En effet, la fixation de mon attention ou de celle de l'autre sur quelque chose est la condition de toute attribution de sens à venir. On retrouve chez Cyrulnik (1995) et Tyvaert (2000) cette idée que le fait de pointer du doigt (geste d'"ostenciation") est un préalable au langage. Cela permet en effet de considérer le principe de la référence (le geste désigne quelque chose d'autre que son doigt) comme antérieur au langage, celui-ci ne faisant qu'exprimer cette relation.

### **Première fois**

Or, la fonction de désignation des objets (pointer du doigt) apparaît spontanément chez l'homme. Cette apparition a été observée (Cyrulnik, 1995) à l'aide du dispositif expérimental suivant : un enfant est assis dans une chaise haute avec un objet désirable hors de sa portée; sa mère est présente à côté de lui (et pourrait lui passer l'objet en question). On observe que jusque vers 10 à 13 mois, l'enfant crie et essaye d'atteindre l'objet par lui-même puis entre dans des conduites d'auto-mutilation, que la mère considère comme "caprice" (!). A partir de 10 à 13 mois (en fonction des enfants, de leur sexe...) l'enfant change de tactique : au lieu de se tendre tout entiers vers l'objet pour essayer de l'attraper par lui-même, il pointe seulement l'index vers l'objet en regardant sa mère et en tentant "l'articulation, toujours d'abord ratée, d'un mot, ou proto-mot" (Cyrulnik, 1995). Pour que le pointer du doigt soit possible, l'enfant doit donc cesser de vouloir attraper l'objet pour se l'approprier immédiatement ; il doit de plus acquérir la représentation très élaborée que, par désignation, il peut renvoyer à quelque chose qui se trouve éloigné dans l'espace et qu'il peut obtenir par l'intermédiaire de sa mère.

Ainsi, "le langage ne commence à apparaître que sur la base d'un ensemble comportemental désignatif, lequel suppose une maturation biologique déterminée, et s'instaure non dans un face à face de l'enfant avec la chose qu'il désigne, mais à la faveur d'une double référence affective à la chose et à la personne d'attachement. Et c'est ainsi que la chose peut devenir 'objet' de désignation ; thème d'une vocalisation qui accompagne régulièrement le pointer du doigt" (ibid.)

### **Co-occurrence et construction de la référence**

A partir du moment où l'enfant peut comprendre l'idée de référence par la désignation, il peut repérer des "schémas sonores" (Boysson-Bardies, 1999) et leur association par co-occurrence avec des événements qu'il perçoit. On retrouve ici l'idée de Condillac (1749) qu'"un son ne deviendrait jamais pour un enfant le signe verbal d'une chose s'il n'avait été entendu, pour le moins, une fois au moment où cette chose est perçue."

Ainsi, pour Tyvaert (2000), l'enfant perçoit le monde par le canal de ses capteurs sensoriels, et perçoit aussi les séquences langagières. Très tôt, l'enfant isole les séquences linguistiques prononcées en sa présence (Boysson-Bardies, 1999) : il y a alors "acquisition" de la langue sur la base d'un travail de mémorisation (des formes et des percepts) d'identification et de comparaison. "Ce qui nous importe ici, c'est que les séquences linguistiques sont énoncées par des locuteurs comme autant d'instances d'énonciation et donc en rapport nécessaire avec un paramétrage du monde local (ressenti par ailleurs)." Les énoncés proposés à son intention sont des phrases simples, renvoyant à des situations concrètes, des déclaratives descriptives de structure syntaxique minimale. "Il y a toujours concomitance entre la phrase comme formulation et la situation comme réalité à représenter en langue" (ibid.) C'est ainsi qu'en présence d'un "objet" ou d'une perception saillante, l'enfant pourrait au moyen de la co-occurrence de l'ostension gestuelle et de l'écoute du nom prononcé à son égard ("ça, c'est un oiseau") arriver à construire peu à peu un vocabulaire. Tyvaert remarque que dans le cas des adjectifs ("ça, c'est jaune"), "il ne s'agit plus d'associer le mot à un existant individuel, mais seulement à une perception considérée comme suffisamment saillante pour être repérée." On retrouve une fois encore cette idée de saillance perceptive à la base de l'explication de lien référentiel. Pourtant,

si un objet saillant semble apparaître comme une évidence, l'idée de saillance perceptive est loin d'être explicitée : pourquoi dans un champ perceptif donné certaines "choses" apparaîtraient spontanément plus facilement que d'autres...

#### **La syntaxe prend le relais**

Néanmoins, il est impossible de concevoir que l'acquisition de chaque mot nouveau nécessite de rencontrer un objet prototypique pouvant lui servir de référent. Il n'y a qu'à mesurer la patience qu'il faudrait pour attendre que tombe sous notre sensibilité un "oiseau jaune" pour conclure à l'absurdité d'une démarche qui n'utiliserait pas, au moment opportun, la structure du langage. La maîtrise progressive des constructions syntaxiques démultiplierait ainsi la capacité référentielle, bien au delà de l'exploitation des termes du lexique (Tyvaert, 2000). On peut comprendre alors la rapidité avec laquelle un enfant maîtrise sa langue maternelle : son propre discours sur le monde est très vite efficient. En effet, "quelques initialisations réussies (c'est-à-dire vérifiées) peuvent suffire à convaincre un jeune cerveau de l'intérêt de l'association référentielle, la démarche d'acquisition elle-même encourageant à assimiler avec méthode (on passe alors en régime d'apprentissage) les couples entités lexicales / entités sensibles qui, par inscription dans des phrases selon une syntaxe, vont démultiplier les associations référentielles bien au-delà des ostensions liées au coup par coup à la sensation. En venant accompagner le développement cognitif de l'enfant, la langue en accélère considérablement le rythme. (...) On peut en effet comprendre la complexification des idées premières, premières représentations mentales associées à des phrases-mots, en pensées secondes, de plus en plus complexes et abstraites, comme un effet des structures de la langue, comme l'avait déjà soutenu Condillac." (ibid.)

#### IV.3. Explosion

En dix mois, l'enfant apprend plusieurs centaines de mots, et se met à jouer avec eux. Entre le vingtième et le trentième mois, sans le secours d'aucune école, le voici qui apprend les mots, l'accent, les règles et même les exceptions aux règles. Non seulement l'enfant comprend que le mot renvoie à la chose, mais il ne tarde pas à découvrir en plus que le mot peut renvoyer non à la chose directement, mais à un autre mot qui renvoie à la chose (Cyrulnik, 1995). Cette explosion des capacités langagières de l'enfant est très impressionnante. Pourtant, elle ne se fait pas sans efforts. Nyckees (1997) souligne que "c'est aux nouveaux venus qu'il revient d'édifier laborieusement leurs compétences linguistiques, au prix de millions d'inférences, à partir des expériences variées que constituent la perception en situation des performances de leurs modèles, les réactions inévitables de leurs tuteurs et modèles linguistiques à l'égard de leurs propres performances, les performances et les réactions de leurs pairs ou de leurs cadets et, d'une manière générale, le succès ou l'échec de leurs interactions de toutes sortes avec les membres de leur communauté linguistique."

#### **Prosodie**



Un enfant peut tenir de grands discours (et être compris !) sans utiliser réellement des "mots" bien formés, mais juste en modulant les intonations de sa voix grâce à la prosodie. Et cela communique déjà beaucoup de choses : émotions, désirs, satisfaction, colère, douleur, questionnement, requête... Et puis, alors que certains s'entraînent rapidement à prononcer des mots isolés mais "ad-hoc", d'autres continuent à pratiquer leurs mélodies babillées qui se structurent insensiblement en vraies phrases. J'ai eu ainsi la surprise de voir un enfant de un an et demi s'adresser à moi en m'indiquant extrêmement clairement "tu peux me passer mes chaussures" (ou "tu pourrais me mettre mes chaussures"), ceci sans aucun mot correctement prononcé, en imitant uniquement les caractères de surface de la phrase (sons, intonations) avec l'intentionnalité, le geste et le sérieux de celui qui dit quelque chose d'important et qui n'a pas de doute que vous allez comprendre et obtempérer !

Il y a communication claire du sens, même sans les mots, le but étant de se faire comprendre. Pourtant, si toute mère fait vite la distinction entre pleurs de faim, de fatigue, d'énervement, de peur, de douleur... de son propre enfant, ces catégories sont assez grossières, et peu compréhensibles par des personnes extérieures à une relation de soins quotidiens. C'est ainsi que l'intervention d'un tiers (le père, l'école...) en plaçant la communication de l'enfant dans le cadre d'un groupe de plus en plus grand, l'oblige à rendre son langage de plus en plus conforme à celui du groupe. L'enfant pour être compris doit alors faire l'effort d'utiliser des mots plus "conventionnels".

#### **Premier mot : "papa" ou "maman" ?**

Et puis (ou Et avant, ou Et en même temps) on voit apparaître les premiers "mots". Là intervient massivement le phénomène de sur-interprétation déjà souligné. Si le mot "papa" est considéré comme le premier mot prononcé en général par un enfant, est-ce parce que c'est leur concept le plus important, est-ce parce que les syllabes "papa" ont été choisies pour désigner le père parce que c'est une des premières syllabes prononcées clairement par les enfants, ou est-ce parce que les adultes avides d'entendre quelque chose dite par leur enfant se précipitent sur ses premières articulations pour leur donner ce sens ? Sans doute les trois à la fois. Le premier signe du singe Loulis est d'ailleurs lui aussi le nom de son instructeur le plus proche (Lestel, 1995).

Anecdote personnelle : Rémi, âgé de quelques minutes, exprimait une plainte pleine de "mam mam mam..." J'en ai immédiatement profité : "il a dit maman !" Ceci peut paraître anodin, pourtant, nul doute que si l'enfant ressent à quel point ces "mam mam" ravissent sa mère, il ne ré-essaye de les produire, même si au départ, cela peut apparaître comme une façon "conditionnée" de se procurer une atmosphère agréable. Ainsi, on pourrait faire un parallèle avec le procédé de changement de sens décrit par Nyckees (1997) qui conçoit celui-ci comme passant par une étape où les deux sens, ancien et nouveau, coïncident. L'enfant pourrait ainsi passer sans discontinuité d'onomatopées liées à ses émotions (bien-être ou désagréments) à de vrais mots, c'est à dire des sons signes d'autres choses. Ainsi on peut transposer l'explication du changement de sens du verbe "*arriver*" décrit ci-dessus sur le fait d'entendre "maman" dans ces "mam mam". Si pour Rémi "mam mam" n'est qu'une onomatopée correspondant à son ressenti mais que pour moi, c'est le fait de dire "maman", notre interaction va pouvoir faire glisser le sens vers le langage. Ces "mam" ne devenant "mots" que quand ils seront prononcés sans lien direct avec le ressenti initial. Ainsi, le plaisir de babiller (faire sonner sa voix) et le

plaisir de l'échange auquel on ajoute la sur-interprétation de l'adulte en empathie permet d'attirer l'enfant dans le langage. Cette idée d'une transition "onomatopées liées aux émotions ressenties - mot" permet en outre de comprendre l'idée de Merleau-Ponty (1945) que les sons d'une langue ne sont pas dus au hasard mais sont une forme de reflet des émotions vécues dans cette langue. En effet, "il faut bien ici que le sens des mots soit finalement induit par les mots eux-mêmes, ou plus exactement que leur signification conceptuelle se forme par prélèvement sur une 'signification gestuelle', qui elle, est immanente à la parole, de même qu'en pays étranger, je commence à comprendre le sens des mots par leur place dans un contexte d'action et en participant à la vie commune" (Merleau-Ponty, 1945).

### **Son - onomatopée - mot**

Il semble que les premiers mots des enfants ont des sens étendus, parce qu'ils ont tendance à étendre l'utilisation des mots à des nouveaux référents qui partagent une des propriétés perceptives saillantes du référent original (Vygotski, 1934 / 1997). Ce comportement est aussi observé chez le singe Loulis qui "attribue d'abord des significations trop larges qu'il réduit peu à peu" (Lestel, 1995). Ceci correspondrait aussi à une observation personnelle, un des premiers "proto-mots" d'Ariane, un an et demi, étant : "boum" et pouvant prendre les sens multiples de "t'as vu le bruit quand ça tombe", "c'est tombé", "ça va tomber", "attention, tu vas tomber" (en voyant quelqu'un monter sur un tabouret). Un autre exemple peut être donné avec "wroum" qui peut désigner une voiture (en jouet, dessinée, réelle) et par extension tout véhicule en marche ou à l'arrêt ! On peut noter que ces deux exemples sont une façon de transformer en mots (donc en signes) les bruits produits par des événements ou des actions. Cette remarque sera à confronter avec le fait que le timbre, en renvoyant à la source sonore qui la produit, se fait signe naturel de cet objet.

L'exemple de "boum" permettrait de comprendre comment on peut apprendre le sens de verbes par ostension. En effet, au contraire des noms et des adjectifs, ceci semble difficile à imaginer pour Tyvaert (2000) : "Certes, on pourrait tenter une 'ostension' en déclarant, en situation, quelque chose comme 'tomber, c'est ça'. Mais comment exhiber une situation où tomber soit saillant, sans que l'attention se précipite sur un objet présent ? (Cordier & Tyvaert, 1999) . En fait, et en mettant les choses au mieux, c'est-à-dire en supposant qu'on dispose d'une situation où l'exophorique désignera sans ambiguïté, il est fort probable qu'on déboucherait sur l'établissement de la capacité référentielle du nom 'chute', ce qui suggérerait plutôt une formulation du type 'une chute, c'est ça' (Tyvaert 2000)." Si le sens des mots est d'abord plus étendu, le même vocable pourra au début désigner l'action la cause et son résultat.

### **De plus en plus de mots**

L'accélération prend des proportions de plus en plus impressionnantes : "les enfants de deux ans apprennent 8 à 10 mots nouveaux par jour" (Soja et al., 1991), et souvent, une seule présentation suffit ! "L'enfant nomme d'abord les objets de son environnement familier important pour lui. Les premiers termes référentiels sont donc des noms d'objets. Nelson (1985) met en évidence que la majorité des étiquettes catégorielles acquises par les enfants entre 1 et 2 ans se rapportent, soit à des

objets sur lesquels l'enfant peut agir (chaussures, jouets, nourriture...), soit à des objets susceptibles de se déplacer dans l'environnement (animaux, voitures...) Il souligne ainsi l'importance du mouvement ou du changement dans l'acquisition du lexique pour le jeune enfant. Clark (1973) accorde, quand à elle, plus d'importance aux caractéristiques visuelles des objets. "L'enfant catégorise également les verbes, mais ce sont certainement les verbes d'action, c'est-à-dire ceux qui impliquent un changement dans l'espace et/ou le temps qui sont les premiers étiquetés" (Cordier, 1987). La représentation que les enfants ont de ces actions est souvent la représentation de l'action terminée, à la condition qu'elle soit maîtrisée -"c'est par exemple le cas pour le verbe 'emplir' mais non pour le verbe 'grimpe' " (Cordier, 1986).

### **"Mots d'enfants" et polysémie**

Quelques anecdotes personnelles comparées entre collègues nous ont permis de réaliser que la plupart des "mots d'enfants" dont les adultes sont si friands, sont souvent le témoignage de la façon dont l'enfant s'approprie peu à peu la polysémie. Ainsi par exemple, Chloé, quatre ans et demi, réagit à ma déclaration : "J'ai vu la secrétaire de M. L." par : "C'est quoi une secrétaire, ça dit des secrets ? Elle t'a dit quoi comme secrets ?". De même, alors que son papa lui disait : "le chameau a 2 bosses", un petit garçon de 3 ans réagit par "où s'est-il cogné ?". Ceci semble faire preuve d'une logique implacable mais très rigide, liée à la mémoire d'un seul cas d'utilisation. En effet, il faut connaître déjà beaucoup de cas pour pouvoir anticiper plusieurs sens possibles. Et la mémoire phénoménale des jeunes enfants (peut-être liée à l'effet de primauté : ils vivent plein de choses pour la première fois ) leur permet peut-être pendant un certain temps de garder tous les exemplaires différents, un peu à la manière du mnémoniste de Luria (1970), dont la mémoire était tellement exceptionnelle qu'il n'avait pas besoin de généralisation, mais aussi qu'il ne pouvait pas catégoriser comme identiques des situations différentes.

Le fait que la plupart des situations d'énonciation entre parents et jeunes enfants concernent d'abord des situations concrètes (Cordier & Tyvaert, 1999), permettrait de comprendre pourquoi on a si longtemps considéré que le sens concret d'un mot était son sens premier, même si cette idée est controversée au niveau historique. En effet, si Nyckees (1997) considère comme normale l'évolution du concret vers l'abstrait à l'échelle de l'histoire des langues, argumentant que la première expérience est l'expérience sensori-motrice immédiate des objets concrets, Victorri (1999) conteste cette idée ("qui prouve que l'on n'a pas appelé l'animal 'lion' parce qu'on avait déjà l'idée 'se battre comme un lion' ?") en donnant pour exemple que le mot *table*, vient de *tabula* dont le sens serait tablettes, support d'inscriptions, et non pas *mensa* qui renvoie au sens concret du meuble. Si le premier sens au niveau du groupe n'est pas forcément concret, au niveau individuel, il est celui qu'on rencontre en premier (par exemple secret avant secrétaire ?).

### **Inventions et non-mots**

Les enfants semblent produire beaucoup de néologismes (Clark, 1973) qu'on peut considérer comme des "non-mots". Pourtant, nous considérons a priori que ces productions ont un sens, nous nous efforçons de les comprendre et nous y réussissons souvent sans difficulté. Par exemple, s'il s'agit de

manger des noisettes "passe moi le *croquoir*" ne posera pas de problème ; de même tout le monde comprend bien et trouve même assez rusé de la part d'un jeune enfant d'appeler *pesette* une balance. Ce comportement ne nous paraît pas différent du fait d'appeler, pour un aphasique, *rougette* un coquelicot. De même, la pluridisciplinarité nous offre la joie de nous heurter régulièrement à des mots que nous ne comprenons pas (par exemple "trope" pour un non linguiste). Pourtant, ces "non-mots" étant utilisés dans des phrases qui ont un sens, ils nous obligent à postuler qu'ils ont un sens, et à le chercher !

Ainsi, dans notre pratique quotidienne, nous sommes amenés, à cause du principe fondamental de coopération ou de la maxime de pertinence de Grice, d'une part à faire des inférences pour comprendre des mots nouveaux ou des utilisations nouvelles de mots connus, et d'autre part, nous sommes obligés d'inventer des mots nouveaux ou des utilisations nouvelles de mots connus pour rendre compte de nos vécus. Cette souplesse forcée est peut-être centrale dans notre façon d'utiliser le langage et permettrait de comprendre pourquoi le phénomène de polysémie est aussi massif dans le langage : c'est ce qui nous permettrait de rendre compte de tout ce que nous vivons avec un ensemble de mots réduit.

#### IV.4. Résumé

Quand on considère l'apparition du langage vers 2 ans, on oublie souvent que c'est seulement le moment à partir duquel chaque enfant réinvente le langage, en étant dedans depuis plus de 2 ans, et en ayant compris la relation de désignation (donc de référence) entre 10 et 13 mois.

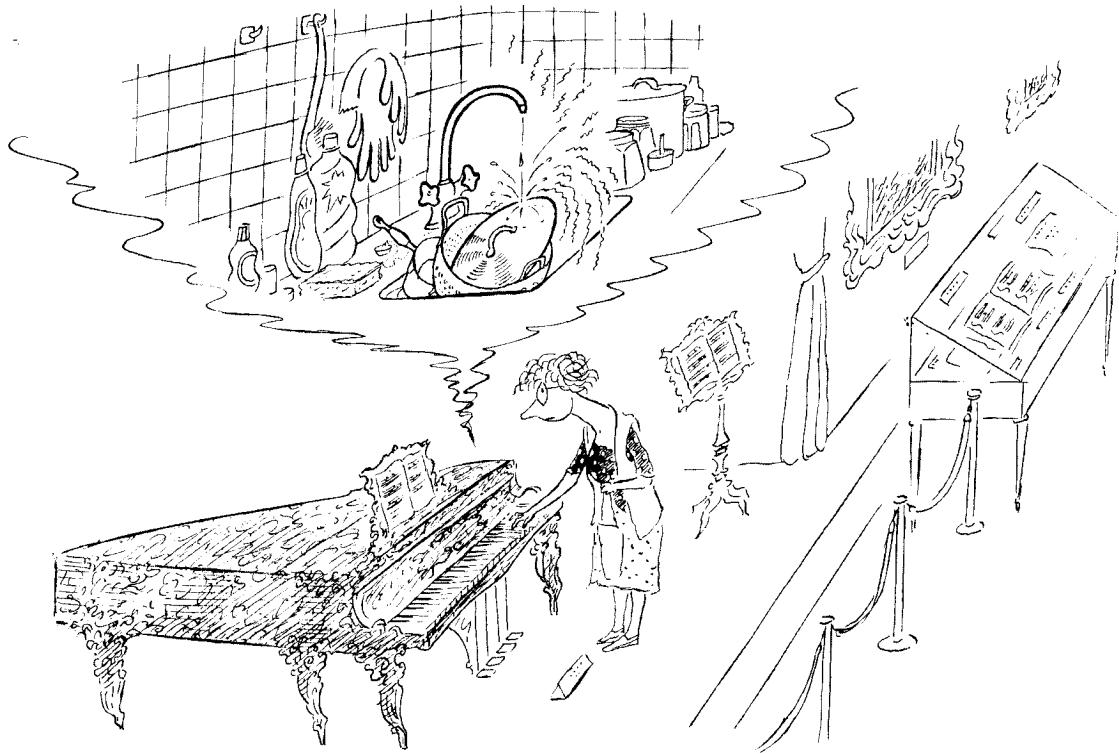
Il est donc intéressant de noter les conditions qui rendent la naissance du langage possible. Il s'agit de pré-requis neurologiques (par exemple l'idée de référence n'apparaît pas avant 10 mois) de pré-requis éthologiques (l'idée de vivre dans un monde de langage) et de pré-requis affectifs (l'idée de rencontrer quelqu'un qui me donne envie de communiquer). En effet, le langage ne peut être que social : "la présence autour de l'enfant de quelque autre être à qui parler, pour qui parler" est absolument nécessaire à son apparition, ce que Cyrulnik (1995) résume par : "Avant de parler, il faut que le développement de mon cerveau d'homme soit correctement programmé, il faut que mes yeux rencontrent une figure d'attachement pour me donner envie de parler, il faut qu'autours de moi le bain parolier social des adultes m'imprègne. (...) Ce nouveau regard éthologique sur la parole en change le statut : la parole ne tombe plus du ciel, elle s'enracine dans le corps, dans l'affectif et dans le social."

Pour finir sur les rapports entre la matière et l'esprit : "On a toujours remarqué que le geste ou la parole transfigurait le corps, mais on se contentait de dire qu'ils développaient ou manifestaient une autre puissance, pensée ou âme. On ne voyait pas que, pour pouvoir l'exprimer, le corps doit en dernière analyse devenir la pensée ou l'intention qu'il nous signifie. C'est lui qui montre, lui qui parle" (Merleau-Ponty, 1945).

## V. Conclusion

Cet apparent détours par une présentation de diverses théories linguistiques explorant les conceptions du langage, comme référentiel ou comme outil de communication, va nous permettre de garder à l'esprit toute sa complexité et sa richesse au sein d'une expérimentation de psychologie. En effet, la plupart des expériences de dénomination effectuées en psychologie sous entendent une conception référentielle du langage proche d'une relation de bijection : objet du monde  $\Leftrightarrow$  étiquette verbale. En particulier, des tests utilisés en neuropsychologie rassemblent toute une série de normes de dénomination d'images qui peuvent faire croire que le lien "image-mot" est fixe, au point qu'on dise de la réponse de quelqu'un qu'elle est juste ou fausse. Au contraire, c'est justement la nature de ce lien "son-mot pour le décrire" qui va être étudiée ici. Nous considérerons donc toute description comme "juste", à nous de comprendre pourquoi, c'est-à-dire comment quelqu'un décrivant ce son a pu utiliser ces mots. On rejoint donc le point de vue de Dubois & Roudy (1997) qui, dans leur étude de la description des odeurs, rejettent une théorie sémantique implicite du langage comme nomenclature. Affranchi de telles théories, on peut en effet concevoir la catégorisation et l'identification du monde comme résultant d'activités humaines plutôt que comme réception passive d'invariants "objectifs".

Enfin, deux résultats nous semblent importants à garder à l'esprit. Notre facilité à comprendre la nouveauté, que ce soit un mot nouveau inventé, ou un changement de sens (c'est-à-dire à attribuer un référent possible à ce "son articulé" dans le contexte de communication considéré), qui apparaît comme une des caractéristiques de notre façon de composer avec le sens des mots. Ainsi, le sens d'un "mot" semble se construire dans une interaction au sein d'un contexte précis. La polysémie des langues naturelles serait donc un moyen très utile d'absorber la nouveauté et la spécificité des expériences vécues sans explosion de vocabulaire. On peut ainsi suggérer qu'un bon système de compréhension artificiel devrait être capable de comprendre lui aussi polysémie et néologismes, par exemple qu'une *pesette* est une balance.



Sempé (1986)

## Chapitre 3. Des sons

Après nous être intéressés à nos rapports avec le monde, notamment à travers la perception que nous en avons, après avoir questionné ce que pouvait comprendre l'idée de parler, nous allons aborder ici la partie théorique de ce travail concernant le domaine sonore. Nous avons déjà défini le timbre en introduction et nous allons présenter ici les études effectuées sur l'identification, la catégorisation, la reconnaissance et la description de nos perceptions, en particulier de nos perceptions auditives : sons, bruits et timbre. Alors que les travaux présentés dans les deux chapitres précédents relevaient souvent de la philosophie et de la linguistique, ceux qui seront présentés ici relèvent en général de la psychologie. En particulier, le paragraphe "Catégorisation..." reprendra cette notion, introduite dans le chapitre 1, mais dans une perspective de psychologues et en lien avec des résultats sur les sons. De même, le paragraphe "Verbalisation..." reprendra l'idée de parler de ses perceptions mais vu sous l'angle de la psychologie (introspection, échelles sémantiques...) alors que le chapitre 2 l'abordait sous l'angle linguistique. Quand au premier paragraphe ("son...") il permettra une introduction à quelques notions d'acoustique élémentaire, nécessaires pour comprendre la partie expérimentale de cette thèse, puis reprendra les études sur les deux principales attitudes d'écoutes présentées dans l'introduction : identifier la source vs les qualités du son.

### I. Son, timbre, représentation perceptive

Nous allons ici reprendre brièvement les définitions de base de ce que représentent son, bruit et timbre, dans le but de faire une introduction à la description acoustique des sons pour des lecteurs peu familiers de ces notions. Puis nous présenterons plusieurs points de vue sur une représentation perceptive des sons. Nous verrons que la ligne de démarcation entre les différents travaux présentés rejoint le débat classique entre les défenseurs d'une perception d'objets et ceux d'une perception de caractéristiques. Nous verrons comment ce débat s'actualise dans le domaine sonore.

#### I.1. Acoustique

Il est parfois des moments où faire appel à une représentation extérieure à un domaine permet de se donner l'illusion d'une objectivité. Nous allons donc définir ici les "objets" de notre travail du point de vue de l'acoustique, la branche de la physique s'occupant du son. Nous verrons dans le cadre de la discussion que, particulièrement dans le domaine sonore, il semble que les outils de visualisation des sons inventés par les physiciens révèlent, autant que les verbalisations de musiciens comme de néophytes, la façon dont nous nous représentons les sons : essentiellement comme des formes temporelles.

La psychoacoustique se donne pour but principal de comprendre le lien entre d'une part nos perceptions et d'autre part la description acoustique des sons perçus. En effet, comme le son d'une part se laisse manipuler (filtrer, couper, additionner...) et que d'autre part, il se propage "lentement" (300 m/s, comparé aux 300 000 km/s de la lumière), on peut envisager de le modifier soit avant de le produire, soit entre le moment où il est produit et le moment où on l'entend. Comprendre le lien entre ces paramètres physiques qu'on peut manipuler et nos perceptions permet donc d'envisager de modifier les sons de notre environnement pour les rendre plus agréables par exemple. C'est un enjeu actuel, notamment dans le design sonore mais aussi dans la lutte contre le bruit.

(On peut trouver de bonnes introductions à l'acoustique et à la psychoacoustique dans Pierce (1984), McAdams (1994), Castellengo (1994) ...)

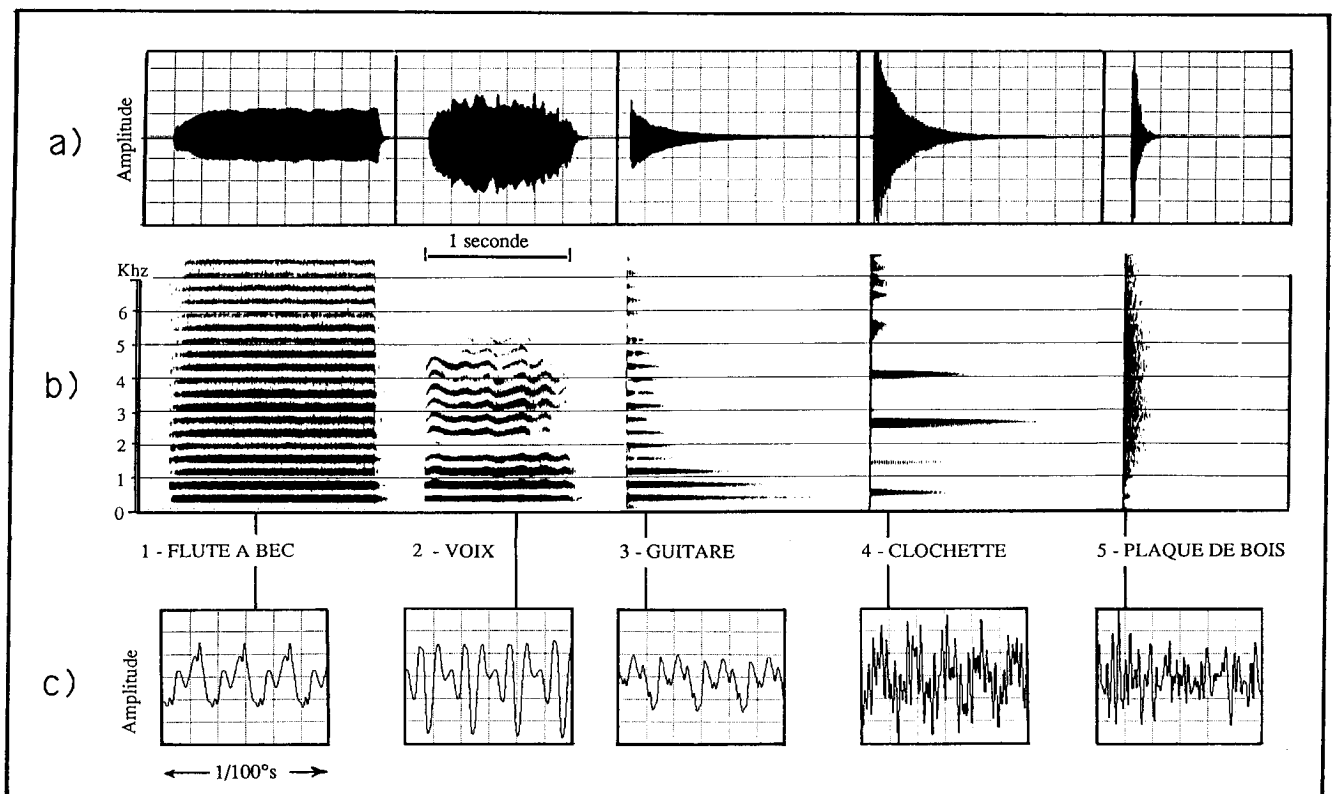


Figure 5 : Extrait de Castellengo (1994) : Analyse acoustique de 5 sons, note sol<sup>3</sup> jouée à la flûte à bec, chantée (voyelle A) et jouée à la guitare, puis clochette de métal et percussion sur une plaque de bois. En a) évolution temporelle de la dynamique opposant les signaux entretenus (1 et 2) et non entretenus (3, 4, et 5). En b) sonagramme des 5 sons. En c) forme d'onde prise au point indiqué sur le sonagramme.

### Son

La définition physique de ce qu'est un son renvoie à une onde de pression engendrée par une action sur un matériau et transmise par un milieu (en général au final l'air) jusqu'à nos oreilles. L'oreille interne reçoit cette onde de pression et la transforme en influx nerveux.

Un son composé d'une seule fréquence (un son "pur") est donc défini dans le monde physique par sa fréquence, son intensité et son évolution temporelle. Un son complexe est composé d'un ensemble de fréquences différentes que la perception fusionne en un seul "évènement". La figure 5 représente



plusieurs types de sons complexes. L'évolution temporelle des différentes fréquences composant chaque son est représentée sur le sonagramme (b).

En caricaturant, la sensation de hauteur d'un son (de quelle note il s'agit) est principalement due à sa fréquence dans le cas d'un son pur, à sa fréquence fondamentale (c'est-à-dire de la fréquence la plus basse) dans le cas d'un son complexe harmonique (comme la flute, la voix et la guitare figure 5). La sensation d'intensité sonore, appelée sonie, est principalement liée à l'amplitude de l'onde de pression, représentée dans la partie supérieure de la figure 5 (a).

#### **Bruit - son complexe harmonique / inharmonique**

Un bruit est une sorte de son complexe qui ne présente en général pas de hauteur saillante (par exemple la plaque de bois, figure 5). Au contraire d'un son complexe harmonique, pour lequel les différentes fréquences qui le composent se répartissent selon une série harmonique décrite par Fourier, les fréquences composant un bruit se présentent "en paquet" sans ordre particulier : sur le sonagramme (b-5), on ne voit pas ressortir de fréquence particulière.

On peut noter aussi l'existence de sons complexes inharmoniques (par exemple le son de clochette, figure 5). Pour ces sons, on entend des fréquences claires (au contraire des bruits) mais leur répartition ne suit pas la loi de Fourier. On peut donc entendre plusieurs hauteurs différentes.

#### **Timbre**

Nous avons déjà donné deux définitions du timbre en introduction : une définition analytique—ce qui différencie deux sons de même hauteur, même sonie et même durée—, et une définition globale—ce qui nous permet de reconnaître la source sonore qui a produit un son. Le timbre d'un son s'il rassemble à la fois l'idée d'"indication de provenance" en plus de représenter "sa forme et sa matière sonore" va en fait être le moyen de "sa complète description" (Schaeffer, 1966).

Du point de vue de l'acoustique, le timbre est dû à la covariation d'un ensemble de paramètres physiques. En effet, on parle de timbre d'un son complexe, composé donc de plusieurs fréquences qui varient chacune en intensité au cours du temps. Des études sur sa perception ont montré que les différentes dimensions perceptives qu'on entendait dans le timbre (Plomp, 1970; Grey, 1977; McAdams et al., 1995) pouvaient être reliées aux paramètres acoustiques suivant : temps d'attaque (Logarithme du temps de montée de l'enveloppe temporelle : LTM), de l'importance relative des différents partiels (Centre de Gravité Spectral : CGS) et de l'évolution temporelle du spectre des sons (Flux Spectral : FL) (Krimphoff et al., 1994). Ces paramètres apparaissent sur la figure 6 qui montre deux types de représentation de deux des sons de l'ensemble étudié dans la partie expérimentale de ce travail. La partie supérieure de la figure montre l'enveloppe temporelle des sons, c'est-à-dire l'évolution de l'intensité au cours du temps. Le temps d'attaque des deux sons représentés est très court. On peut remarquer une remontée de l'intensité à la fin du son de clavecin (hcd.aiff), qui correspond à la perception du retour de souterault (petit étouffoir) sur cet instrument. On retrouve ces deux moments sur la partie basse de la figure qui représente un sonagramme de ces deux sons. On peut remarquer que le son de clavecin est plus riche en harmoniques que le son de "vibron" (vbn.aiff), qui présente, par contre, une double attaque.

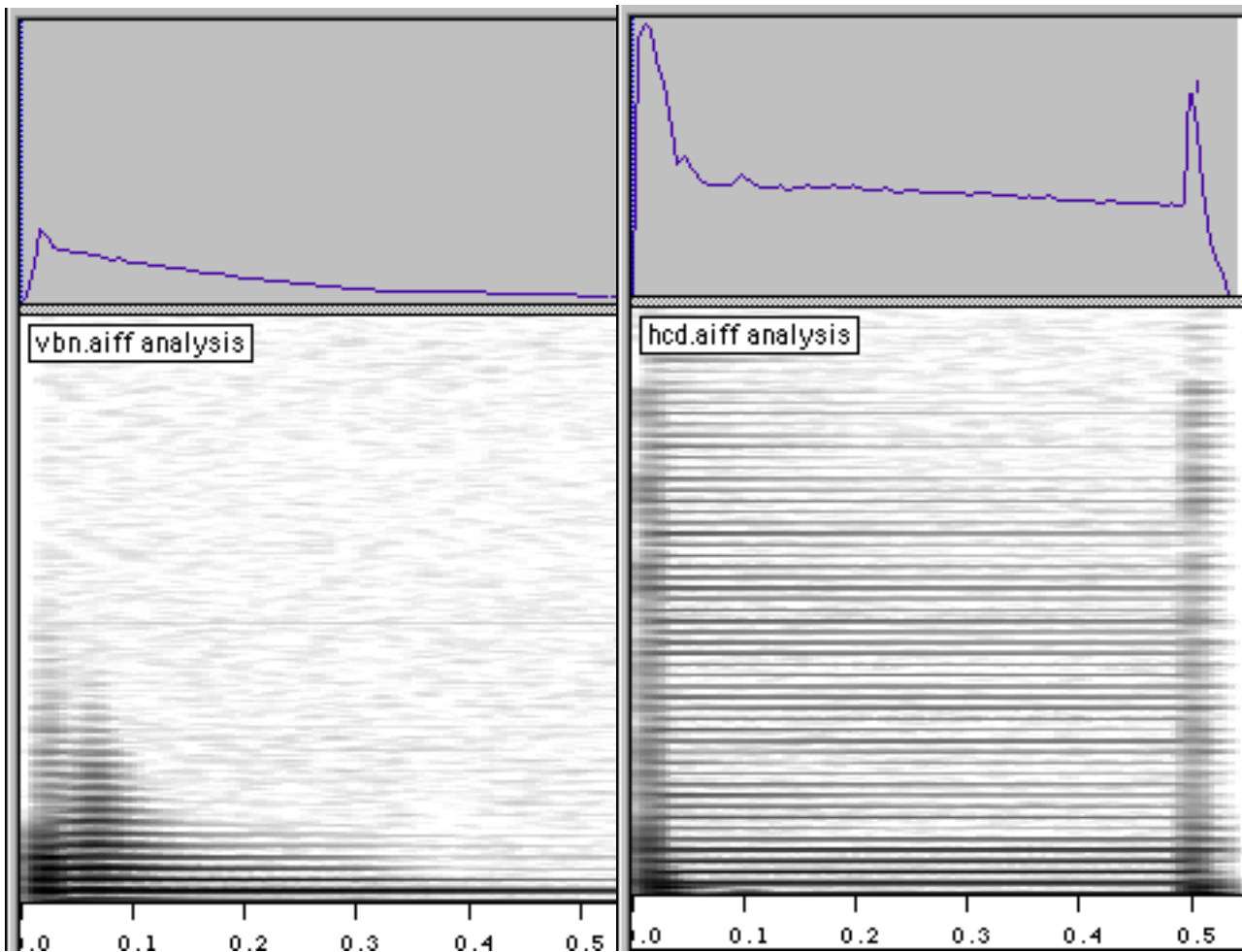


Figure 6 : Représentations acoustiques (enveloppes temporelles et sonagrammes) de 2 des sons de notre ensemble : le son hybride de vibrone (vbn) et le son imitant le clavecin (hcd)

On notera néanmoins que l'identité de la source ne semble pas directement liée à l'analyse fréquentielle de chacun des sons qu'elle produit. En effet, on peut écouter (Bregman & Ahad, 1995) un basson qui produit successivement tous les sons de sa tessiture, du plus grave au plus aigu. On garde la sensation que c'est du basson malgré le fait que chacun de ses sons n'est pas défini acoustiquement par les mêmes paramètres (chaque son à un "timbre" différent). Si au contraire, on prend le "timbre" du son le plus grave du basson et qu'on le "fait monter" en transposant sa fréquence fondamentale (changeant de note), on obtient la perception d'un instrument qui change au cours de la montée. On ne peut plus reconnaître le basson quand les notes sont aiguës.

### I.2. Perception d'un objet versus de qualités

Le psychologue expérimentaliste américain George A. Miller imaginait, en 1962, cette anecdote pour illustrer le débat classique entre les défenseurs d'une perception d'objets et ceux d'une perception de caractéristiques, suivant qu'on considérait la représentation de l'objet comme une connaissance perceptive ou, au contraire, postérieure à la perception immédiate (cité dans Jimenez, 1997).

"Imaginons que quelqu'un est en train de visiter un laboratoire de psychologie vers 1915. Un psychologue l'approche et, sans faire de présentation, lui demande ce qu'il voit sur la table.

- Un livre.

- Oui, naturellement que c'est un livre, dit le psychologue, mais qu'est-ce que vous voyez *en réalité* ?

- Que voulez-vous dire par là ? Je viens de vous dire que je vois un livre, un petit livre avec une couverture rouge."

Le psychologue insiste : "Quelle est votre perception réellement ? Je vous demande de la décrire avec la plus grande précision possible.

- voulez-vous dire que ce n'est pas un livre ? De quoi s'agit-il, d'un piège ?" Le visiteur commence à s'impatienter.

- "Oui, c'est un livre, il n'y a pas de piège. Ce que je veux c'est que vous me décriviez ce que vous voyez exactement, ni plus ni moins."

Le visiteur devient méfiant. "Eh bien, dit-il, de l'endroit où je me trouve, la couverture du livre ressemble à un parallélogramme rouge foncé.

- Oui, dit le psychologue très content, oui, vous voyez une partie rouge foncée en forme de parallélogramme. Et quoi d'autre ?

- Au-dessous il y a un bord blanc grisâtre et sous ce bord une ligne étroite de la même couleur rouge obscure. Au-dessous je vois la table..."

Le psychologue sursaute.

"Au dessous je vois quelque chose de marron avec des franges irrégulières plus ou moins parallèles d'un marron plus clair.

- Très bien, très bien." Le psychologue remercie le visiteur de sa collaboration."

Miller (1962) concluait sagement sur une pluralité d'attitudes perceptives ou, mieux, d'activités perceptives et de leurs représentations subséquentes. Pourtant, quand quelqu'un présente spontanément ce genre de description, s'attachant uniquement à des qualités visuelles sans pouvoir nommer l'objet perçu, c'est le signe de troubles de la perception qu'on appelle "agnosie visuelle". L'observation d'un tel patient a été par exemple décrite par O. Sacks dans L'homme qui prenait sa femme pour un chapeau (1988) :

"J'ouvris un exemplaire du National Geographic Magazine et lui demandait d'en décrire quelques photos. Ses réponses furent très curieuses. Ses yeux sautaient d'un point à l'autre, il remarquait des détails imperceptibles (...) une brillance, une couleur, une forme arrêtaient son attention et lui tiraient un commentaire, mais en aucun cas il ne voyait une scène dans son ensemble. Il ne parvenait pas à voir le tout, mais seulement des détails qu'il enregistrerait comme des tâches sur un écran radar. Il ne considérerait jamais l'image dans son ensemble - il n'affrontait pour ainsi dire jamais la physionomie de l'image : le paysage ou la scène n'avait pour lui aucun sens."

On trouvera un exemple complet de description ci-dessous montrant que cet homme ne peut accéder au "sens" de l'objet par la vision, altérée par cette agnosie visuelle, mais qu'il parvient à le faire à l'aide de son odorat, intact :

"En venant, je m'étais arrêté chez un fleuriste et avait acheté une extravagante rose rouge pour ma boutonnière. Je l'enlevais et lui la tendis. Il la prit comme un botaniste ou un morphologiste s'empare d'un spécimen et non comme une personne reçoit une fleur.

- environ quinze centimètres de long, commenta-t-il. Une forme rouge enroulée avec une attache linéaire verte. - Oui, dis-je encourageant. Et que pensez-vous que ce soit ? - Pas facile à dire. (...) Ça manque de la simple symétrie des corps platoniques, bien que ça puisse avoir une symétrie propre... Je pense que ce pourrait être une inflorescence ou une fleur. - Une fleur, vraiment ? demandais-je. - Oui, confirma-t-il. - Sentez-la", lui suggérerais-je, et de nouveau il prit l'air intrigué, comme si je lui avait demandé de sentir une symétrie supérieure. Mais il s'y soumit avec courtoisie et porta la rose à son nez. Soudain, il s'anima. "C'est beau ! s'exclama-t-il. Une rose précoce. Quelle odeur divine ! " Il commença à fredonner : "Die Rose, die Lillie..." On aurait dit qu'il percevait la réalité de la rose par l'odorat et non par la vue."

Il semble donc que sans la possibilité d'associer l'idée d'un objet à nos sensations, nous serions comme dans un monde privé de sens et peuplé de formes ayant des traits bizarres. Pourtant l'attitude inverse, qui ne verrait dans nos perceptions que des indices d'objets semble tout aussi handicapante :

"Karl Pribram, venu donner une conférence à Paris, (...) réclame un volontaire pour faire une démonstration, je me précipite. Il me demande de tendre la main ouverte et de fermer les yeux.. Il pose quelque chose sur ma paume : "que sentez-vous ?" Sure de moi, je réponds : "un crayon". Il pose une autre chose. "Que sentez-vous ?" Fière d'identifier si facilement ce qu'il a mis dans ma main, je réplique : "une pièce de monnaie". Il a l'air déçu et me renvoie à ma place sans autre forme de procès. Un autre volontaire se présente et l'exercice se répète. Pribram sort une clé de sa poche, l'applique sur la paume du volontaire. "Que sentez-vous ? - une clé - Je vois qu'on est en France, je n'arriverai pas à faire ma démonstration. J'ai remarqué que les français ne savent pas sentir. Je vous pose des objets dans la main et je vous demande ce que vous ressentez. Et plutôt que de me répondre : un contact froid, dur... vous identifiez l'objet. Vous faites ça pour paraître intelligents, mais vous êtes coupés de vos sensations." (Filliozat, 1997)

Il semble donc que ces deux attitudes de perception : perception d'objets vs de ses qualités n'aient pas du tout le même "sens", ni le même type d'implication dans notre vie. Le fait de reconnaître un objet permet d'accéder à son "sens" lié à la manière dont on inter-agit avec lui dans la vie quotidienne, et est donc nécessaire pour vivre dans un monde d'objet "pragmatique". Le fait de percevoir les qualités des stimuli qui nous entourent nous garantit un "rapport" sensible à notre environnement. En effet, la plupart des arts sont dus à des modulations de ces qualités. Par exemple, la musique naît de variations subtiles de sons produits par la même source, la peinture est un jeu de couleurs abstraites de leur source initiale et fixées sur des pigments, le langage est possible grâce aux modulations de timbre et de hauteur de notre voix...

### I.3. Dans le domaine sonore

Nous avons déjà lié lors de l'introduction les deux définitions du timbre à deux attitudes d'écoute. Ces deux attitudes relèvent en fait des deux positionnements définis ci-dessus. "En effet, si nous prêtons un peu d'attention à la manière dont nous écoutons les bruits, nous constatons bien vite deux modalités : tantôt nous identifions la source et tantôt nous analysons la qualité du son" (Guyot et al., 1998).

#### **Ecouter, ouïr, entendre, comprendre**

Pierre Schaeffer (1966) définit plutôt quatre attitudes d'écoute, liées aux 4 verbes écouter, ouïr, entendre et comprendre. "*Ecouter*, c'est prêter l'oreille, s'intéresser à. Je me dirige activement vers quelqu'un ou quelque chose qui m'est décrit ou signalé par un son. *Ouïr*, c'est percevoir par l'oreille. Par opposition à écouter qui correspond à l'attitude la plus active, ce que j'ouïs, c'est ce qui m'est donné dans la perception. *D'entendre*, nous retiendrons le sens étymologique : 'avoir une intention'. Ce que j'entends, ce qui m'est manifeste, est fonction de cette intention. *Comprendre*, prendre avec soi, est dans une double relation avec écouter et entendre. Je comprends ce que je visais dans mon écoute, grâce à ce que j'ai choisi d'entendre. Mais, réciproquement, ce que j'ai déjà compris dirige mon écoute, informe ce que j'entends." Une façon de récapituler ces quatre types d'écoute serait contenu dans la phrase suivante : "Je vous ai ouïs malgré moi, sans que j'ai écouté à la porte, mais je

n'ai pas compris ce que j'entendais" ou "j'écoute un moteur, j'ouïs un bruit, j'entends un basson, je comprends un accord parfait." Ainsi ces quatre activités semblent avoir quatre vis-à-vis : la cause, la chose, la qualification, le sens. Une fois choisi l'objet privilégié (j'écoute), je tends l'oreille (j'ouïs), j'apprécie ce que j'entends, et je me réfère à ce que je sais déjà (je comprends).

### **Identification - Qualification**

Nous avons déjà souligné la tendance à identifier les sons que l'on entend, c'est-à-dire de se focaliser sur leur origine (source sonore). On retrouve cette attitude aussi au niveau du nom donné à chacune des notes. Ainsi, si on est formé au solfège, "dès que nous entendons les notes d'une mélodie, notre conditionnement nous les fait nommer" (Schaeffer, 1966). Le mécanisme qui induit cet étiquetage des sons en noms de note étant appelé "oreille absolue" par comparaison à une "oreille relative" qui entendrait d'abord les intervalles entre les sons. Toute personne, musicienne ou non, procéderait de l'une ou l'autre attitude. Ainsi, pour pouvoir apprécier réellement la qualité d'un son, il faudrait pouvoir s'abstraire de son origine, c'est ce que propose Schaeffer avec une "collection d'objets dépourvus de sens musical" afin de les entendre mieux (en tant que tels et non en référence à un objet) : "l'un plus *rauque*, l'autre plus *strident*, les uns plus *brefs*, les autres interminables, les uns *claironnants*, les autres *râpeux*." Il définit ainsi, au contraire de l'"écoute naturelle" (identification) une écoute musicienne, ou "écoute réduite" (qualification) : "on n'écoute plus le son pour l'événement, mais l'événement sonore lui-même" (ibid.).

On retrouve d'une part la priorité à la dénomination de sources, et d'autre part l'existence de qualifications de ces sources, soit les deux types d'écoutes menant à l'identification ou à la qualification, dans une étude linguistique sur des définitions d'odeurs, de sons et de bruits donnés par des sujets (David, 1997) : "Pour l'audition, le français possède des dénominations lexicalisées, se rattachant d'ailleurs plutôt au bruit qu'au son. Par exemple : *vacarme*, *brouhaha*, etc. On a aussi des noms déverbaux (construits sur une base verbale), pour la plupart suffixés en *-ment* (on ne trouve que deux noms suffixés en *-is*, *gazouillis* et *cliquetis* ; ils sont employés chacun une seule fois), tels que *crépitement*, *grésillement*, *claquement*, *grondement*, etc. Cependant, toutes ces formes sont (...) moins utilisées que les dénominations construites sur la source du bruit, telles que *bruit de la vie urbaine*, *bruit de vaisselle*, *aspirateur*, *télévision*, etc. (...) Les sujets privilégient les dénominations de source ou les dénominations construites sur des adjectifs renvoyant à des propriétés physiques, tels que *aigu*, *grave*, etc." David retrouve donc l'hypothèse de Guyot, Castellengo et Fabre (1998) selon laquelle l'usage des noms déverbaux serait lié à des situations dans lesquelles la source n'est pas ou ne peut être identifiée. Pourtant, les noms déverbaux sont construits sur des verbes d'action permettant de produire les sons considérés, donc d'une certaine façon à la source de ceux-ci. On peut donc aussi voir dans cette forme de description une identification de la source du son.

### **Chez l'enfant**

On peut ici aussi se poser la question de l'origine de ce comportement dans l'enfance. Ainsi le son est considéré comme l'attribut d'un objet dès 4 mois, (Dumaurier, 1982). On observe dès cet âge une coordination vision - audition (l'enfant reconnaissant l'image d'un objet correspondant à un son

entendu par exemple) où le son est considéré comme l'attribut d'un objet donné. L'association entre une voix et un visage étant observée dès 30 jours (ibid. p 37). Le timbre semble déjà perçu comme trait d'un objet pour la localisation (l'enfant se tourne vers la voix) et l'identification d'une source sonore (ibid. p 36).

Quand aux productions sonores des enfants, il semble que dès les premiers jours, ils modulent leur voix en fonction des auditeurs présents (Cyrulnik, 1995) : la composition en harmoniques est différente s'ils ne sont entourés que d'autres bébés ou si des adultes sont présents et parlent autour d'eux, et ce dès le quatrième jour de vie. Ainsi, "dans un milieu stable, leurs cris s'enrichissent en basses fréquences. Mais si quelqu'un vient à entrer, ou si une infirmière se présente pour changer le lit, dans le cri suivant les hautes fréquences seront plus nombreuses!" (ibid.). La présence de hautes ou de basses fréquences dans ces cris engendre d'ailleurs des réactions très différentes chez des auditeurs adultes, voir même chez des animaux : les cris comportant beaucoup de haute fréquences sont perçus comme transmettant un malaise ("les chiens se mettent à gémir, les chats à miauler d'inquiétude" (...) les adultes ressentent des angoisses : "je me sens mal, ces cris me donnent des crampes d'estomac") alors qu'on obtient des commentaires "très neutres, très détachés lorsque les mêmes personnes avaient écouté les cris en basse fréquence : 'tiens ! c'est un garçon... il doit avoir faim ! ça me rappelle mon fils'" (ibid.). Il semble donc que les enfants très petits savent déjà moduler la qualité du timbre de leur voix pour la rendre porteuse de sens, qui semble d'ailleurs compris.

Si on considère les productions sonores d'enfants plus grands, utilisant des "instruments" (cuillère par exemple !), il semble que "la production musicale répond à trois motivations fondamentales : 1- jouissance de la motricité productrice du son, 2- dépassement de cette origine matérielle par une valeur symbolique, 3- plaisir de soumettre l'organisation formelle à une règle du jeu" (Delalande, 1982). Ceci semble donc lié aux trois formes successives que prend le jeu chez l'enfant d'après Piaget (1945) : jeu sensori-moteur, jeu symbolique et jeu de règles. Ainsi, est soulignée l'idée de la primauté d'une fusion entre motricité et perception dans la petite enfance (Céleste et al., 1982).

Schaeffer (1966) décrit pour sa part un enfant plus grand expérimentant les sons qu'il peut produire en soufflant dans une herbe : "l'enfant expérimente ses sons les uns après les autres, et le problème qu'il se pose est moins celui de l'identification que celui du style de fabrication. D'autre part, son intention est visiblement musicienne. Si le résultat ne paraît pas musical à ses auditeurs exaspérés, on ne saurait contester à l'auteur une intention esthétique (...). Non content de façonner des sons, il en joue, il les compare, il les juge (...) s'il ne fait pas de musique, que fait-il donc ?" C'est ainsi qu'on peut s'abstraire de l'écoute naturelle qui renvoie à la source sonore. "A peine en effet a-t-il identifié la causalité herbe, que les essais successifs du virtuose ne lui apprennent plus rien, ni sur une anecdote d'herbe, ni sur un improbable morceau de musique herbe. (...) on n'écoute plus le son pour l'événement, mais l'événement sonore lui-même." (ibid.)

L'enfant, dans son rapport au monde, semble donc utiliser très tôt le son pour identifier et localiser les personnes et les objets qui l'entourent, tout en modulant les sons qu'il produit pour transmettre par exemple son état de stress. Dans son expérimentation de jeux sonores, il semble pratiquer à la fois une observation des liens entre motricité et production sonore et une écoute de type musicienne.

### **Position des compositeurs contemporains**

On retrouve ici aussi, ce partage entre "abstrait et concret" dans le rapport entre le son et les choses (Risset, 1994). Si le son renvoie à la source, les sons organisés dans un "discours" musical sont sensés transmettre autre chose. Ainsi, pour Chion & Reibel (1976) : "la musique instrumentale peut être dite abstraite parce que conçue et notée préalablement dans l'abstrait avant d'être réalisée concrètement". De même, "Schaeffer attribue un rôle décisif à l'incident du sillon fermé, répétant inlassablement le même fragment sonore, dans sa prise de conscience du caractère étrange et potentiellement musical d'un son arraché de son contexte." (Risset, 1994), ou comme le dit Michel Chion (1982), "*L'Etude aux chemins de fer* est une 'Etude de rythmes' et pas une 'nature morte au train' ; une oeuvre abstraite plutôt qu'une évocation."

Ainsi, pour Risset (1994) les recherches sur l'objet sonore "s'inspirent de la démarche phénoménologique : elle prône l'écoute réduite, détachée des références à l'origine physique du son et à son sens. Pour que le son soit apprécié pour lui-même et non pas comme une étiquette, un renvoi à un corps sonore particulier", ceci pouvant aboutir à "un solfège des effets et non des causes" (ibid.).

Pourtant, au début de la mise en pratique de la synthèse de sons complexes à partir de sons simples dans le courant de musique électronique : "Les sons électroniques étaient ternes, tristes - mortuaires, disait Varèse. (...) Stables, mornes, il était difficile de leur conférer dynamique et identité. (...) Ce n'est qu'en abdiquant leur volonté de contrôle total des paramètres sonores que les musiciens sont parvenus à enrichir, à musicaliser le son électronique." (Risset, 1994).

En effet, alors que l'idée de synthèse laissait entrevoir la possibilité de créer des sons détachés de toute production matérielle, ces sons renvoyaient.. à leur source électronique ! "Le timbre d'un instrument électronique renvoie à la nature artificielle du son". (Schaeffer, 1966) "Les résultats sonores surprenaient (...) parce qu'un certain nombre d'effets, vite reconnaissables, révélaient à coup sûr leur origine 'électronique' (...) Cette source électronique apparaissait comme un instrument parmi d'autres. Alors qu'elle prétendait, soit reconstituer des timbres préexistants, soit créer des timbres inouïs convenablement variés, elle marquait les uns et les autres de son 'timbre' propre, au sens pragmatique où nous avons défini ce terme." (ibid.)

On pourrait donc conclure, comme Schaeffer (1966) : "le phénomène musical a donc deux aspects corrélatifs ; une tendance à l'abstraction, dans la mesure où le jeu dégage des structures ; l'adhérence au concret, dans la mesure où il reste attaché aux possibilités instrumentales". Ceci pourrait expliquer deux tendances musicales : l'orchestration, qui renvoie aux événements, versus Klangfarbenmelodie (c'est-à-dire "mélodie de timbres"), qui est une tentative de réincorporer le timbre des notes, en faire une valeur.

Enfin, on ne peut oublier que même si les systèmes musicaux construisant un langage uniquement sur des modulations de timbre ne sont pas jugés comme complètement convaincants (Risset, 1994), il existe un système de communication fondé sur la modulation temporelle du timbre, que nous utilisons quotidiennement : les langues parlées ! Dans la succession des voyelles et des consonnes, on s'intéresse à leur dimension porteuse de sens, c'est-à-dire à leur qualité et non pas uniquement au renvoi à un objet : l'identification du locuteur.

## I.4. Résumé

Après avoir introduit quelques notions acoustiques décrivant le timbre, nous avons repris la question du rapport entre perception d'objet et perception de qualité, en remarquant que dans le domaine sonore les deux étaient présentes. Le timbre renvoie à l'objet qui l'a produit et ce sont des modulations de qualités sonores qui permettent la musique.

## **II. Catégorisation, reconnaissance, identification**

Ayant abordé dans le chapitre 1 les arguments expliquant la nécessité de catégoriser le monde, nous ne reprendrons ici qu'un résumé succinct des différentes théories de la catégorisation, ainsi que les applications dans le domaine du sonore : comment sont catégorisés, reconnus ou identifiés les sons, bruits et timbres. Il est nécessaire de rappeler ici que ces trois "choses" correspondent déjà à trois niveaux de catégorisation différents. Les bruits sont des sons, et comme les sons complexes harmoniques et inharmoniques, ils ont des timbres.

### II.1. Catégorisation

Nous allons donc présenter ici trois modèles successifs de catégorisation : par Conditions Nécessaires et Suffisantes (CNS), par ressemblance à un prototype (qu'il soit considéré comme une abstraction ou une forme de résumé des traits des exemplaires les plus fréquents) et par air de famille. Puis nous présenterons l'idée de perception catégorielle et son application dans le domaine du timbre.

#### **CNS, prototypes et air de famille**

Dans Catégories sémantiques et historicité des significations, Nyckees (1997) propose un résumé des différentes théories de construction de catégories perceptives, sémantiques et langagières : "Les catégories sémantiques peuvent être différentes des catégories psychologiques : la catégorie perceptive ne correspond pas forcément à celles utilisées dans le langage", et ce, bien que pour lui, il n'y ait pas de catégories perceptives pures.

Ainsi, dans les années 1970 (Rosch et al., 1976), la catégorisation repose sur la base du partage de traits à travers le modèle dit des Conditions Nécessaires et Suffisantes (CNS), qui considère les traits partagés par les exemplaires de la catégorie. "Bruner a ainsi fondé sur cette hypothèse le classement de figures géométriques selon leurs traits relatifs à leur forme, leur couleur, leur taille" (Nyckees, 1997). "Cette première version de Roch (...) était clairement réaliste : les catégories étant corrélées à des regroupements d'attributs dans la réalité." (ibid.)

La seconde génération de la psychologie cognitive a considéré la catégorisation comme la saisie d'une forme globale en opposition à une démarche analytique. "La version standard du prototype s'inscrit dans le cadre du conceptualisme (...) L'objectif de la catégorisation étant considéré comme un moyen



de maximiser l'information des riches variations du réel. (...) Dans la généralisation de cette idée, nous ne percevons jamais que les particuliers (en opposition à la conception d'Idée Générale) ce qui pose la question de comment avoir une image mentale d'un triangle qui ne soit pas le reflet d'un exemplaire particulier" (ibid.). Ainsi par exemple, dans le cadre de la théorie des CNS, un oiseau serait considéré comme tel après avoir par exemple vérifié qu'il ait des ailes, un bec, qu'il ponde des oeufs... (avec tous les problèmes rencontrés pour définir un ensemble de traits qui soient réellement des conditions nécessaires et suffisantes pour que "la chose" soit un oiseau... par exemple les tortues ont un bec, les serpents pondent des oeufs, l'autruche ne vole pas, peut-on considérer que le pingouin a des plumes ?...). Dans la théorie du prototype par contre, est considéré comme oiseau, toute chose se rapprochant plus où moins du "prototype" d'un oiseau, par exemple un moineau pour les occidentaux (ou un pigeon pour les Parisiens !). Ce type de théorie permet notamment de prendre en compte les mesures montrant que les gens mettent plus de temps à décider si une autruche est un oiseau que pour décider qu'un canari est un oiseau (Collins & Quilian, 1969). Il y aurait donc des oiseaux plus "oiseau" que les autres...

Enfin, la dernière version de la théorie étendue du prototype est appelée : modèle de ressemblance de famille. "C'est une sorte de version eucuménique : il y a des CNS conservées comme possibilité. (...) Il faut et il suffit que chaque membre de la catégorie partage au moins un trait avec un autre membre de la catégorie. Il n'est plus considéré la référence à un prototype qui serait central à la catégorie. On peut donc avoir A ressemble à B et B ressemble à C sans que A ne ressemble à C." (Nyckees, 1997).

Il faut garder à l'esprit que ces catégories sont essentiellement formées à partir de considérations humaines : "Tout semble indiquer que les sciences du vivant ont hérité leurs catégories fondamentales d'une problématique fonctionnelle issue de la chasse, de la cueillette, de l'élevage et de l'agriculture." Et "Les significations mises en oeuvre dans les langues (...) sont nécessairement relatives aux déterminations qui régissent notre espèce en même temps qu'aux traditions culturelles et linguistiques qui organisent les divers groupes humains." (ibid.)

### **Drôles d'oiseaux**

Je ne résiste pas au plaisir de retranscrire ici une anecdote familiale entrant dans le débat prototype contre traits définitoires. De retour de l'école de sémantique des langues d'Asnelle, Chloé (3 ans) retrouve ses jeux et exhume une espèce d'oiseau en plastique ayant pour fonction de cracher de la pâte à modeler par le bec grâce à une vis qui lui sert de pieds... (Il se présente donc debout sur ses pattes de derrières qui sont jointes... ). Moi : - "Tiens, tu as retrouvé l'oiseau ?" Chloé - "C'est pas un oiseau, c'est un pingouin !" Moi, sautant sur l'occasion - "Un pingouin, c'est pas un oiseau ?" Chloé, après réflexion - "Si, ça vole... ça marche et ça vole." Moi, regardant l'engin et l'air sans doute songeur : "ah bon, ça marche et ça vole..." Chloé, presque condescendante : "Non, ça, ça ne vole pas, c'est un truc à pâte à modeler !"

Pour elle, l'appartenance à la catégorie se faisait donc en terme de traits définitoires. Je retransmets cette conversation à son père le soir qui me dit : "C'est évident, pour elle, un oiseau, c'est un pigeon (en vraie petite parisienne) et un pigeon, ça marche et sa vole..." Le (bon) sens commun pense en

terme de prototype (sans doute formé sur des fréquences d'exemplaires : le pigeon étant rencontré plus souvent contribuant à plus de traits définissant la catégorie oiseau).

## II.2. Reconnaissance et identification

De la même façon, plusieurs approches sont possibles en matière de théorie sur la reconnaissance des sources et événements sonores. Dans une perspective de théorie de l'information (McAdams, 1993), on peut ainsi concevoir plusieurs étapes de traitement dans la reconnaissance auditive : le son donnerait lieu tout d'abord à une représentation auditive périphérique du signal acoustique, puis des processus auditifs de groupement permettraient de former des "événements". Puis interviendraient des analyses des propriétés et traits pertinents de l'activité auditive (invariants structuraux : nature de l'objet, et invariants transformationnels : nature des modifications subies par l'objet). Des ajustements des propriétés auditives aux représentations mnésiques permettraient de sélectionner la représentation la plus proche en mémoire puis une activation du lexique verbal et des structures sémantiques associées permettraient de rendre compte de cette reconnaissance par l'identification proprement dite. Il est néanmoins évident que cette vision des choses est trop simplifiée, étant donné que de nombreuses interactions entre ces étapes du traitement ont déjà été mises en évidence (McAdams, 1993).

On peut trouver une autre façon de concevoir l'identification et la reconnaissance, dans le cadre de la perception directe. Les invariants structuraux et transformationnels seraient directement extraits de l'environnement et permettraient d'identifier directement les "objets".

### **Bruits**

On trouve ainsi des études sur l'identification des bruits définies dans un cadre théorique basé sur un schéma de perception qui suppose que l'identification d'un événement sonore a lieu grâce à l'existence physique d'invariants de structure de transformation. Ainsi, Van Derveer (1979) a étudié l'identification libre et la classification d'un ensemble de bruits divers (cartes battues, cliquetis de clés, cognements ou bruits de scie...). "Les sons classés dans les mêmes catégories présentaient des patterns temporels et/ou une configuration spectrale comparables" (Donnadieu, 1997).

Depuis, Warren & Verbrugge (1984) ont étudié l'identification de sons de bocal qui rebondissent ou se cassent, Repp (1987) s'est intéressé à la possibilité d'identifier une personne à travers ses applaudissements (notamment son sexe et la position de ses mains), Li et al. (1991) ont repris cette idée à partir de bruits de pas. Enfin, Ballas (1993) a montré que les bruits impulsifs, comme les pas sont difficilement identifiables hors contexte au contraire des bruits d'alarmes, et ce, en dépit de fréquences "écologiques" très différentes dans l'environnement habituel (Guyot et al., 1998). Ces résultats rejoignent peut-être l'importance et la force émotionnelle des hautes fréquences dans les sons d'alerte produits dès la naissance (Cyrulnik, 1989).

### **Influence du label sur la mémorisation**

Dans la lignée d'une longue tradition de recherches sur l'influence du codage verbal sur la mémorisation de matériel non-verbal (on peut trouver une revue de ces travaux dans Ellis, 1973), Bower & Holyoak (1973) ont étudié ce phénomène sur la reconnaissance de 80 sons. Les résultats montrent que la reconnaissance du son semble conditionnée par la reconnaissance du label. Elle est donc en partie due à son interprétation. Dans le même ordre d'idées, Light & Carter-Sobel (1970) ont montré qu'on se souvient d'une séquence de sons non-musicaux en se souvenant du scénario d'images visuelles des objets produisant les sons.

### **Instruments de musique**

En ce qui concerne l'identification d'instruments de musique, il semble qu'on ait remarqué depuis longtemps (Stumpf, 1883 u. 1890) que le fait de supprimer l'attaque du son, c'est-à-dire les quelques millisecondes de transitoires présentes au début du son, rend l'identification difficile. Il semble pourtant que cette notion de "moment d'attaque" soit plus complexe. Ainsi Risset (1994) déclare que "l'attaque subjective peut correspondre à l'intégration par l'oreille d'un fragment postérieur à l'attaque physique, mais perçu comme attaque : Schaeffer (1966) a montré, dans le cas du piano grave, qu'on modifiait moins l'attaque en la coupant qu'en coupant un fragment ultérieur du son. L'importance de l'enveloppe dynamique sur l'identité de l'instrument - le timbre - a été démontrée par des expériences de 'transmutation' assez probantes." Ainsi par exemple on peut écouter comment un son de piano peut être transformé en son d'harmonium si l'enveloppe temporelle de chaque son est inversée (Bregman & Ahad, 1995). Schaeffer (1966) propose une "loi du piano" pouvant expliquer que l'identité est maintenue au travers des variations de registre : "Raideur dynamique \* Richesse harmonique = constante". De même, Risset (1965) a montré que la "signature" du timbre cuivré de la trompette était non pas un spectre de Fourier caractéristique, mais une relation entre paramètres : le fait que le spectre s'enrichisse dynamiquement avec l'intensité. Il semble d'ailleurs que dans la plupart des cas l'identification se fait grâce à des détails (crissement de l'archet...)

Pourtant, les pourcentages d'identifications correctes observés en conditions expérimentales sont en général faibles, même si les sujets sont musiciens, et même si les sons isolés sont présentés sans attaque coupée. Ainsi, Berger (1964) présente les enregistrements de sons de 10 instruments à vent (flûte, hautbois, clarinette en Sib, saxophone ténor, saxophone alto, cornet, trompette, cor, trombone et baryton) selon 4 modalités différentes : son réels entiers, son sans attaque ni fin, à l'envers, à travers un filtre passe bas à des musiciens membres d'un Jazz Band. Il obtient seulement 59% d'identifications correctes (82% si on accepte les familles : trombone - baryton - cor, cornet - trompette et saxophone - clarinette). Se pose la question du registre normal de l'instrument, les sons étant égalisés en hauteur. De même, Saldanha & Corso (1964) montrent à partir de 10 sons d'instruments variés (violon, violoncelle, flûte, hautbois, basson, clarinette, saxophone alto, cor trompette et trombone), et 5 types de sons (entiers, raccourcis par le milieu, sans attaque, sans fin, partie intermédiaire seule) que des musiciens entraînés n'identifient correctement que 40% de ces sons (ils avaient à leur disposition une liste de 39 noms d'instruments). Par contre, la matrice de confusion est ici aussi intéressante car elle présente un organisation des sons par classe d'instruments (cordes,

bois à vent, cuivres, percussion). On retrouve l'idée que la famille d'instrument est plus facile à reconnaître que l'instrument lui-même quand les sons sont égalisés en hauteur. Cela peut se comprendre dans la mesure où, par exemple les sons de cor, trompette et trombone sont émis de la même manière dans le même type de matériau. On peut avancer l'hypothèse qu'à part pour des experts familiers avec chacun des timbres de ces instruments, le fait de les entendre produire la même note enlève les indices de tessiture (le trombone étant plus grave que la trompette par exemple) et rend donc leur identification plus difficile.

Wedin & Goude (1972) font écouter neuf sons réels et naturels présentés avec et sans attaque, par paires. Ils demandent d'une part une identification des instruments de façon libre et d'autre part des jugements de similarité sur une échelle de 0 à 10. Enfin, pour avoir une mesure de la structure cognitive de ces sons, ils demandent une évaluation des similarités entre les instruments dont on ne présente que le nom. L'espace perceptif construit à partir des jugements de dissemblance présente trois dimensions perceptives : I (violon, violoncelle, clarinette), II (trombone, cor anglais, flûte), III (trompette, hautbois, basson). La structure cognitive contient aussi trois classes, mais pas les mêmes : les bois, les cuivres et les cordes. Les confusions entre les noms des instruments sont reliées à la distance entre ces deux instruments dans l'espace perceptif. Ils notent aussi une influence de l'occurrence de cet instrument dans la vie de tous les jours. De plus, les seuls instruments qui sont reconnus aussi facilement sans les transitoires d'attaque sont ceux présentés avec vibrato.

D'autre part, une étude effectuée afin de construire une norme associative sur les noms d'instruments de musique (Dubois, 1986) donne les instruments les plus cités et les propriétés associées en fonction de la "typicalité" de l'instrument. On observe notamment plus de descriptions des qualités sonores dans le cas des instruments moins typiques.

### II.3. Perception catégorielle

Les psychologues, après s'être intéressés à la catégorisation des "objets" du monde (avec des nuances suivant que ces objets étaient "naturels" ou "manufacturés"...), ont mis en évidence un phénomène qu'ils ont appelé "perception catégorielle", et qui consiste à comprendre comment un continuum sensoriel est "découpé" en deux ou plusieurs catégories distinctes. En effet pour que ce soit possible, il faut qu'il se passe des choses différentes à l'intérieur des 2 catégories et à la frontière entre les deux. Le phénomène de perception catégorielle est donc lié en théorie à deux choses co-occurentes (Petitot-Cocorda, 1985) : un brusque changement d'identification au niveau de la frontière située sur le continuum du stimulus lié à un pic de sensibilité dans la perception des différences entre stimuli au niveau de cette frontière (alors que 2 stimuli d'un même côté de la frontière ne devraient pas être perçus comme différents). On trouve une bonne revue de l'ensemble de ces questions dans Categorical perception de Harnad (1987), et de ce qui se réfère plus spécifiquement à l'audition dans Listening de Handel (1989).

En effet, ce phénomène a été tout d'abord étudié à propos de la perception des différentes consonnes sur le continuum formé par /b/ /d/ /g/ (Lieberman et al., 1957). Ont suivi un grand nombre d'études visant à décider si une perception était ou non catégorielle en fonction des deux critères présentés ci-

dessus. Cet engouement était aussi lié au fait que des mathématiciens ont essayé de modéliser ces brusques transitions avec des modèles issus de la théorie des catastrophes (Petitot-Cocorda, 1985). L'idée de perception catégorielle a été aussi testée sur des sons imitant des sons de parole (Miller et al., 1976), la perception de la hauteur (Siegel & W., 1977), celle des intervalles (Burns & Ward, 1978), et chez les enfants de 2 mois (Jusczyk et al., 1977). Pourtant, dès 1971, ce processus apparaît comme spécifique à la perception de la parole et ne semble pas exister pour les autres sons (Mattingly et al., 1971).

#### **Le timbre : temps d'attaque**

Néanmoins, en 1974, Cutting & Rosner testent l'éventualité d'une perception catégorielle du temps d'attaque de timbres. (On peut trouver un bon résumé de la problématique dans Donnadieu, (1997)). En effet, il semble que l'identification des sons par leur timbre soit liée à leur temps d'attaque, qui permet notamment de distinguer les sons pincés des sons frottés (Cutting & Rosner, 1974). Une série d'études tend donc à montrer que la perception du temps d'attaque des sons serait catégorielle comme les consonnes (Cutting & Rosner, 1976; Cutting et al., 1976). Puis Remez et al. (1980) étudient l'adaptation des frontières de catégorisation qui semble se faire différemment pour la parole et les sons de violon. Ces études nient donc la spécificité de la perception de la parole et la replacent dans le cadre général de la perception auditive. Deux théories s'affrontent alors : une liée à des détecteurs de traits dans le signal, l'autre liée à la capacité de produire les intermédiaires du continuum (qui sont toujours fabriqués artificiellement lors des expérimentations). La distinction pincé / frotté (avec un archet) viendrait-elle de la connaissance implicite qu'on ne peut *produire* effectivement un son entre les deux sur un violon ?

#### **Controverse**

On s'achemine donc vers une reconnaissance générale de ce phénomène jusqu'à l'intervention de Rosen & Howell (1981), qui déclarent que les résultats obtenus par Cutting & Rosner (1974) sont dus à un artefact de construction des stimuli : les temps d'attaques ne sont pas répartis linéairement de façon rigoureuse. Les résultats peuvent s'expliquer par une meilleure discrimination dans les temps d'attaque quand ils sont courts que quand ils sont longs (cf. loi de Weber). S'ensuit une controverse : Cutting (1982) remercie Rosen et al. de lui avoir fait remarquer que ses stimuli étaient faux mais maintient que la perception est catégorielle si on considère des incréments logarithmiques de temps d'attaque. Rosen & Howell (1987) tentent alors trois types d'explications de ce phénomène : avec une théorie motrice (analogue de la prononciation pour la parole avec ici la référence aux modes de productions possibles), une théorie nécessitant des détecteurs de traits auditifs et une théorie qui lie la catégorisation perceptive à l'existence de mots pour désigner les catégories. Enfin, l'affaire semble close par Hary & Massaro (1982) qui montrent que des résultats de type catégoriels (aboutissant à des catégories) n'impliquent pas forcément le mécanisme désigné par "perception catégorielle". Néanmoins d'autres études ont continué dans ce sens, notamment Saldanha & Rosenblum (1993) qui mettent en évidence l'analogie d'un effet McGurk sur la perception de sons pincés ou frottés : le

cerveau "entend" un stimulus entre celui vu et celui entendu. Ils soutiennent une théorie écologique montrant une relation entre le son et son mode de production.

Enfin, Donnadieu (1997), en reprenant de nouveau un continuum de temps d'attaque entre des sons de vibraphone frotté et vibraphone frappé, a montré que les temps d'attaques n'étaient pas perçus de manière catégorielle. En effet, la discrimination entre deux sons était toujours aussi bonne, tout le long du continuum, même si les sujets pouvaient catégoriser les deux types de sons de manière claire lors de la phase d'identification.

#### II.4. Résumé

Après avoir rappelé les principales théories de la catégorisation en psychologie et leur exemple prototypique de la catégorie des oiseaux, nous avons présenté les principales études sur l'identification des bruits et des sons d'instruments de musique. Celle-ci semble particulièrement sensible au début des sons. Nous avons alors présenté brièvement les études ayant testé une éventuelle perception catégorielle de l'attaque des sons complexes.

### **III. Les verbalisations en psychologie**

Nous abordons enfin l'ensemble des études qui ont essayé de décrire nos perceptions à l'aide du langage. Nous présenterons tout d'abord deux types de méthodes de recueil de verbalisations : libres ou dirigées sous forme d'échelles sémantiques ou de listes de mots à cocher. Puis nous présenterons les études concernant les modalités visuelles, tactiles et l'odorat, avant d'aborder celles qui analysent la façon dont on parle de la musique en générale, puis des sons, des bruits et enfin du timbre.

#### III.1. Méthodes extérieures / introspection

La psychologie expérimentale se veut une démarche scientifique tout en étudiant les facultés supérieures de l'homme. Elle a donc essayé de construire des modes d'observations extérieures (en troisième personne) pour décrire les différentes activités humaines. Néanmoins, comme le sujet humain est à la fois l'observateur et l'observé, on a très tôt demandé à ce sujet humain ce qu'il pensait de ce qu'il faisait, voire comment il réalisait une tâche (observations en première personne). Même si le statut de l'introspection, a souvent fait l'objet de violentes controverses, elle est souvent à l'origine des recherches en psychologie : comment avoir l'idée de chercher dans une voie si on n'a pas une petite idée de ce qu'on cherche ? De plus, la psychologie a hérité de questions traitées antérieurement par la philosophie (notamment le langage, la perception...), dont la méthode se rapproche plus d'une réflexion en première personne que d'une observation méthodique en troisième personne...

#### **Introspection**

Droit (1998) considère que c'est avec Saint Augustin que naît le fait de témoigner par écrit de sa subjectivité. "Les confessions inaugurent une nouveauté littéraire : l'autobiographie spirituelle. Elles innovent aussi dans le domaine psychologique, en inventant l'introspection, que les Grecs ignoraient. L'apparition de l'intériorité modifie les relations du sujet au temps, à la mémoire, au désir, à la culpabilité. Nietzsche, venant de lire l'ouvrage, n'a pas tord d'écrire : "avec ce livre, on arrive à voir l'intérieur de l'estomac de la chrétienté". Richard (1992) quand à lui attribue à Descartes, l'orientation donnant le *prima* à l'intériorité, de part son affirmation que "tout ce qui se passe en nous a un retentissement dans la conscience".

On peut en effet définir l'introspection comme "la connaissance de soi par soi : elle consiste dans la connaissance intérieure que nous avons de nos perceptions, actions, émotions, connaissance qui est différente de celle que peut avoir un spectateur étranger. (...) La psychologie introspective pose que l'introspection a non seulement valeur de témoignage mais est la seule méthode qui permette de saisir la vie psychologique : elle affirme que son authenticité ne peut être mise en doute." (ibid.)

Ce type de position repose sur deux hypothèses sous-jacentes : 1- la psychologie est définie comme l'étude de la conscience (on ne peut avoir accès par l'introspection qu'à ce qui est conscient) ; 2- la conscience est claire et transparente à elle-même (le *je* qui observe et le *moi* qui est observé ne font qu'un). C'est ainsi que la psychologie, à la fin du XIXème et au début du XXème siècle repose sur l'évidence de l'auto-observation, notamment dans la lignée de James, Brentano et Wundt.

### **Contre l'introspection**

C'est bien évidemment contre ces deux hypothèses que se sont construits les courants psychologiques opposés. Parmi les précurseurs, on peut citer l'empirisme de Hume qui repose sur des méthodes objectives telles que l'observation comportementale et Kant, qui "développe une psychologie réflexive, dont l'idée principale est que le *je* ne s'atteint pas directement mais s'atteint seulement à partir de ses oeuvres." (ibid.) Mais le courant qui s'est opposé directement à la psychologie introspective pratiquée au début du siècle est principalement le Comportementalisme (né entre 1913 et 1919). L'homme y est considéré comme un être réagissant passivement à des excitations. Il s'agit donc de comprendre les liens directs entre stimuli et réponses. La méthode introspective est remplacée par des mesures aussi précises que possible des réactions humaines. En même temps, le domaine d'étude évolue : il ne s'agit plus ici de s'intéresser à la conscience, la volonté, l'âme, voire la nature de l'homme. On observe une émancipation de la métaphysique pour s'orienter vers un pragmatisme "américain" (De la Motte-Haber, 1994). La psychologie est restreinte à l'étude du mode de fonctionnement des facultés psychiques.

### **Sémantique différentielle**

Pourtant, dès les années quarante (Mosier, 1941), des chercheurs avaient déjà essayé de réintégrer des formes de descriptions en langue des expériences psychologiques. Mais ce sont les travaux d'Osgood et al. (Osgood, 1952; Osgood & Suci, 1955; Osgood et al., 1957; Osgood, 1979) qui systématisent cette méthode baptisée sémantique différentielle. Ne demandant aucune production personnelle aux sujets, cette méthode pouvait prétendre à plus d'"objectivité" et ainsi échapper à la comparaison avec

l'introspection. Il s'agit en effet de présenter aux sujets une série d'échelles, par exemples en 10 points, dont les extrémités opposent deux attributs verbaux tels que blanc-noir, chaud-froid. Ceux-ci doivent positionner chaque stimulus sur chacune des échelles proposées. Par une analyse factorielle des données, on cherche ensuite à réduire le nombre de dimensions en éliminant les échelles qui ne contribuent pas ou peu à l'explication de la variance des jugements. Les dimensions retenues peuvent ainsi être considérées comme une description du sens des stimuli ainsi mesurés.

Osgood et al. (1957) s'étaient attachés à proposer une série d'échelles qui permette de considérer la plus grande étendue possible de stimuli différents. Par la suite pourtant, les utilisateurs de cette technique ont formé leur propre ensemble d'échelles, souvent déterminées avant l'expérience par un comité d'experts. Les études sur le son utilisant ces méthodes seront détaillées plus loin.

### **Les verbalisations libres remplacent l'introspection**

Pourtant, chassée par la porte, l'introspection revient par la fenêtre. En effet, le champ de recherche s'intéressant à la résolution de problèmes va recommencer à utiliser comme données ce que les gens disent de ce qu'ils font (Newell & Simon, 1972). Pourtant, afin de s'affranchir des critiques liées à l'introspection, toute une argumentation est mise en place (Ericsson & Simon, 1984; Caverni, 1988; Wilson, 1994) : on analyse des protocoles (écrits à partir des verbalisations des sujets) ; les données sont considérées comme des données fiables si elles ne sont pas modifiées ou induites par la consigne, si elles sont produites pendant la tâche ou juste après mais sans faire appel à la mémoire à long terme.

Pourtant, même si les verbalisations sont de nouveau utilisées, la controverse continue. Nisbett & Wilson (1977) montrent que les sujets disent plus que ce qu'ils peuvent savoir et qu'on peut influencer leur jugement. En fait, ceci semble dû au fait qu'ils posent la question du "pourquoi". En demandant au sujet non pas seulement de décrire ce qu'il fait mais d'expliquer sa conduite, ils touchent en effet à des motifs pas forcément conscients. Crutcher (1994) répond à cet article en précisant à nouveau les conditions méthodologiques nécessaires : "penser à haute voix" ne rapporte que ce qui est en mémoire à court terme. Il n'y a pas de modification de la tâche due à la verbalisation, et les données gardent un caractère d'objectivité si le sujet n'essaye pas d'analyser, d'expliquer ou d'interpréter ses pensées. Cette méthode permet en outre d'obtenir : 1- une séquence d'observations (accès aux stratégies utilisées par le sujet) ; 2- des informations sur le type de connaissances nécessaires pour effectuer la tâche ; 3- un moyen de vérifier ce que fait le sujet par rapport aux instructions.

### **Faut-il trancher ?**

Ces approches sont-elles forcément antagonistes ? En effet, on ne peut nier que "le point commun entre les partisans de l'introspection et les comportementalistes (et qui est la base de l'approche scientifique) est l'expérimentation comparative" (Richard, 1992). Peut-on, comme Cyrulnik (1995) refuser l'alternative qui semble proposée en ces termes : "Choisissez, nous dit-on, l'étude de la chose, l'objectivité qui donne accès à la maîtrise du monde, ou bien choisissez d'écouter l'autre et de le respecter, mais alors ce ne sera pas scientifique". Il semblerait plutôt que le véritable problème



d'utiliser les verbalisations libres comme données est de construire des outils qui permettent de rendre-compte et de comprendre la complexité de ces données.

Une tentative de conserver la richesse des verbalisations libres tout en les catégorisant dans une approche systémique est proposée par Nosulenko & Samoilenko (1997), dans la lignée de toute une série de travaux des psychologues russes (Samoilenko, 1986; Samoilenko, 1987; Nosulenko, 1989; Nosulenko, 1990; Samoilenko, 1991). Selon cette approche, la cognition humaine ne peut être étudiée indépendamment des processus de communication qui lui sont associés. La méthodologie utilise les données verbales pour analyser les processus psychiques et consiste en une analyse du contenu des protocoles verbaux afin d'en inférer les propriétés des représentations des sujets. Elle considère pour cela que "les protocoles verbaux ont un double contenu : 1- un contenu communicationnel, dans la mesure où toute production verbale est orientée vers un destinataire, réel ou imaginaire ; 2- un contenu cognitif qui correspond aux propriétés perçues des objets et des événements décrits" (Nosulenko & Samoilenko, 1997) .

### III.2. Modalités autres qu'auditives

Nous allons tenter une brève présentation des travaux présentant des descriptions de stimuli appartenant aux différentes modalités sensorielles : vision, goût, odorat... avant de nous pencher plus particulièrement sur l'audition.

#### **En vision**

Le domaine visuel a souvent été beaucoup plus étudié que tous les autres. Dans les sciences cognitives, il bénéficie d'une avance par rapport aux autres modalités sensorielles due au fait que des équipes de philosophie, perception, neurosciences, modélisation (etc.) travaillent en synergie depuis plus longtemps que dans les domaines relevant des autres modalités. Ceci engendre parfois le fait que certains problèmes, notamment le lien entre percepts et concepts (Meini, 1996), ou la plupart des études sur la catégorisation des objets, n'ont été souvent considérées que du point de vue de la vision (Dubois, 1998).

Parmi les premiers psychologues à avoir utilisé des protocoles verbaux pour définir les caractéristiques perçues des objets visuels, on trouve Bartlett (1932), qui a analysé les caractéristiques de la perception de dessins simples et complexes. Nous avons déjà cité Loewenthal (1967), qui a étudié comment le type de descriptions de combinaisons de formes et de couleurs différentes dépendait du destinataire supposé de ces descriptions : soi-même immédiatement ou quelques semaines après, ou quelqu'un d'autre.

Les études sur la catégorisation de la couleur (Berlin & Kay, 1969) ont souvent été considérées comme prototypiques. On en trouve une bonne présentation critique dans Dubois et al. (1997). Clark & Chase (1972 ; 1974) ont notamment étudié la formation et la vérification de descriptions de dessins, Izmailov & Sokolov (1991) un modèle de discrimination des couleurs et de la brillance. On trouve plus récemment des études de description de paysages (Vaillant, 1993) qui s'inscrivent dans toute une

perspective de description d'itinéraires. On peut enfin considérer une étude sur lexicologie des textures (Bhushan et al., 1997).

### Sur les odeurs

L'étude des odeurs est assez récente. On trouve une revue critique générale de leur ensemble dans Beguin & Costermans (1994). Parmi les premières études utilisant des verbalisations par descripteurs sémantiques, on trouve celle de Land et al. (1970). 14 substances odorantes sensées représenter les classes principales d'odeurs sont proposées comme standard pour l'étude de la perception des odeurs. 12 sujets expérimentés (!) doivent donner une note de 0 à 5 (absent - très présent) pour chaque odeur et pour 44 mots proposés. On observe déjà des différences de vocabulaire entre les parfumeurs et les gens qui travaillent dans l'alimentaire. D'autres études utilisant la méthode de sémantique différentielle suivent. Le but est de construire un profil sémantique des odeurs étudiées (Dravnieks & Bock, 1978; Dravnieks, 1982). Les relations entre ces profils sémantiques et des critères de perceptions des mêmes odeurs (par exemple à partir de mesure de similarités) sont étudiées et modélisées (Callegari, 1994). Les profils sémantiques obtenus semblent stables. On peut noter que presque toutes les échelles sont construites à partir de noms de sources odorantes, par exemple orange, miel...

Dans une autre perspective, on trouve des études sur la mémoire des odeurs et de leurs noms (Engen, 1987), et dans un numéro spécial d'*Enfance* consacré à l'odorat chez l'enfant (Schaal, 1997), des études sur les performances olfactives précoces et leur lien avec le langage (Engen & Engen, 1997), sur la cognition olfactive et le lexique olfactif (Roudy et al., 1997) et sur la dimension sociale de l'olfaction (Dubois et al., 1991). L'étude de Dubois & Roudy (1997) prend le contrepied de la démarche habituelle en se positionnant "du linguistique au neuronal". Il s'agit de critiquer la conception du langage comme nomenclature souvent implicite dans ces études où il est considéré comme "un répertoire d'étiquettes posées sur les choses ou les concepts dans une relation de vérité". Leur étude montre en effet le problème d'une description objective de ce que serait une odeur. Ils proposent au contraire "que l'odeur serait non pas objet du monde extérieur, mais davantage un effet (du monde certes) produit sur le sujet (et donc ressenti par lui)" (ibid.). Les verbalisations seraient le reflet d'une "représentation en langue de leur expérience subjective. L'odeur comme effet sur un sujet resterait ainsi un invariant individuel dont l'objectivité n'a pas été construite, dans notre culture tout au moins, par la négociation d'un partage du sens dans l'interaction verbale." (ibid.)

Dans la même lignée, on trouve des études linguistiques de définitions de l'odeur (David, 1997; David et al., 1997), soulignant le fait que "pour l'olfaction, on ne dispose en français que de très peu de termes lexicalisés (...) à l'exception de *odeur* et *parfum*." Et montrant que "la langue française traite l'odeur comme une propriété caractéristique d'un objet (...) pas comme une propriété autonomisée ('odeur de cheval' mais pas 'odeur cheval')", et ce, malgré l'existence de "produits" comme les désodorisants, dont la fonction centrale est de produire (et/ou masquer) une odeur, qui autonomisent l'odeur de son support "naturel". Ceci conduit à "développer un tel usage en langue, et donc à rendre acceptable en français "odeur + nom", de manière identique à ce qui s'est historiquement produit depuis longtemps dans le domaine de la couleur" (ibid.). David et al. (1997) dégagent aussi des axes

de classification des odeurs liés : à la source, à l'intensité, à la mémoire, à la valeur hédonique (ou à l'effet).

### **Sur le goût**

Enfin, le domaine du goût commence à se développer, principalement dans le cadre de départements "d'analyse sensorielle" de grandes entreprises alimentaires, même si on trouve depuis longtemps un vocabulaire utilisé pour décrire le vin (Paris-Grignon, 1996). L'étude du goût est complexe, d'une part à cause du fait qu'il est étroitement lié à l'olfaction : "la perception rétro-nasale et la sensation du goût sont en fait complémentaires et synergiques et portent le nom de flaveur" (Urdapilleta, 1998), d'autre part à cause de son caractère évanescent. Au contraire du domaine visuel, il n'existe pas de nuancier standardisé des odeurs et des flaveurs reproduisant l'exhaustivité de l'espace olfactif.

### III.3. Sur la musique

La plupart des études utilisant des verbalisations concernant la modalité auditive se situent dans une recherche du sens de la musique. Il s'agit souvent d'essayer de montrer en quoi la musique est un langage dont le sens est compris objectivement, en lien notamment avec les émotions qu'elle suscite. (On trouve une bonne bibliographie de ces études dans Wedin, 1972).

#### **Majeur / mineur**

Ainsi, plusieurs études essaient de montrer le lien entre les modes majeurs et mineurs et des émotions de joie ou de tristesse. Ce genre de travaux est déjà présent chez Rameau et d'Alembert (cités par Helmholtz, 1877). Leurs arguments sont du type physique et métaphysique : la tierce majeure étant la plus naturelle et la plus parfaite, bénéficierait d'une "beauté cosmique". Britan (1911) considère que "la conformité au standard est le secret de la valeur émotionnelle du mode majeur, la non-conformité, le secret du mode mineur". Beaunis (1918) discute des éléments émotionnels en musique et de leurs relations avec le rythme, la durée, l'intensité, la hauteur et la vibration générale du corps, plus intervention des états d'âme du sujet.

Des études plus expérimentales apparaissent avec Heinlein (1928) : les sujets doivent choisir des adjectifs dans une liste (et éventuellement rajouter les leurs), à propos de ce que chacun des 48 accords présentés leur font ressentir. Les adjectifs de la liste ont été considérés comme adéquats par les sujets, bien que certains aient rajouté leur propre vocabulaire. Ces adjectifs sont répartis en deux classes : ceux habituellement utilisés pour décrire un mode majeur, et ceux pour le mode mineur. Ce classement fait office de réponse "correcte". On observe moins d'"erreurs" d'accords mineurs décrits avec des adjectifs considérés comme décrivant des modes mineurs que l'inverse. L'intensité est responsable de beaucoup d'erreurs, un accord joué "piano" étant décrit avec les mêmes termes qu'un accord mineur. Pour Heinlein, la réponse commune de majeur / gai, mineur / triste est apprise par association lors des premières présentations.

Au contraire, Hevner (1935) considère que les résultats historiques et expérimentaux donnent l'évidence de l'existence d'un sens à la musique relativement constant et universel bien qu'il y ait de

fortes variations entre les gens. Ainsi, Sorentin (1932) montre des associations entre la joie et des harmonies simples et des rythmes circulaires, ou association entre la douleur et des tempos lents et des dissonances. Ces résultats sont moins évidents si le sentiment à trouver n'est pas pur ou immédiat. Gundlach (1935) montre à partir de trois types de chants indiens (pour soigner, d'amour ou de guerre) que l'intention peut être reconnue. Leur principale différence, d'ordre rythmique, semble de première importance pour la reconnaissance de l'expressivité. Une des questions posées est de savoir par exemple si ce sens (ou cette intention) est généralement compris par toute sorte d'auditeurs ou seulement par un petit pourcentage d'auditeurs entraînés. Y a-t-il un système d'expressivité avec des éléments de structure musicale, comme le rythme ou les accords, auxquels certaines idées ou émotions seraient toujours associées ? En utilisant deux types de méthodes : introspection ou comparaison avec une liste de huit sortes d'émotions ou d'humeurs, Hevner (1935) montre que le sens de la musique est le même pour tous les auditeurs même si ceux qui ont le plus d'entraînement font des différenciations plus fines et plus facilement. La reproductibilité des résultats augmente si les extraits musicaux sont courts et ne présentent qu'un seul type d'humeur. Elle conclue tout de même que beaucoup de paramètres très liés les uns aux autres interagissent et ne sont jamais présentés indépendamment dans un morceau. Ceci rend impossible la construction d'une sorte de dictionnaire du sens musical, d'autant plus que la musique ne produirait que des humeurs générales et qu'en tant qu'art temporel : la succession et la répétition peuvent changer le sens. Elle affirme pourtant : "Nous croyons qu'un tempo rapide combiné à un rythme vivant, des harmonies majeures simples et une intensité légère dans un registre aigu produiront un effet d'humeur gaie et drôle, et que cette humeur se transformera en agitation en changeant le rythme, ou en étrangeté si on modifie le mode, ou en délicat en changeant l'intensité, ou en sérénité en baissant le registre ou en modifiant le rythme, etc..." (ibid.) D'autres études suivent dans la même lignée (Hevner, 1936; Hevner, 1936; Hevner, 1937; Rigg, 1937) s'intéressant aussi aux différences individuelles (Sopchak, 1955).

#### **Jugements continus en écoute d'une oeuvre**

On peut noter que l'idée de faire choisir des adjectifs parmi une longue liste lors de l'écoute d'extraits d'oeuvres musicales est déjà présente chez Hevner (1935). On retrouve ce type d'expérimentation plus récemment (Campbell, 1942; Wedin, 1972; Asada & Ohgushi, 1991; Namba et al., 1991). Ainsi, par exemple, (Wedin, 1972) propose de choisir 5 à 10 adjectifs dans une liste de 125 présentés au hasard, en rapport avec 40 extraits d'oeuvres entendues d'environ une minute. A partir de la fréquence d'utilisation et des similarités, il classifie les adjectifs utilisés en trois niveaux. Huit grandes catégories d'adjectifs émergent : gaieté, mélancolie, rage, peur, solennité, religiosité, exaltation et douceur. Asada & Ohgushi (1991) font écouter le Boléro de Ravel en utilisant pour chacune des 18 parties du morceau 25 échelles sémantiques. Ils obtiennent deux dimensions principales liées à l'intensité (*fort* en décibel) et à un côté *brillant*. Namba et al. (1991) proposent quatre interprétations différentes (piano ou orchestre) des 5 promenades des Tableaux d'une exposition de Moussorgsky et demandent de choisir parmi 60 adjectifs ceux qui correspondent le mieux à chaque mouvement. Trois facteurs principaux apparaissent : la dynamique (énergie du son), la tranquillité (tempo ou vitesse d'exécution) et la tristesse (mélodie). Ceci rappelle les trois facteurs principaux d'Osgood et al. (1957)

: "*potency, activity, evaluation*". La spécificité du travail de Namba et al. est qu'ils comparent des jugements effectués en continu pendant l'écoute et en fin de morceau. Les résultats semblent globalement convergents.

### III.3. Sur les sons

Le domaine sonore se situe du point de vue du "statut des objets" dans une position intermédiaire entre le domaine visuel et l'olfaction. En effet, le son est maintenant facile à reproduire à l'identique ou en contrôlant des modifications, au contraire des odeurs qui sont encore difficiles à fixer sur des supports. Par contre, c'est un phénomène d'ordre temporel qui ne peut donc s'affranchir d'être toujours lié à la mémoire qu'on en a, au contraire de la vision qui présente un caractère fixe. Nous allons essayer de présenter séparément les études effectuées sur les sons, les bruits et le timbre, bien que certains auteurs ne se soient pas spécialisés sur un de ces domaines.

Il s'agit tout d'abord de rappeler l'ampleur du travail de Schaeffer (1966), à la recherche d'un "solfège ou d'une typologie des objets musicaux", dans son Traité des objets musicaux. En effet, bien que ce ne soit pas un travail relevant directement d'une discipline scientifique comme la psychologie expérimentale, il impressionne par sa qualité et l'étendue de ses conclusions. Dans le même genre de démarche, il faut noter le fantastique travail de Frize (2000) sur les références culturelles du bruit et de l'audition à travers l'analyse de nos représentations et nos constructions mentales (idéologiques, sensibles, éducatives...) du concept de bruit (tout ce qui est sonore y est de fait inclus), ainsi que sur la façon dont nous envisageons collectivement l'action auditive, "active" ou "passive". "L'étude a porté d'une part sur l'analyse de textes ressources qui participent de nos connaissances et de nos références (les dictionnaires du XIIe siècle à nos jours, l'intégralité de la Bible, les sept volumes des Lagarde et Michard, des livres illustrés pour enfants traitant de cette thématique, divers manuels scolaires, l'ensemble des campagnes 'anti-bruit' ministérielles ou communales...), d'autre part sur des observations in situ (les pratiques bruyantes et auditives dans huit disciplines sportives, une série d'entretiens avec un panel de personnes, une analyse qualitative et quantitative de 120 jouets sonores...), enfin sur une recherche - action dans une école primaire de la Seine-Saint-Denis, objet d'une expérimentation sur une pédagogie alternative de l'écoute." (Frize, 2000).

Nous comparerons nos résultats avec ces deux sources dans la discussion.

#### **Pas de vocabulaire spécifique**

Dans une étude sur diverses sonneries de téléphones à décrire par écrit afin que l'identification du sens soit certaine, Wright (1971) montre qu'il n'apparaît pas de vocabulaire spécifique disponible pour décrire ces sons. Il observe en effet deux catégories dominantes de verbalisations : des onomatopées et des descriptions imagées de la source. Taylor et al. (1974) retrouvent ce résultat lors de la description de vingt sons purs d'amplitudes et durées différentes par des sujets appartenant à trois groupes sociaux différents (ingénieurs, musiciens, secrétaires). Les réponses sont regroupées dans trois catégories : des onomatopées (surtout produites par les secrétaires) : "*bleep, buzz, pop, pip*", des

noms de sources sonores (surtout produits par les ingénieurs) : "*fog horn, tug boat, bass instrument, siren*" et caractéristiques physiques (surtout produites par les musiciens) : "*high or low, long or short, slow or fast*". De plus, les musiciens utilisent des descriptions de qualités émotives : "*eerie, anxious, spooky, frightening, haunting*".

### **Linguistique**

On peut citer deux séries d'études de linguistique sur la façon dont on parle des sons. Celle de David (1997), que nous avons déjà citée au §I.3. sur une analyse de définitions en langues des sons et bruits, et celle de Dupuy-Engelhardt (1990; 1994), sur une étude comparée du champ lexical de l'audible en allemand et français. Cette dernière étude présente de manière quasi-exhaustive l'ensemble des verbes, adjectifs et substantifs allemands et français ayant une connotation "sonore". Elle note par exemple qu'en ce qui concerne les adjectifs "le français caractérise avant tout la qualité du phénomène audible (*sonore* vs *sourd, rauque*), à laquelle se joint l'intensité (*bryant*) et, en plus sa hauteur (*strident*). (...) Tous les autres adjectifs sont, soit des emprunts à d'autres champs, soit des lexèmes apparaissant dans plusieurs champs", citant comme exemple l'adjectif *aigu* (Dupuy-Engelhardt, 1994). Elle répertorie aussi une soixantaine de substantifs, surtout centrés sur l'intensité des bruits (*brouhaha, chahut, fracas, rumeur, tumulte, vacarme...*), et note que les verbes se subdivisent en trois classes, "selon leur appartenance classématique, en phénomènes audibles produits par des êtres humains (12 verbes comme *gronder, grogner, grommeler, zézayer, bégayer...*), des animaux (64 verbes comme *aboyer, beugler, rugir, bêler, blatéer...*), et le reste (bruits de la nature et artefacts : environ 60 verbes français comme *clapoter, crépiter, grésiller, glouglouter, cliqueter, sonner, tintinnabuler...*)".

Il semble donc au contraire apparaître un grand nombre de mots utilisés pour parler des événements sonores. On peut noter l'importance des verbes en lien avec le mode de production des sons.

### **Sémantique différentielle**

Dans le cadre de la sémantique différentielle, la première étude dans le domaine sonore est celle de Solomon (1957) sur des sons de sonar écoutés et jugés par des "oreilles d'or" de la US Navy. 20 sons sont jugés selon 50 échelles bipolaires inspirées de celles d'Osgood et al. (1957). Il obtient sept facteurs principaux qui expliquent 40% de la variance : "*magnitude, aesthetic, clarity, security, relaxation, familiarity, mood*". On trouve dans la même veine l'étude de Kerrick et al. (1969) sur 16 sons brefs de musique, voitures et sons artificiels, jugés selon 15 échelles bipolaires. Un des intérêts de cette étude c'est que Kerrick et al. cherchaient non pas un profil de chacun de leurs sons ou un ensemble de dimensions (ou facteurs principaux) permettant de définir l'ensemble des sons. Ils s'intéressaient au contraire au sens que donneraient des participants aux deux échelles "*noisy-quiet*" et "*soft-loud*" correspondant aux notions de "bryant" et de "fort", considérées comme deux aspects différents des sons par les acousticiens. Ces deux échelles sont utilisées comme synonymes, malgré une présentation renversée. L'étude de Nielzén & Olsson (1993) sur 7 sons (instruments, voix humaine, réveils, électronique) présentés en deux durées différentes (0.5 s et 3 s) et jugés selon 12 échelles sémantiques, s'intéresse à d'éventuelles différences de positionnement des mêmes sons

suivant si l'extrait écouté est long ou très bref. Ils n'obtiennent pas de différences significatives liées à la longueur des sons et concluent que les sons très courts ne dénaturent pas le sens véhiculé.

#### **Bruits de voiture et d'aspirateurs**

Parmi les bruits dont la dénomination est étudiée, on trouve, outre les sons de sonar déjà cités, des sons de voiture, de circulation et d'aspirateurs. Ainsi par exemple, pour les bruits produits par le moteur des voitures variant en continu, Weber (1990) utilise des jugements continus sur des échelles catégorielles équipées de curseurs, et Schulte-Fortkamp & Weber (1990, 1993) comparent des jugements sur l'intensité de bruits de voitures sur le bord de la route et dans leur laboratoire. On trouve aussi une étude sur les bruits de voisinage (Guski, 1990) et une étude de la catégorisation d'un corpus de bruits domestiques, grattements, frottements, mais aussi bruits d'aspirateurs (Guyot, 1996; Guyot et al., 1998).

Enfin, pour caractériser les bruits produits par les voitures du constructeur Renault, Nosulenko et al. (1998, 1999) utilisent des verbalisations libres pour construire des échelles sémantiques, comme nous l'avons fait aussi dans notre deuxième expérience. Les résultats obtenus dans les deux conditions, verbalisations libres et utilisations des échelles sémantiques, sont comparables. Les auteurs attribuent cette adéquation au fait que les échelles sont construites à partir du vocabulaire extrait de verbalisations libres sur ces mêmes sons. Néanmoins, ils présentent aussi chacun des attributs verbaux proposés comme échelle sémantique dans plusieurs phrases permettant de comprendre le contexte d'utilisation de ces mots. Nous verrons que ceci peut être une explication suffisante de la réussite de l'utilisation de ces échelles.

L'ensemble de ces travaux, ainsi que toute une série de travaux identiques en préparation notamment pour (France Télécom, EDF, SNCF...) reflètent l'engouement actuel d'entreprises qui commencent à se préoccuper de design sonore ou du moins de limiter les nuisances sonores des appareils et machines qu'ils produisent.

#### III.4. Sur le timbre

On trouve déjà des descriptions langagières de qualités de timbres de sons dans l'étude de Helmholtz (1877) : les sons simples sont considérés comme "*sweet, pleasant*", ceux contenant des basses fréquences sont "*dull*", ceux contenant des hautes fréquences sont "*brighter and sharper*". Beaucoup d'amplitude dans les harmoniques basses donnent des sons *plus musicaux et riches* alors que beaucoup d'amplitude dans les harmoniques hautes les rende "*sharp and penetrating*". Enfin, la présence d'harmoniques impaires induit le caractère *nasal* de la clarinette et si toutes les harmoniques sont présentes avec des amplitudes égales, les sons ressemblent aux "*softer tones of the horn*".

Cette étude sera à l'origine à la fois de caractérisations du timbre en utilisant des échelles sémantiques, mais aussi de descriptions moins systématiques qui seront ensuite recueillies avec cette méthode.

#### **Sémantique différentielle**

L'étude de von Bismarck (1974 a, b) s'inscrit dans la continuité de celles de Solomon (1957) et de Kerrick et al. (1969), dont les échelles sémantiques seront à la base de son choix, mais aussi des psychologues introspectionnistes comme Stumpf (1883 u. 1890) et Lichte (1941), qui cherchaient un ensemble d'attributs verbaux comme "*brightness, fullness, sharpness, roughness, volume and density*", qui puissent décrire le timbre. Pour la première fois, cette étude se consacre uniquement à des sons égalisés en hauteur, durée, intensité et "steady-sounds". En effet, il demande de positionner 35 timbres sur 30 échelles sémantiques bipolaires (de type *chaud - froid*), sélectionnées à partir d'un ensemble de 69 échelles pour leur "pertinence pour décrire le timbre" (ceci a priori, sans écouter les sons). L'ensemble de sons est représentatif de sons de voix, de musique et de bruits.

Kendall & Carterette (1992 a,b, 1996) reprennent le principe de cette étude sur 10 sons de mélange de 2 instruments à vent à partir de sons de flûte, clarinette, saxophone, trompette et hautbois. Trois expériences sont rapportées. Une utilisant 8 échelles sémantiques bipolaires choisies parmi les 28 échelles de Bismarck (*hard-soft, sharp-dull, loud-soft, complex-simple, compact-scattered, pure-mixed, dim-brillant, heavy-light*) auxquelles ils ajoutent deux autres concernant l'intensité et la hauteur. Les profils obtenus pour les sons sont parallèles avec peu d'écart-type et ne permettent donc pas de comprendre la distinction perceptive entre les timbres. Ils se posent alors la question des échelles bipolaires : est-ce que les opposés choisis le sont vraiment ?

Leur deuxième expérience présente donc une transformation des échelles bipolaires (de type *chaud - froid*) en échelles unipolaires (de type *chaud - pas chaud*) en choisissant une des deux extrémités de chacune des échelles bipolaires précédentes (*hard, sharp, loud, complex, compact, pure, dim, heavy*). Cette technique sera baptisée VAME pour *Verbal Attribute Magnitude Estimation*. Ils remarquent que certaines échelles (*loud, heavy et hard*) permettent de différencier les sons contenant du saxophone des autres sons. Ils sont troublés de voir que *loud* est très utilisé pour différencier les timbres bien que ceux-ci soient égalisés en sonie !

Pour leur troisième expérience, ils cherchent donc à modifier à nouveaux les échelles utilisées en les construisant à partir de termes plus "musicaux" et donc plus pertinents pour décrire des timbres. Pour cela, ils extraient tous les adjectifs de qualités sonores du traité d'orchestration de Piston (1955). La liste totale est ensuite réduite à 61 adjectifs, puis 21 d'entre eux sont choisis par procédure de liste ouverte (les sons sont rejoués régulièrement pendant que les sujets doivent choisir dans une liste les adjectifs qui permettent de décrire ces sons). Une analyse en composantes principales des résultats obtenus sur les 21 échelles sémantiques unipolaires ainsi constituées présente 4 dimensions principales (*power, strident, plangent and reed*).

Nous verrons comment ces études ont inspirées notre deuxième expérience et nous comparerons nos résultats à ceux-ci.

On peut noter enfin quelques études de la qualité des appareils de reproduction des sons (Enomoto & Yoshida, 1968; Gabrielsson & Sjögren, 1979). Ainsi, Enomoto & Yoshida recueillent des jugements sur des appareils de reproduction du son selon trois méthodes différentes : 1- à partir de 60 échelles sémantiques choisies par des spécialistes comme les plus appropriées parmi une liste de 200; 2-



jugements de similarité; 3- descriptions verbales libres. Ils obtiennent ainsi 3 facteurs pour décrire les haut-parleurs, cinq facteurs pour décrire les écouteurs (*headphones*) et 3 facteurs pour décrire les prothèses auditives (*hearing aids*).

### Verbalisations libres

Enfin, nous allons présenter ici le principe de l'étude de Samoylenko, McAdams & Nosulenko (1996) sur l'analyse systématique de verbalisations produites en comparant des timbres musicaux dont notre première expérience présente la continuité directe. Leurs résultats seront comparés aux nôtres dans la discussion du chapitre 6.

Les données brutes sont des enregistrements de verbalisations libres de sujets russes non musiciens qui ont à décrire "avec le plus de détails possible les différences et similarités entre les deux sons" de chacune des paires qui leurs sont proposées. Les protocoles verbaux sont transcrits sans omissions de l'oral à l'écrit et sont entrés dans une base de données développée par Nosulenko & Samoylenko (1997) dans une perspective développée à l'Institut de Psychologie de l'Académie des Sciences de Russie depuis plus de 20 ans. Les différences d'interprétation sont résolues par consensus entre différents experts. Dans une première phase d'analyse, les unités verbales ayant un sens ont été extraites du texte. Chaque unité correspond à une caractéristique différente du son. Par exemple, dans la description "un son très riche, aigu", il y a deux unités verbales : "très riche" et "aigu", alors que dans la description "ressemble à un marteau qui tombe et rebondit", la phrase entière constitue une unité verbale. Chaque unité verbale correspond à un sujet, une des paires écoutées et le timbre décrit dans la paire.

Puis commence l'entrée dans la base de données proprement dite, qui consiste en trois niveaux d'analyse principaux correspondant à l'analyse du sens logique de chaque unité verbale, de leurs relations aux objets comparés, puis des aspects sémantiques.

Une description de l'ensemble des niveaux d'analyse est présentée dans (Nosulenko & Samoylenko, 1997). Comme nous ne nous étions intéressés dans notre première expérience qu'aux trois étapes du premier niveau (sens logique), dont l'arborescence est résumée dans la figure 3, nous ne détaillerons ici que le principe de codage de celles-ci.

- étape 1 : Les unités verbales sont séparées en deux : celles décrivant des similarités entre les deux sons de la paire présentée sont codées **/sim/** dans la base de données, et celles décrivant leurs différences **/dif/**. Ainsi, si on considère par exemple, l'extrait suivant du protocole verbal brut : "..similaires dans le sens... les deux attaques sont lentes... même si la seconde est un peu plus courte et il y a à la fin une sorte de bruit". Il contient 3 unités verbales. La première : "les deux attaques sont lentes" sera codée **/sim/** car elle correspond à des similitudes des deux sons. Les deux autres : "la seconde est un peu plus lente" et "il y a à la fin une sorte de bruit" seront par contre codées **/dif/**.

- étape 2 : Chacun des deux groupes précédents est scindé en deux. Si les unités verbales expriment des aspects concrets des sons, elles sont codées **/con/**, si elles expriment des aspects généraux, elles sont codées **/gen/**. Ainsi par exemple dans l'énoncé : "les sons ont une couleur très semblable", il manque des unités verbales permettant de comprendre en quoi cette couleur est semblable, elle sera donc codée **/sim/gen/**. Part contre, la première partie du protocole de l'exemple précédent : "les deux

attaques sont lentes" sera codé **/dif/con/** dans la mesure où le type de différence est spécifié concrètement.

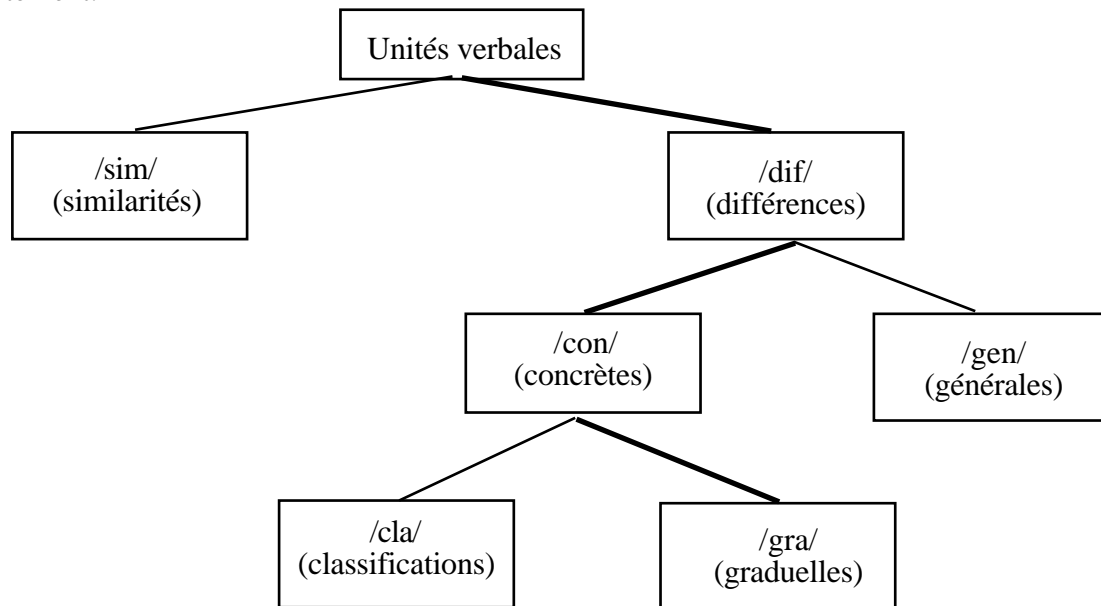


Figure 3 : Structure des trois étapes du niveau du "sens logique" de la base de données

- étape 3 : Les unités verbales classées précédemment dans les différences concrètes, code **/dif/con/** sont encore séparées en deux groupes : celles de type classificatoire, codées **/cla/**, et celles de types graduel, codées **/gra/**. Ainsi par exemple, sont considérées comme classificatoires les unités verbales suivantes : "le premier son est très brillant" ou "ressemble à un violon". Par contre une unité verbale est considérée comme graduelle si les deux stimuli sont comparés par rapport à la même classe sémantique. Ces unités verbales rapportent un degré de différence entre deux objets. Par exemple : "le second son paraît plus fort (que le premier)".

La technique se montre efficace pour révéler dans quelle mesure des différences accordées aux attributs perceptifs des timbres contribuent aux divergences inter-sujets et intra-sujets dans l'évaluation de la dissemblance. De plus, les données suggèrent que les catégories de base utilisées par ces sujets non-musiciens pour identifier les sources sonores produisant les événements sont très générales.

### III.5. Résumé

Après avoir rappelé les arguments des partisans et des opposants aux méthodes introspectives, nous avons présenté diverses méthodologies utilisant comme données ce que disent les participants : lors de protocoles de type "sémantique différentielle" ou dans des verbalisations libres. Nous avons ensuite résumé les travaux visant à décrire le contenu de nos perceptions visuelles, olfactives et gustatives, avant de nous focaliser plus particulièrement sur les études décrivant nos perceptions auditives.

Dans ce domaine, nous avons parcouru les études sur la musique, sur les sons en général puis plus spécifiquement sur le timbre des sons.

#### **IV. Conclusion**

Le timbre apparaît donc comme une qualité multidimensionnelle des sons. Cette qualité peut être modulée pour créer les différentes voyelles et consonnes de la parole ou pour servir de matériau à la musique. Elle donne aussi accès à l'identité de la source ayant produit le son, et sans doute particulièrement grâce à une de ses composantes : la rapidité de l'attaque. Cette propriété en fait le sujet sous-jacent de toute étude sur l'identification d'objets sonores, que ce soient des bruits ou des sons produits par des instruments.

D'autre part, il existe plusieurs points de vue à l'étude de la façon dont on parle des sons. On trouve ainsi des études linguistiques, par exemple celles de Dupuy-Engelhardt (1990) et de David (1997), qui analysent les usages de la langue au niveau du champ lexical ou des structures syntaxiques présentes dans des textes littéraires ou des définitions de son et bruit. A l'autre extrémité, on trouve des études expérimentales de psychologues, qui cherchent un nombre d'axes sémantiques minimum pour définir et décrire un ensemble de sons, par exemple à la suite de l'étude de Solomon (1957) sur les sons de sonars, celles de von Bismarck (1974) ou de Kendall et Carterette (1992 a, b) sur le timbre de sons d'instruments. Ces études utilisent le langage dans une version réduite à une vingtaine de mots au maximum, souvent des adjectifs. Elles visent plutôt à expliciter une représentation mentale des sons considérés. Entre ces deux types d'études, on trouve des descriptions des usages des musiciens, ingénieurs du son, architectes acousticiens... représentées par les études de Schaeffer (1966) et de Frize (2000). Elles cherchent plutôt à décrire les sons ou le phénomène sonore. Nous cherchons dans ce travail une position intermédiaire entre celles des expérimentalistes et celles des linguistes ou musiciens. Notre démarche, inspirée de celle de Samoylenko et al. (1996) et de Nosulenko et al. (1998), sera expérimentale, mais nous essayerons de garder la richesse des descriptions des sons obtenues lors de verbalisations libres. Nous chercherons à la fois à décrire les sons et à comprendre comment nous les percevons à travers notre façon de les décrire. Enfin, et c'est l'originalité de notre travail, nous allons poser la question de la possibilité de reconnaître les sons décrits à partir de leur description de manière expérimentale.

## Hypothèses

Nous allons donc tester les implications des théories présentées ci-dessus au cours des 3 expériences décrites en introduction et dont les analyses seront développées dans la partie II. On rappelle qu'il s'agit d'une expérience de comparaison de paires de sons et de description libre de leurs différences et ressemblances, d'une expérience demandant de positionner ces mêmes sons sur des échelles sémantiques construites à partir des termes verbalisés spontanément et d'une expérience d'association de ces mêmes sons à des descriptions construites à partir des verbalisations de la première expérience.

Le chapitre 1, "Être au monde", nous a permis de comprendre l'influence de l'homme sur la nature du monde qu'il croit percevoir comme extérieur à lui.

Nous nous attendons donc à trouver dans les descriptions des sons une trace de cette place de l'homme. Les sons ne devraient donc pas être décrits "en-soi", mais en liaison avec les préoccupations et les activités humaines.

Le chapitre 2, "Parler", nous a permis de comprendre que le langage est un code inscrit dans une société et qu'il possède deux propriétés principales : celle de rendre compte du monde et celle d'être un outil de communication.

Nous nous attendons donc à trouver dans les descriptions une trace du fait qu'utiliser un mot, c'est toujours faire entrer sa perception particulière dans un code défini socialement. Les sons devraient être décrits avec des mots dont le sens ne recouvrerait pas exactement la même expérience à chaque occurrence. Ce sens devrait notamment varier (tout en restant compréhensible donc "conventionnel") en fonction des locuteurs et des conditions expérimentales (notamment verbalisations libres vs échelles sémantiques).

Nous nous attendons aussi à trouver une trace du fait que le langage a une fonction de communication et nous testerons l'efficacité de cette communication dans notre troisième expérience.

Enfin, nous comparerons nos résultats avec ceux présentés dans le chapitre 3.

## **Partie II : Expériences effectuées**

La partie expérimentale de cette thèse comporte trois expériences, effectuées sur des sujets différents ou en partie différents mais toujours sur les mêmes 12 sons (qui sont décrits dans la partie stimuli de l'expérience I) et que vous pouvez écouter dans le CD en annexe.

Dans la recherche de liens entre le timbre des sons que l'on perçoit et les mots que nous utilisons pour en parler, le but principal était de pouvoir expérimenter :

- d'une part, la façon dont les gens parlent de sons qu'on leur propose d'entendre, soit de manière libre en comparant les sons par paires (expérience 1) ou en les décrivant successivement selon diverses consignes (expérience 3), soit en utilisant des échelles sémantiques (expérience 2).
- d'autre part, la possibilité de reconnaître un son à partir d'une de ses descriptions (expérience 3).

Par ailleurs, nous nous sommes donnés deux autres buts :

- rechercher des corrélats verbaux des dimensions perceptives du timbre qui avaient été dégagées dans les études précédentes (Grey, 1977 ; McAdams et al., 1995), soit dans les verbalisations libres (expérience 1) soit à partir des échelles sémantiques (expérience 2).
- tester la pertinence ou la précision du sens du vocabulaire proposé aux auditeurs dans les expériences 2 et expérience 3.

La plupart de nos objectifs sont testés à cheval sur plusieurs expériences. Plutôt que de reprendre le même type d'analyse dans différentes sous-parties de l'analyse de chacune de nos 3 expériences, il nous a paru plus intéressant de regrouper les données et les analyses qui nous permettaient de répondre à chacune de nos quatre questions dans les chapitres 5 à 8. Nous présenterons dans le chapitre 4, l'ensemble des trois méthodes utilisées, ainsi que quelques analyses préliminaires spécifiques de chacune des trois expériences.

En effet, la première expérience comporte deux tâches co-occurentes : un jugement de dissemblance sur des paires de sons et des verbalisations libres sur toutes les dissemblances et les similarités entre les deux sons de chacune des paires. Elle nous permettra donc de modéliser notre représentation perceptive du timbre sous forme d'espace multidimensionnel (chapitre 4), de rechercher des corrélats verbaux des dimensions perceptives (chapitre 5) et de présenter un panorama du lexique utilisé pour comparer les timbres des sons (chapitre 6).

La deuxième expérience comporte aussi deux tâches : positionner chacun des 12 sons sur chacune des 23 échelles sémantiques construites à partir des verbalisations libres recueillies dans l'expérience 1 et donner un jugement de pertinence de chacune de ces échelles de manière générale et pour chacun des sons. Elle nous permettra donc d'une part, de rechercher des corrélats verbaux des dimensions perceptives de manière plus quantitative mais plus dirigée et de comparer les résultats obtenus avec ceux dégagés à partir de l'expérience 1 (chapitre 5) et d'autre part, de tester la pertinence pour les

sujets d'un vocabulaire qu'on leur propose sous forme d'échelles sémantiques et qui a été produit par une partie d'entre eux (chapitre 7).

Enfin la troisième expérience comporte 3 tâches. Les sujets devaient tout d'abord décrire chacun des 12 sons en se répartissant selon 4 consignes différentes (de manière libre, en donnant un son de source, d'instrument ou de famille d'instruments de musique ou de qualité sonore). Puis ils devaient associer à chacun des portraits verbaux proposés un ou plusieurs sons du groupe. Enfin, ils devaient donner leur jugement sur le caractère précis, vague ou dépourvu de sens des descriptions qu'on leur avait proposées. Cette dernière expérience nous permet donc de compléter la description du champ lexical utilisé pour parler des sons (chapitre 5), de tester la possibilité de reconnaître un son à partir de sa description (chapitre 8) et de tester la pertinence des différentes façons de décrire un son (chapitre 7).

## Chapitre 4 : Présentation et Méthodologie des trois expériences réalisées

Dans ce chapitre, nous allons rassembler les données techniques de méthodologie ainsi que quelques analyses préliminaires qui ont été effectuées pour chacune des trois expériences que nous avons réalisées au cours de cette thèse. Les trois grandes parties correspondront donc tout naturellement à chacune de nos trois expériences, intitulées respectivement : I. Jugements de dissemblance entre paires de sons et verbalisations libres, II. Echelles sémantiques et III. Reconnaissance de portraits verbaux.

### **I. Expérience 1 : Jugements de dissemblance entre paires de sons et verbalisations libres**

#### I.1. Sujets

21 sujets ont participé à l'expérience mais l'enregistrement des commentaires verbaux d'un des sujets ayant été défectueux, nous l'avons écarté de nos analyses. Les 20 sujets restants (voir annexe A4.1) se répartissaient en deux groupes. Le premier comporte 10 musiciens dont 3 instrumentistes professionnels (S1.pro, S2.pro et S3.pro) et 7 musiciens amateurs ayant au moins 10 ans de pratique instrumentale. Parmi ceux-ci 3 avaient un métier lié à la musique (S4.am1, S5.am1, S6.am1), les 4 autres étant notés S7.am2, S8.am2, S9.am2 et S10.am2. Le second groupe comporte 10 non-musiciens dont deux musiciens amateurs depuis moins de 2 ans sans formation musicale théorique (S11.am3 et S12.am3), les 7 autres étant de vrais néophytes (notés Si.néo, i allant de 13 à 20). Les sujets avaient entre 22 et 35 ans ( $M=25.0$ ), comprenaient 12 femmes et 8 hommes, et, mis à part les musiciens professionnels, étaient tous étudiants dans des disciplines scientifiques variées. Les sujets ont été testés au Laboratoire de Psychologie Expérimentale de l'Université René Descartes. Ils étaient indemnisés pour leur participation.

#### I.2. Stimuli

Les 12 sons présentés aux sujets étaient synthétisés sur un Yamaha TX802 FM Tone Generator. Ils viennent d'un ensemble de 21 sons développé par Wessel et al. (1987) (voir Tableau I.2.1). Ces sons sont censés soit imiter assez précisément certains instruments traditionnels occidentaux (comme par exemple, la trompette, le trombone, le vibraphone, le clavecin, le cor anglais, la clarinette, la corde frottée et le piano), soit représenter des hybrides d'instruments connus (par exemple, le *striano* est un hybride de corde frottée et de piano, le *vibrone* est un hybride de vibraphone et de trombone et l'*obochord* est un hybride de hautbois et de clavecin). Tous ces sons sont produits à la même hauteur (fréquence fondamentale 311 Hz, soit un Mi bémol 3) et ont été égalisés perceptivement pour la sonie (mezzo forte : environ 50 sur l'échelle de vélocité MIDI qui contrôle l'intensité sur les synthétiseurs numériques) et la durée (environ 830 ms). Ces timbres ont été choisis pour permettre des comparaisons avec les études perceptives multidimensionnelles (Krumhansl, 1989 ; McAdams et al.,

1995) et les études des corrélats acoustiques des dimensions perceptives (Krimphoff et al., 1994) effectuées dans le laboratoire.

Tableau 4.1. Noms, amplitude maximum et durée totale pour les 12 sons utilisés. Tous les sons avaient pour fréquence fondamentale 311 Hz (Mib 4).

Noms (origines des hybrides)	Code	Niveau max (dBA)	Durée totale (ms)
trompette	TPT	62	603
trombone	TBN	65	564
vibraphone	VBR	58	798
<i>striano</i> (corde frottée/piano)	SNO	60	896
piano échantillonné	PNE	64	798
clavecin	CVC	55	797
cor anglais	CRA	72	647
clarinette	CLA	65	647
<i>vibrone</i> (vibraphone/trombone)	VBN	55	1197
<i>obochord</i> (hautbois/clavecin)	OBC	66	747
corde frottée	CFT	56	1197
piano	PNO	59	1145
Moyenne		61.2	836
Écart type		5.2	228

### I.3. Procédure

L'expérience comportait trois phases : une phase de familiarisation, une phase d'entraînement et une phase test. Lors de la phase de familiarisation, les sons étaient présentés deux fois aux sujets dans un ordre aléatoire. Lors de la phase d'entraînement, 6 essais choisis au hasard parmi ceux du groupe, étaient présentés, en présence de l'expérimentateur. Enfin, les 66 essais expérimentaux constitués des 66 paires de sons différents possibles à partir des 12 timbres utilisés étaient présentés. L'ordinateur se chargeait à chaque nouvel essai de tirer au sort la paire présentée et l'ordre des deux timbres à l'intérieur de chaque paire. L'expérience durait entre une heure et une heure et demie suivant les sujets. Une pause était prévue au bout de 33 essais.

Dans chaque essai, le sujet devait écouter la paire de sons présentée dans un casque à une intensité confortable, en moyenne 61 dB SPL. La paire pouvait être répétée autant de fois que le sujet le jugeait nécessaire durant un essai. Il lui était demandé successivement de donner un jugement numérique de dissemblance entre les deux sons (de 1, pour deux sons peu dissemblables, à 9, pour des sons très dissemblables), puis de "décrire verbalement les similarités et les différences entre les deux sons entendus avec le plus de détails possibles". A la fin de la verbalisation, il a été proposé au sujet de modifier son jugement numérique s'il le jugeait nécessaire. Ces protocoles ont été enregistrés sur cassette, transcrits, puis analysés suivant la procédure décrite plus loin.



#### I.4. Matériel

Les auditeurs étaient assis dans une salle traitée phoniquement, en face d'un ordinateur. L'expérience était contrôlée par un programme LISP qui tournait sur un Macintosh SE/30 et qui commandait le synthétiseur TX802 Yamaha à travers une interface midi. Les stimuli étaient présentés de façon diotique dans un casque Sony Monitor K240 connecté directement à la sortie du synthétiseur. Les jugements de dissemblance ont fait l'objet d'une analyse multidimensionnelle de proximités (MDS). Celle-ci est décrite ci-dessous (§I.5). Les verbalisations libres ont tout d'abord été codées dans une base de données développée par Nosulenko & Samoylenko (1996 ; 1997) afin de choisir une partie du vocabulaire pour y rechercher des corrélats verbaux. Ce travail avait été commencé lors de mon DEA (Faure, 1994). L'ensemble des analyses effectuées est décrit dans le chapitre suivant (Chapitre 5 §I). Puis ces mêmes données de verbalisations libres ont été entrées dans un arbre hiérarchique de propriétés, STONE, développé par Poitrenaud (1995 ; 1998). Ces analyses sont décrites dans le chapitre 6.

#### I.5. Principe et résultats de l'Analyse MDS

Les corrélations entre les vecteurs de dissimilarité de toutes les paires de sujets ont été calculées. Ces vecteurs ont été soumis à une analyse de clusters hiérarchique utilisant l'algorithme du voisin le plus proche (*nearest neighbor, single link*). Le sujet numéro 16 étant clairement isolé du reste des sujets et la moyenne de ses corrélations avec chacun des 19 autres sujets était très inférieure à la moyenne générale ( $M_{cor} = .40 \pm .08$ ,  $M_{16} = .21$ ). Les données de ce sujet ont donc été écartées lors du calcul de l'espace multidimensionnel.

Nous avons donc effectué une analyse multidimensionnelle de proximités EXSCAL à partir des jugements des 19 sujets restants. Cette analyse, développée par Winsberg & Carroll (1989a), permet de modéliser la dissemblance perceptive  $D_{ij}$  entre les sons  $i$  et  $j$  par la distance  $d_{ij}$  dans un espace multidimensionnel de  $N$  dimensions,  $d_{ij}$  répondant à la formule suivante :

$$d_{ij} = \left( \sum_{k=1}^N (x_{ik} - x_{jk})^2 + s_i^2 + s_j^2 \right)^{\frac{1}{2}}$$

où  $x_{jk}$  est la coordonnée du timbre  $i$  sur la dimension  $k$ ,  $N$  est le nombre de dimensions partagées par tous les timbres et  $s_i$  est la "spécificité" du timbre  $i$ . Cette spécificité peut être conçue soit comme la coordonnée sur une ou plusieurs dimensions qui caractérisent uniquement ce timbre, soit comme la "force" d'un trait discret que possède uniquement ce timbre. Le nombre de dimensions et la présence ou absence des spécificités sont choisis sur la base du critère d'information statistique BIC (Schwartz, 1978) qui est calculé à partir du logarithme de vraisemblance et du nombre de paramètres nécessaires dans chacune des solutions considérées (voir McAdams et al., 1995, pour une revue de ce genre de technique).

Les jugements de dissemblance obtenus avant verbalisation sont expliqués par une solution en 4 dimensions sans spécificité. On retrouve une solution en 4 dimensions sans spécificité si on tient compte des modifications des jugements de dissemblance effectuées par les sujets après avoir donné leurs verbalisations sur la paire jugée (ce qui concerne en moyenne 7% des paires et en général une modification d'une seule unité dans l'échelle de 1 à 9). Cette solution est très fortement corrélée à la solution des jugements spontanés avant verbalisation ( $r(10) > .99$ ,  $p < .01$ ). Ainsi, après leur jugement de dissemblance sur une paire de sons, les sujets réécoutent plusieurs fois cette même paire (en moyenne 7.5 fois) et sont amenés pour élaborer leur verbalisation à se focaliser sur des aspects différents des deux sons considérés. Cela les amène parfois à modifier leur jugement numérique de dissemblance. Cependant, ces modifications n'affectent ni la structure (4 dimensions) ni la répartition des timbres dans l'espace perceptif modélisé. Les coordonnées de chaque timbre le long de chacune des 4 dimensions de cet espace sont rassemblées dans le tableau 4.2.

Tableau 4.2. Coordonnées des timbres le long des 4 dimensions perceptives de l'espace dérivé des jugements de dissimilarité de 19 sujets sur le 12 timbres. Les extrêmes de chacune des dimensions sont en gras.

Timbres	Dimension 1	Dimension 2	Dimension 3	Dimension 4
trompette	-2.8	0.7	-1.6	<b>2.6</b>
trombone	-3.1	-1.7	-0.1	0.8
vibraphone	<b>3.7</b>	-2.5	1.0	-1.1
<i>striano</i>	-0.6	1.4	-1.8	0.1
piano échantillonné	2.4	-0.5	-1.3	-0.3
clavecin	3.5	<b>3.0</b>	0.1	-0.4
cor anglais	-2.9	1.3	<b>2.6</b>	0.0
clarinette	<b>-3.1</b>	-1.3	1.8	<b>-2.3</b>
<i>vibrone</i>	1.3	<b>-2.7</b>	-0.6	2.0
<i>obochord</i>	2.0	2.5	2.0	1.6
corde frottée	-1.4	1.4	<b>-2.1</b>	-1.9
piano	1.1	-1.7	-0.1	-1.2
Étendue	6.8	5.7	4.7	4.9

Nous avons effectué les corrélations entre nos 4 dimensions et celles du modèle de McAdams et al. (1995) (voir tableau 4.3).

Tableau 4.3. Corrélations ( $dl = 10$ ) entre les coordonnées du modèle en trois dimensions de McAdams et al. (1995) et celles de notre modèle en 4 dimensions sans spécificité.

Modèle de McAdams et al. (1995)	Notre Modèle			
	D1	D2	D3	D4
Dimension 1	.97*	.02	-.28	-.01
Dimension 2	-.05	-.96*	.08	-.23
Dimension 3	-.06	-.02	-.56	-.55

\* :  $p < .0001$  (Fisher's  $r$ -to- $z$  ajustés pour tests multiples selon Bonféroni)

Les seules corrélations significatives ( $dl=10$ ,  $p < .0001$  ajustées pour tests multiples selon Bonféroni) sont celles de nos dimensions 1 et 2 fortement corrélées avec respectivement les dimensions 1 ( $r=.97$ ) et 2 ( $r=-.96$ ) de l'espace de McAdams et al. (1995) qui étaient elles-mêmes fortement corrélées aux

dimensions "Enveloppe temporelle" et "Enveloppe spectrale" de Krumhansl (1989). Ces deux premières dimensions ne sont donc pas affectées par la tâche de verbalisation co-occurente, contrairement à ce que les détracteurs de l'utilisation de verbalisations libres avancent souvent (Nisbett & Wilson, 1977). Nous avons essayé de trouver le meilleur corrélat acoustique de chacune de nos dimensions en effectuant directement les corrélations avec les 24 paramètres acoustiques dégagés par Krimphoff (1993) et Krimphoff et al. (1994). On retrouve le même type de corrélats acoustiques que McAdams et al. (1995) pour nos dimensions 1 et 2. Les coordonnées des timbres sur la dimension 1 sont corrélées avec une mesure du logarithme du temps de montée de leur enveloppe temporelle (LTM) ( $r=-.93$  ;  $p<.01$ ) et les coordonnées sur la dimension 2 avec le centre de gravité spectral ( $r=.90$  ;  $p<.01$ ). Notre dimension 3 est plus faiblement corrélée ( $r=.80$  ;  $p<.05$ ) à la déviation spectrale, mesure de l'écart entre la structure fine réelle du spectre et l'enveloppe spectrale globale de celui-ci. On ne trouve pas de corrélat acoustique de notre quatrième dimension parmi les paramètres acoustiques calculés par Krimphoff (1993). On retrouve donc le caractère temporel de la première dimension et le caractère spectral de la seconde. Ces deux dimensions perceptives, qui sont les deux principales, ne sont donc pas affectées par la tâche de verbalisation co-occurente.

## **II. Expérience 2 : Echelles sémantiques**

L'idée de cette expérience est née de la difficulté à tester de manière quantitative la présence de corrélats verbaux de dimensions perceptives du timbre musical à partir de verbalisations libres (cf. chapitre 5 § I). En effet, si les verbalisations libres ont l'avantage d'être toutes pertinentes pour celui qui les produit dans le contexte où il le fait, elles ne présentent aucun caractère de reproductibilité, ce qui rend leur analyse délicate, particulièrement quand on essaye de trouver des parallèles entre l'utilisation de mots et la position des sons décrits dans un espace perceptif qui modélise notre représentation perceptive de ces sons. Ainsi, pour pouvoir rechercher des corrélats verbaux des dimensions perceptives, nous avons dû :

- sélectionner les mots utilisés le plus fréquemment dans le contexte de comparaison graduelle entre les deux sons de la paire entendue ("le son A est plus (ou moins) XXX que le son B"), qui ne représentent qu'environ 3% de ces verbalisations libres, puis
- mélanger leur utilisation par l'ensemble des sujets qui avaient utilisé le même mot.

L'importance de la présence de corrélats verbaux composites de plusieurs dimensions perceptives dans les résultats a ainsi été attribuée :

- soit à une réelle différence d'orientation entre l'utilisation des mots désignant les concepts associés aux sons d'une part et les principales dimensions perceptives nous permettant de les comparer d'autre part.
- soit à une différence de sens attribuée à ces mêmes mots par les différents sujets dont les données avaient été jointes pour faire les analyses.

Cette expérience devrait nous permettre de trancher entre ces deux hypothèses.

## II.1. Présentation

### **II.1.1. Sémantique différentielle**

Nous avons choisi ici de reprendre et de tester le vocabulaire extrait précédemment des verbalisations libres dans des conditions plus quantitatives à l'aide du paradigme de la sémantique différentielle. Chacun des mots testés a été présenté à chacun des sujets pour évaluer chacun des sons. De plus, la théorie de "la mesure du sens" d'Osgood et al. (1957) nous a permis d'avoir un cadre théorique dans lequel interpréter une représentation sémantique de notre ensemble de sons à partir des profils sémantiques obtenus. Enfin, cela a permis de comparer directement notre étude à celles effectuées sur le timbre dans le cadre de la sémantique différentielle : celles de Solomon (1957) sur les sons complexes de sonar, celles de Kerrick et al. (1969) sur des sons complexes de musique, de voiture et des sons artificiels, celles de von Bismarck (1974a) sur un ensemble de timbres musicaux et de bruits, ainsi que celles de Kendall & Carterette (1992a ; 1992b) sur des sons d'instruments à vent joués en duos (cf. chapitre 5 § I.4). Pourtant, fort de certaines critiques de ce paradigme, notamment dues à Kendall & Carterette (1992a), nous avons tenu à actualiser ici plusieurs variantes du paradigme habituel de la sémantique différentielle.

### **II.1.2. Échelles unipolaires**

En effet, lors de leur étude sur les attributs verbaux des timbres de duos d'instruments à vent, Kendall & Carterette (1992a ; 1992b) ont montré la nécessité d'utiliser des échelles unipolaires. Ceci nous permet en effet de nous affranchir des problèmes liés à la pertinence des antonymes choisis comme extrêmes des échelles bipolaires utilisées habituellement en sémantique différentielle. En effet, lors de la définition du paradigme, Osgood et al. (1957) cherchaient à obtenir un ensemble de termes qui permettrait de tester le sens de l'ensemble de nos perceptions, à partir d'échelles du type : "chaud-froid", "blanc-noir"... Quand les chercheurs ont commencé à utiliser cette technique pour tester la perception de sons, ils ont modifié le vocabulaire testé, gardant certaines échelles telles quelles et remplaçant d'autres par des termes qui leurs semblaient plus spécifiques aux stimuli qu'ils présentaient (voir par exemple Solomon (1957) pour des sons de sonar). Le problème de la recherche d'antonymes se pose donc avec acuité : les musiciens ont par exemple l'habitude de parler de sons plus ou moins "*sec*", mais y a-t-il pour autant des sons *mouillés*, ou *humides*, ou *mous* ? De même, quel peut être l'opposé d'un son *doux* (utilisé spontanément très souvent) : un son *dur* ?, un son *agressif* ? De plus, lorsqu'on demande de positionner des objets sur des échelles bipolaires, on mesure en pratique deux choses différentes sur chaque moitié d'échelle qui ne sont pas utilisées de manière symétrique.

### **II.1.3. 23 échelles choisies**

Nous avons choisi de construire 23 échelles sémantiques unipolaires à partir des critères de sélection du vocabulaire testé dans l'expérience 1, afin que l'expérience puisse être effectuée en une séance de 1H30 environ. En effet, à partir du moment où ce vocabulaire avait été testé dans le cadre de

l'expérience 1, il était intéressant de le tester à nouveau dans un autre contexte. Ensuite, les conditions de recueil nous garantissaient sa pertinence dans le contexte précis de notre expérience puisque ce vocabulaire avait été produit spontanément (et avec le plus grand nombre d'occurrence) pour comparer les mêmes sons que ceux testés ici. Enfin, pour que cette expérience se déroule dans des conditions raisonnables, il n'était pas envisageable de tester un plus grand nombre de mots au sein d'une même expérience.

#### **II.1.4. Jugements de pertinence**

Nous avons ajouté une deuxième variante au paradigme de sémantique différentielle. En effet, étant donné que nous demandions à tous les sujets d'utiliser le même vocabulaire, nous avons jugé intéressant de recueillir leur jugement sur la pertinence pour eux du vocabulaire proposé dans le contexte donné. Ceci s'est imposé à nous du fait que l'adjectif le plus fréquent dans le contexte de la comparaison graduelle entre deux sons de notre ensemble testé s'est trouvé être "*aigu*" et que, pour des musiciens, "*plus ou moins aigu*" ne pouvait pas être pertinent pour comparer des sons égalisés en fréquence. Comme nous souhaitions proposer ce vocabulaire à des sujets de compétence musicale variée, il était important, sous peine de se faire rire au nez par les musiciens, de leur laisser la possibilité de dire que ce mot était ou non adapté pour eux. Ceci nous permettait aussi de tenter de comprendre pourquoi cet adjectif était tout de même dans les plus fréquemment utilisés : pour combien de personnes était-il pertinent ? Quelles étaient ces personnes ? Pour quel sous-ensemble de sons ce mot pouvait-il être pertinent ?

Nous avons demandé deux types de jugements de pertinence aux sujets pour chacune des 23 échelles sémantiques proposées. A la fin des jugements d'entraînement pour une échelle donnée, les sujets devaient dire si ce mot leur paraissait adapté pour juger l'ensemble des sons proposés. Puis, lors des jugements réels, après chaque positionnement d'un son sur l'échelle proposée, le sujet devait dire si ce mot était adapté ou non pour juger ce son. Toutes ces questions étaient posées à travers l'interface PsiExp de gestion de l'expérience et étaient enregistrées sous forme d'appuis sur des boutons ("*adapté*" / "*pas adapté*") à cliquer grâce à la souris de l'ordinateur.

#### **II.1.5. Groupes de sujets**

Enfin, nous avons souhaité interroger des sujets appartenant à plusieurs groupes. D'une part des sujets de compétences musicales variées (musiciens ou non-musiciens) et d'autre part des sujets "*anciens*" qui avaient participé à l'expérience 1, à partir de laquelle le vocabulaire testé avait été récolté, et des sujets nouveaux. En effet, à partir du moment où le vocabulaire a été extrait de verbalisations libres d'un groupe de 20 sujets, il pouvait être intéressant :

- d'une part de reposer systématiquement ce vocabulaire à ces sujets (groupes des "*anciens*") afin de tester une éventuelle différence de comportement entre une utilisation libre et spontanée d'un vocabulaire et une utilisation plus contrainte et systématisée de ce même vocabulaire ;
- d'autre part de tester une éventuelle différence d'utilisation de ce vocabulaire par rapport à un groupe de sujets "*nouveaux*".

Sachant que si des différences apparaissaient dans les jugements de pertinence des échelles proposées entre les deux groupes "anciens" / "nouveaux", notre hypothèse serait que les mots sont *a priori* jugés plus pertinents pour les groupes des sujets "anciens". Différentes analyses des données de cette expérience II ont fait l'objet d'une présentation orale au 4ème congrès ICMPC à Montréal (Faure et al., 1996) et d'une présentation orale au 4ème congrès de la Société Française d'Acoustique à Marseille (Faure & McAdams, 1997).

## II.2. Sujets

Trente-deux sujets ont participé à cette expérience. Seize d'entre eux avaient participé à notre première expérience, les seize autres étaient nouveaux. Nous avons numéroté ces sujets de a1 à a16 pour les "anciens", et de n1 à n16 pour les nouveaux (voir annexe A4.2). On peut noter aussi que 7 de ces sujets ont participé à notre troisième expérience de reconnaissance de portraits verbaux.

Les sujets ont été répartis dans deux groupes suivant leur compétence musicale. Dans le groupe des musiciens, nous avons rassemblé 4 instrumentistes professionnels (codés "pro"), et 12 musiciens amateurs ayant plus de 10 ans d'expérience musicale et continuant cette pratique ("am2"), dont 2 possédant de plus un métier lié à la musique ("am1"). Dans le groupe des non-musiciens, nous avons rassemblé les 5 vrais néophytes, n'ayant jamais pratiqué la musique et n'ayant aucune connaissance théorique à son propos ("néo"), ainsi que 11 musiciens amateurs ayant moins de 10 ans d'expérience musicale et ayant commencé à l'âge adulte ("am3") ou n'ayant pratiqué que dans l'enfance ("am4"). Les sujets considérés comme appartenant au groupe des "musiciens" ont été numérotés a<sub>ij</sub> ou n<sub>ij</sub> (i, j entre 1 et 16). L'âge moyen des sujets était de 27 ans. Il y avait une équirépartition des femmes et des hommes, 21 d'entre eux (65%) étaient étudiants dont 11 en thèse dans des disciplines variées. Tous avaient un niveau d'étude (musicale ou traditionnelle) élevé.

## II.3. Stimuli

Les 12 sons étudiés étaient les mêmes que ceux de l'expérience 1 (cf. Annexe A.4.3).

## II.4. Procédure

Nous avons utilisé la procédure VAME (*Verbal Attribute Magnitude Estimation*) que Kendall et Carterette (1992a) ont développée en s'inspirant de la Sémantique différentielle d'Osgood (1957). 23 échelles sémantiques unipolaires (cf. Annexe A.4.4) du type *peu aigu-très aigu* ont été présentées, sur lesquelles les sujets ont positionné chacun des 12 sons testés. Pour chacun des sujets on effectuait un tirage au sort de l'ordre de passation des 23 échelles sémantiques. Pour chaque échelle, les sujets devaient juger une première fois les 12 sons pour s'entraîner. Ils effectuaient ensuite un second jugement sur cette même échelle pour chacun des 12 sons, entrecoupé après chaque positionnement

sur l'échelle par un jugement de pertinence de l'échelle pour le son considéré. A la fin de l'utilisation de chacune des 23 échelles testées, on demandait un jugement de la pertinence globale de cette échelle par rapport à l'ensemble des sons écoutés. Enfin, un certain nombre de sujets ont indiqué, à la fin de l'expérience, les échelles qu'ils préféraient et nous avons noté les commentaires qu'ils avaient faits sur leur utilisation des différentes échelles.

## II.5. Matériel

Les auditeurs étaient assis dans une cabine audiométrique en face d'un ordinateur. L'expérience était contrôlée par le programme PsiExp\* version 2.5 développé par Bennett Smith (1995), qui contrôlait l'interface entre une console NeXT et les sons enregistrés sur le disque dur. Les stimuli étaient présentés diotiquement\* à travers un casque Sennheiser HD520 II.

## II.6. Analyses préliminaires

Nous avons effectué sur ces données brutes un certain nombre d'analyses préliminaires, ceci en vue de traiter (éventuellement) séparément les données de sujets dont les réponses s'écarteraient trop de celles du groupe et d'apporter d'éventuelles corrections à ces données brutes. Nous avons donc tout d'abord effectué des corrélations inter-sujets sur ces données brutes, ce qui nous a amené à observer certaines inversions d'échelles. Nous avons alors repris les données individuelles de chacun des sujets avec leur commentaire pour modifier les données sur lesquelles ils pensaient avoir commis des erreurs. Enfin, nous avons essayé de systématiser la recherche d'inversion d'échelle éventuelle afin de les corriger.

Les corrections effectuées finalement ont été rares : 6 échelles ont été entièrement inversées (sur les  $23 \times 32 = 736$  soit 0.8% d'entre elles) et 16 jugements de pertinence (sur  $23 \times 12 \times 32 = 8832$  soit 0.2%) ont été modifiés d'après les remarques faites par les sujets. Ces corrections améliorent la cohérence des jugements entre les sujets. On remarque aussi une amélioration de cette cohérence entre les séances 1 et 2. Ces corrélations inter-sujets sont globalement fortes si on considère l'ensemble des jugements de chaque sujet sur les 23 échelles sémantiques proposées. Elles deviennent moins bonnes dès qu'on les considère échelle par échelle (les degrés de liberté passent alors de 550 à 10). L'ensemble de ces analyses préliminaires est développé ci-dessous.

### **II.6.1. Jugements bruts**

#### Corrélations inter-sujets

Ces analyses de corrélation inter-sujets ont été effectuées sur l'ensemble des jugements bruts des 23 échelles (séance 1 et séance 2, soit  $2 \times 12 \times 23 = 552$  jugements par sujet) à l'aide du logiciel Statview 4. Les probabilités associées aux coefficients de corrélations ont été multipliées par 32 pour tenir compte de la correction de Bonferroni pour 32 analyses. Globalement, sur 496 corrélations inter-sujets, 62 (soit 13% d'entre elles) sont non significatives ( $p > .05$ ,  $r < 0.128$ ), les autres corrélations

inter-sujets se répartissent comme suit : 28 (6%) avec  $.01 < p < .05$  ; 13 (3%) avec  $.0032 < p < .01$  ; les 393 restantes (soit 79% des corrélations inter-sujets) sont significatives avec  $p < .0032$ . On peut donc considérer que les corrélations inter-sujets sont bonnes.

Pour les corrélations plus mauvaises, nous avons cherché quels étaient les sujets représentés (annexe A.4.5). Au total, 9 sujets ont des corrélations non significatives avec plus de 5 autres sujets. En fait, sur les 62 corrélations inter-sujets non significatives, 25 (soit 40% des corrélations non significatives) font intervenir le sujet n12m qui a donc 81% de ses corrélations avec les autres sujets qui ne sont pas significatives. Les autres sujets qui ont des corrélations non significatives avec d'autres le sont toujours moins de 10 fois (soit pour au maximum 32% de leurs corrélations avec les autres sujets). Ceux qui présentent le plus de corrélations non significatives sont les sujets a9, a4, n6 donc tous trois dans le groupe des non-musiciens. Ces trois sujets ajoutés au sujet n12m représentent 85% des corrélations non significatives.

Plutôt que d'éliminer les données du sujet n12m, (par ailleurs un musicien et psychoacousticien !) nous avons regardé ses données de près et les avons comparées aux données moyennes des autres sujets. Nous avons constaté que les données du sujet n12m différaient de celles du groupe de deux façons : pour un certain nombre d'échelles, les résultats avaient été manifestement inversés. Pour un autre nombre d'échelles, les données étaient globalement ordonnées de la même façon que pour la moyenne des autres sujets mais l'utilisation de l'échelle était beaucoup plus extrême : la totalité de l'amplitude proposée étant utilisée. Ainsi, par exemple sur les 23 échelles proposées, le sujet n12m utilise 11 échelles de manière très dilatée, 8 de manière inversée et seulement 6 à partir d'un critère manifestement différent de l'ensemble des autres sujets. Les échelles inversées correspondent à tous les synonymes et antonymes de *aigu* (jugé sans inversion), c'est-à-dire *bas*, *grave*, *haut*, mais aussi *riche* et *nasillard* qui correspondent eux aussi en général à une perception de hauteur spectrale (cf. chapitre 5).

Il semble donc que notre protocole puisse induire des inversions dans l'utilisation des échelles sémantiques. Comme nous avons voulu tester dans ces échelles certains termes qui sont synonymes ou antonymes les uns des autres tout en gardant pour chaque échelle la même orientation : "peu XXX" à gauche et "très XXX" à droite, un sujet qui vient de juger tous les sons entre "peu *aigu*" et "très *aigu*" par exemple peut inverser inconsciemment l'échelle suivante, si c'est par exemple entre "peu *bas*" et "très *bas*" s'il juge qu'il s'agit de la même qualité sonore et qu'il positionne les mêmes sons au même endroit. Ces observations nous ont amenées d'une part à reprendre toutes les données brutes et à essayer de les corriger, soit en considérant les remarques faites par les sujets eux-mêmes en fin de session, soit à partir d'une recherche systématique d'échelles éventuellement inversées (voir ci-dessous). D'autre part, nous avons analysé, dans la partie utilisation des échelles sémantiques (§II.6.), la répartition des jugements que les sujets ont effectués sur chacune des échelles sémantiques, l'utilisant de manière normale, dilatée ou au contraire resserrée voir très resserrée.



Certains sujets ont noté au cours de leur passation les erreurs qu'ils ont effectuées. C'est le cas par exemple du sujet n12m qui avait signalé à la fin de sa passation qu'il avait inversé certaines échelles sans préciser lesquelles de façon formelle. Ce fut aussi le cas pour d'autres sujets (voir texte en annexe A.4.6). Ainsi, sur les 32 sujets qui ont passé cette expérience, 8 ont noté avoir fait des erreurs qu'ils souhaitaient corriger en fin de séance. Il est intéressant de remarquer que 6 de ces 8 sujets étaient dans le groupe des musiciens et que les deux sujets dans le groupe des non-musiciens étaient en fait des amateurs ayant pratiqué un instrument de musique mais ne pratiquant plus. Est-ce que les musiciens passant l'expérience de manière plus rapide font plus facilement des fautes ou est-ce qu'ils ont une idée très précise de ce qu'ils veulent juger et qu'ils peuvent plus facilement corriger ces jugements si après coup ils souhaitent le modifier ?

Ces erreurs étaient globalement de deux types :

- soit une inversion du sens d'une échelle pour l'ensemble des sons jugés ou seulement une partie (par exemple une inversion pour les 4 premiers sons jugés). C'est le cas pour 5 sujets. Ces erreurs sont faciles à repérer et à corriger quand l'inversion est faite pour l'ensemble des sons ou quand elle concerne des échelles qui ont été jugées de manière similaire par la plupart des sujets (*grave, bas...*). Elle n'a pas été possible dans deux cas où l'inversion ne concernait que quelques sons et pour des échelles présentant plus de diversité de jugement inter-sujets (*doux et sourd*).
- soit une erreur dans un jugement de pertinence d'une échelle pour un ou deux sons. C'est le cas pour 4 sujets. Ces erreurs ont toutes été corrigées.

L'analyse des types d'erreurs notées par les sujets à la fin de leur passation confirme la possibilité d'inverser le sens de certaines échelles. Nous avons donc fait une recherche systématique de ces éventuelles inversions sur l'ensemble des jugements de la séance 2.

#### Inversions d'échelles

Nous avons vu que 5 des 32 sujets qui ont passé cette expérience (soit 16%) ont indiqué avoir inversé le sens d'une échelle pour au moins une partie des sons jugés, jugeant "très XXX" à gauche au lieu de le faire à droite comme cela était mentionné sur l'échelle. D'autre part, en faisant les corrélations inter-sujets sur les jugements bruts, un sujet (n12m) sortait beaucoup du groupe à cause de données seulement inversées. Nous avons donc considéré que le matériel proposé pouvait engendrer ce genre d'inversions. En effet, des échelles liées à des antonymes pouvaient alterner (par exemple *aigu* puis *grave*) ou les mêmes sons devaient donc être positionnés de manière inversée vu que le sens des échelles avait toujours la même orientation, par exemple : "peu *aigu*" à gauche et "très *aigu*" à droite, suivi de "peu *grave*" à gauche et "très *grave*" à droite. Nous avons donc cherché à dépister d'éventuelles inversions de manière systématique, afin de corriger ces données (voir texte en annexe A.4.7). Ceci nous amènera finalement à corriger les résultats de 6 échelles dont 5 pour le même sujet (n12m) et une pour le sujet a8m sur l'échelle qu'il avait signalée dans ses remarques.

Trois échelles (*attaque, clair* et *doux* ) n'avaient aucun sujet dont les jugements de la séance 2 étaient corrélés négativement avec plus de 10 autres sujets. Pour les échelles ayant moins de 10 sujets dans ce

cas, la majorité d'entre elles (*aigu, bas, brillant, fort, grave, haut, métallique, nasillard, net, pincé, résonnant, rond, sec, soufflé, sourd et vibré*), on a remarqué 6 cas où l'échelle avait manifestement été inversée (On considère qu'il y a une réelle inversion quand l'échelle est utilisée de façon suffisamment dilatée et que les sons positionnés aux extrêmes de manière "habituelle" sont inversés. On vérifie de plus qu'il y a une forte corrélation avec une dimension perceptive qui est inversée) . Dans ces 6 cas, 1 correspond à une inversion notée par le sujet et les 5 autres correspondent au même sujet qui avait indiqué en fin d'expérience qu'il s'était trompé pour quelques échelles.

Pour les 4 échelles ayant plus de 10 sujets dans ce cas (*diffus, large, long, riche*), nous avons effectué une analyse de cluster sur les sujets afin de regarder si cela est dû au fait que ce mot a peu de sens pour tous les sujets ou si plusieurs groupes de sujets se dégagent. (voir texte en annexe A.4.8 et figures en annexe A.4.9 à A.4.12)

Pour l'échelle liée à *riche*, on observe donc 4 groupes de sujets qui semblent se focaliser sur des aspects différents des sons, jugeant par exemple plutôt le caractère *aigu, cuivré, vibré* ou *clair* des sons. Pour l'échelle liée à *diffus*, à part le fait que le son CFT est unanimement considéré comme un des plus *diffus*, il est par contre difficile de dégager les critères de jugements des 3 groupes de sujets qui apparaissent. Il se dégage 4 façons de juger la qualité *large* des sons. En la considérant comme proche de *long*, en jugeant la *résonance* ou au contraire le coté *coupé* de la fin des sons, ou en jugeant les sons maintenus *constants sans évolution* soit comme très larges, soit comme peu larges selon les groupes de sujets. Enfin, pour l'échelle liée à *long*, on voit apparaître 3 groupes de sujets qui semblent juger, plutôt de la *longueur*, de la *résonance* et de l'aspect *continu* des sons.

### II.6.2. Jugements corrigés

On peut considérer qu'après toutes ces étapes préliminaires, les données brutes sont corrigées et donc prêtes à être analysées.

#### Comparaison des corrélations inter-sujets avant et après correction

Nous avons recalculé les corrélations inter-sujets effectuées sur l'ensemble des jugements d'échelles sémantiques (des séances 1 et 2). L'ensemble des résultats avant et après corrections sont rassemblés dans le tableau 4.4.

**Tableau 4.4.** Corrélations Inter-sujets : nombre et pourcentages correspondant aux différents seuils de significativité (correction selon Bonféroni pour 32 analyses). dl=550 pour les séances 1 et 2 ensembles, dl=274 pour les séances séparées.

% Corrélations I.S.		Séances 1 et 2		Séance 1	Séance 2
		brut	corrigé		
significatives	p<.0032	79	83	55	82
	.0032<p<.01	3	3	3	3
	.01<p<.05	6	5	7	5
non significatives	p>.05	13	9	36	10

Les corrections effectuées vont donc dans le sens d'une amélioration des corrélations inter-sujets. Plus que 9% d'entre elles sont non significatives, contre 13% précédemment. Cette amélioration se fait au bénéfice des corrélations très significatives ( $p < .0032$ ) vu que les pourcentages de corrélations avec des seuils de significativité intermédiaires ne changent pas avant et après correction. Si on reprend la comparaison sujet par sujet, on remarque là aussi une amélioration globale : pour 17 sujets (soit 53%) les corrections diminuent le nombre de corrélations non significatives au profit de corrélations très significatives ( $p < .0032$ ), pour 14 sujets (44%) elles ne changent rien et seulement pour un sujet, elle font passer une de ses corrélations significatives avec  $.01 < p < .05$  à non significative. Malheureusement, sur les 4 sujets qui avaient le plus grand nombre de corrélations non significatives, on n'observe d'amélioration que pour le sujet n12m, les données des sujets a4, a9 et n6 restant inchangées. Ces améliorations sont par contre spectaculaires : le sujet n12m n'a, après correction de ses données, plus que 8 de ses corrélations avec les autres sujets qui ne sont pas significatives (soit plus que 26% au lieu de 74%).

Les sujets ont tous moins de 10 corrélations non significatives avec les autres sujets (alors que le sujet n12m en avait 25) et 7 sujets ont plus de 5 corrélations non significatives avec d'autres sujets contre 9 avant correction. Les corrections effectuées sur les données ont donc globalement augmenté la cohérence inter-sujets.

#### Comparaison des corrélations inter-sujets entre les séances 1 et 2

Nous avons tenté de comparer les corrélations inter-sujets entre les séances 1 et 2. Le sujet a1 n'ayant pas de donnée pour sa séance 1, les comparaisons se font sur les 31 autres sujets sur leur  $12 \times 23 = 276$  jugements sur les échelles sémantiques. Les résultats sont rassemblés dans le tableau ci-dessus. On remarque une nette amélioration de la cohérence inter-sujets lors de la séance 2. En effet, le pourcentage de corrélations inter-sujets non significatives passe de 36% à 10% et parallèlement, celui des corrélations significatives à  $p < .0031$  passe de 55% à 82%. L'idée de faire faire systématiquement aux sujets une passation d'entraînement nous a donc permis de stabiliser les résultats.

#### Corrélations inter-sujets échelle par échelle

Nous avons repris les analyses de corrélations inter-sujets échelle par échelle à partir des jugements corrigés de la séance 2. Ces analyses sont faites à chaque fois sur les jugements d'un sujet pour une échelle pour les 12 sons, soit avec un degré de liberté,  $dl=10$ . Les résultats sont rassemblés dans le tableau de l'annexe A.4.13. Si on corrige les seuils de significativité pour 23 analyses (d'après Bonferroni), la plupart des corrélations inter-sujets ne sont pas significatives : entre 63% et 98% (pour 496 corrélations inter-sujets). La différence avec les corrélations inter-sujets calculées sur l'ensemble des jugements est sans doute liée ici au faible nombre de degrés de liberté. En effet, la plupart des échelles (16 sur 23 soit 70% d'entre elles) ont moins de 10% des corrélations inter-sujets qui sont significatives. Les 7 échelles qui ont le plus fort pourcentage de corrélations inter-sujets qui sont significatives (à  $p < .05$ ) sont celles liées à *attaque rapide - attaque lente* (37% des corrélations inter-

sujets significatives à  $p < .05$ ), puis celles liées à *pincé* (26%), *soufflé* (17%), *sec* (15%), *résonnant* (14%), *sourd* (11%) et *net* (10%).

Si on compare ces résultats à ceux sur le nombre de sujets qui, pour chaque échelle, ont leurs données corrélées négativement avec plus de 10 autres sujets, on remarque que parmi ces 7 échelles, celle liée à la notion d'*attaquerapide* était une des 3 qui ne présentaient aucun sujet dont les données étaient corrélées négativement avec celles de plus de 10 autres. *Pincé* et *sourd* étaient les deux échelles ayant un seul sujet dans ce cas. Les 4 autres échelles étant toutes avec moins de 6 sujets dans ce cas. Globalement, sur ces 7 échelles seule celle liée à *sec* à été inversée par un sujet. On peut donc émettre l'hypothèse qu'une échelle présentant une forte cohérence inter-sujets est aussi une échelle sur laquelle peu d'inversions ont lieu.

Vu le nombre faible de degrés de liberté et les corrections effectuées sur les seuils de significativité selon Bonferroni (\*23), nous nous sommes autorisés à regarder aussi les corrélations significatives à  $p < .1$ . Même si les pourcentages de corrélations inter-sujets significatives avec ce seuil augmentent, elles ne changent pas radicalement la nature des résultats et globalement l'ordre des échelles n'est pas modifié : on retrouve les 7 même échelles les plus cohérentes à travers les sujets. On retrouve aussi comme échelles faisant le moins consensus entre les sujets (moins de 10% des corrélations inter-sujets sont significatives à  $p < .1$ ) celles liées à *riche* (2% des corrélations inter-sujets significatives à  $p < .1$ ), *large* (3%), *fort* (4%), *vibré* et *diffus* (5%), *grave* et *rond* (7%) et *brillant* et *nasillard* (8%). On retrouve donc 3 des 4 échelles (*riche*, *diffus* et *large*) présentant plus de 10 sujets dont les données étaient corrélées négativement avec celles de plus de 10 autres sujets. Ces échelles ne semblent donc pas présenter de consensus dans leur utilisation par les sujets pour juger des sons proposés. Les cas des 6 autres échelles sont plus difficiles à interpréter, le nombre de sujets dont les données sont corrélées négativement avec plus de 10 autres variant entre 2 et 8.

## II.7. Utilisation des échelles sémantiques

Si nous avons considéré comme important de discuter de l'utilisation des échelles par les sujets, et si nous choisissons de le faire à cette place (c'est à dire avant toute analyse des résultats des positionnements sur les échelles sémantiques) c'est parce qu'en général, dans ce genre de paradigme inspiré de la sémantique différentielle (Osgood et al., 1957), le choix des échelles sémantiques est fait après consultation d'experts préalablement à l'expérience et n'est plus questionné ultérieurement. Ce questionnement, lié aux mesures directes de pertinences demandées aux sujets (pertinence globale de chaque échelle sémantique proposée par rapport au contexte de l'expérience et pertinence de chaque échelle pour chacun des sons jugés) qui seront analysées dans le chapitre 7, est donc une des originalités de notre démarche.

Dans ce paragraphe et le suivant, nous allons analyser ce que nous pourrions considérer comme des mesures indirectes de cette pertinence : la manière resserrée ou dilatée dont un sujet a utilisé chacune

des échelles proposées et le nombre de re-écoutes de chacun des sons jugés dans chacune des situations de jugement. Ces résultats seront comparés aux jugements de pertinence des échelles, analysés dans le chapitre 7.

En effet, lors de l'analyse des corrélations inter-sujets (§II.5), nous avons remarqué qu'une des raisons qui entraînait parfois de faibles corrélations inter-sujets est que les sujets n'utilisaient pas toutes les échelles proposées de façon discriminante entre les sons. Dans certains cas, ils choisissaient de positionner tous les sons au même endroit d'une échelle sémantique. Bien que ce type de comportement ait un sens, que nous allons essayer de dégager dans ce paragraphe, il ajoute du bruit quand on considère des jugements moyens sur l'ensemble des sujets. Nous allons donc essayer de décrire les différents types d'utilisations des échelles sémantiques et proposer plusieurs modes de calculs de moyennes sur ces données. Afin d'appréhender ce phénomène, nous allons commencer par en présenter un exemple, puis nous en ferons une description plus systématique.

#### II.7.1. Un exemple : l'échelle *peu aigu* - *très aigu*

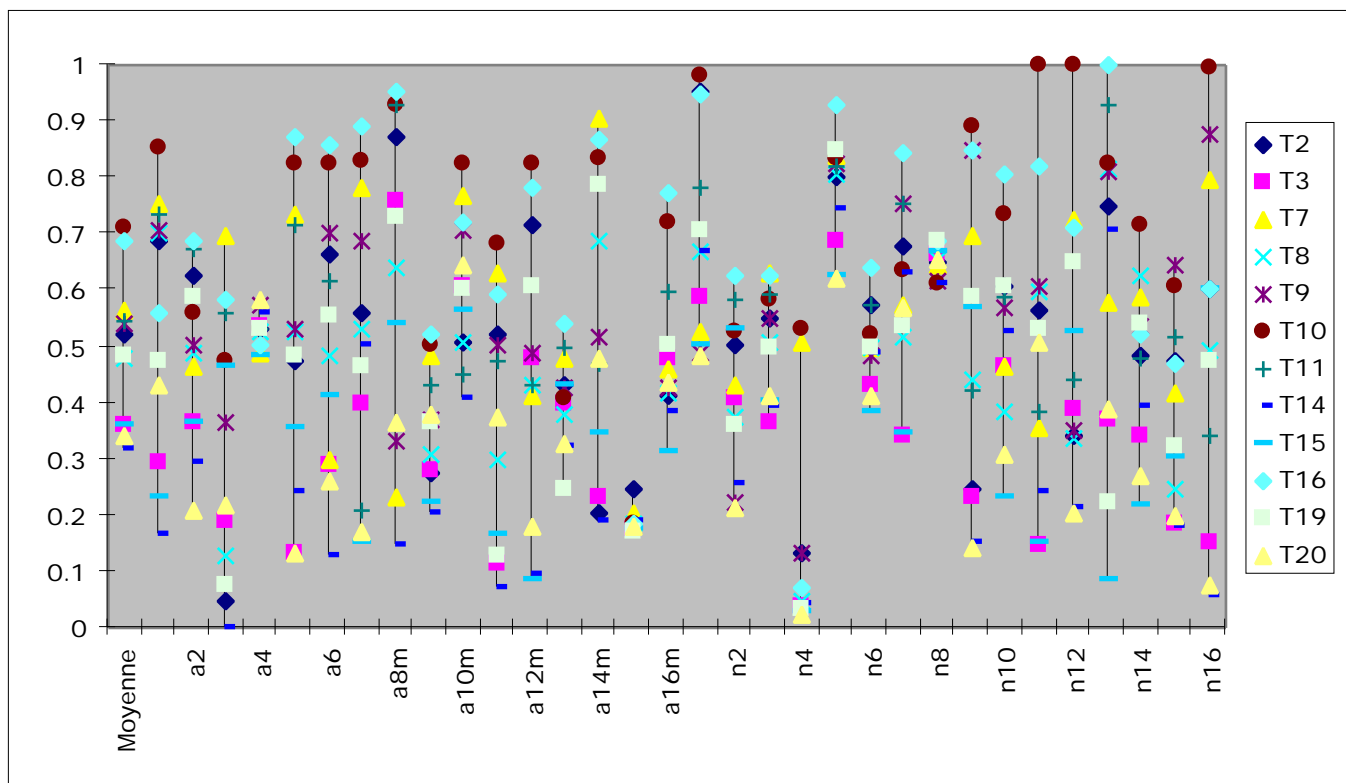
La figure 4.1 présente le positionnement sur l'échelle sémantique liée à *aigu* des 12 sons par les 32 sujets, ainsi que la moyenne effectuée sur l'ensemble des sujets. Ces jugements sont codés entre 0 (pour l'extrémité de l'échelle *peu aigu*) à 1 (pour l'extrémité *très aigu*). On observe immédiatement d'énormes différences de comportement entre les sujets, deux extrêmes pouvant être caractérisés par exemple par l'utilisation du sujet a4 (tous les sons positionnés autour de la valeur 0.55) et du sujet n16 (les sons sont répartis entre les valeurs 0.1 et 1), le sujet n6 présentait un comportement intermédiaire (jugements répartis entre 0.4 et 0.6).

On observe donc que, même si la plupart des sujets jouent le jeu de répartir les sons sur l'échelle sémantique proposée, pour l'échelle liée à *aigu*, 3 sujets ont mis la même valeur pour tous les sons. Ceci peut s'expliquer si ces sujets ont jugé sur cette échelle la hauteur tonale des sons (qui ne varie pas : tous les sons sont joués sur la même note) considérant le mot *aigu* dans son acception la plus courante en musique et si les autres sujets ont jugé la distribution spectrale (liée au Centre de gravité du spectre CGS\*, voir chapitre 5), la seule acception permettant de différencier les sons dans notre contexte.

De plus, parmi les trois sujets qui utilisent l'échelle liée à *aigu* de manière très resserrée (en positionnant tous les sons au même endroit) il y a une musicienne (le sujet a15m) et deux sujets appartenant au groupe des non-musiciens (les sujets n8 et a4). Ces sujets n'ont d'ailleurs pas regroupé les sons au même endroit, le sujet a15m jugeant tous les sons relativement *peu aigu* (moyenne  $m=0.19$ , écart type  $e=0.02$ ), le sujet a4 : *plutôt aigu* ( $m=0.53$ ,  $e=0.03$ ) et le sujet n8 *assez aigu* ( $m=0.65$ ,  $e=0.03$ ). Il est intéressant de remarquer que ces trois sujets avaient aussi indiqué que pour eux cette échelle n'était pas pertinente dans le contexte de notre expérience, le sujet n8 ajoutant même comme commentaire : "pas adapté pour les distinguer les uns des autres, comme *haut*", et sujet a15m : "il ne s'agit pas de fréquence mais de 'l'angulosité' du son qui serait qualifiable pour moi de

‘pointu’”. Ce choix dans l'utilisation de l'échelle a donc un sens qui peut être lié à la pertinence de cette échelle pour le sujet considéré.

Figure 4.1. Répartitions des jugements des sujets sur l'échelle sémantique *peu aigu - très aigu*



Nous avons essayé de décrire systématiquement l'utilisation par les sujets des échelles sémantiques proposées. Il sera intéressant de comparer ces utilisations avec les jugements de pertinence de chacune des échelles effectués par les sujets (cf. chapitre 7).

### II.7.2. Recherche systématique

Il nous est apparu qu'on pouvait classer les sujets dans différents groupes en fonction de leur utilisation des échelles sémantiques proposées : de très resserrée à très dilatée. Le critère proposé étant la valeur de l'écart type de la moyenne de ses jugements sur l'échelle considérée. Nous avons donc considéré de manière empirique des valeurs seuil des écarts types : l'utilisation de l'échelle étant considérée comme très resserrée si la valeur de l'écart type est inférieure à 0.05 ( $e < .05$ ), resserrée si  $.05 < e < .1$ , normale si  $.1 < e < .2$  et dilatée si  $e > .2$ . Les résultats sont rassemblés dans le tableau 4.5. ci-dessous.

Même si la plupart des sujets (plus de 75%) utilisent ces échelles de manière normale ou dilatée, le fait d'utiliser une échelle de manière resserrée (voir très resserrée) concerne plusieurs sujets et plusieurs échelles. C'est le cas pour au moins un sujet pour les 13 échelles liées à *aigu*, *grave*, *haut*, *soufflé*, *fort*, *bas*, *long*, *riche*, *vibré*, *rond*, *large*, *résonnant* et *sec* (soit 56% des échelles).

Tableau 4.5. Nombre de sujets (sur 32) ayant utilisé chaque échelle de façon :

très resserrée (Écart type  $e < .05$ ), serrée ( $e < .1$ ) normale ( $.1 < e < .2$ ) et dilatée ( $e > .2$ ).

Échelle	très serrée	serrée	normale	dilatée	Total $e < .1$	Total $e > .1$
aigu	3	4	11	14	7	25
attaque			11	21	0	32
bas	2		20	10	2	30
brillant			16	16	0	32
clair			12	20	0	32
diffus			15	17	0	32
doux			11	21	0	32
fort	2	6	13	11	8	24
grave	3	4	14	11	7	25
haut	3	5	11	13	8	24
large	1	1	16	14	2	30
long	2	3	15	12	5	27
métallique			11	21	0	32
nasillard			12	20	0	32
net			11	21	0	32
pincé			7	25	0	32
résonnant	1		16	15	1	31
riche	1	4	14	13	5	27
rond	2	1	8	21	3	29
sec	1		7	24	1	31
soufflé	3	4	8	17	7	25
sourd		2	11	19	2	30
vibré	1	3	10	18	4	28

On retrouve ici les 4 échelles liées directement à des termes décrivant en général la hauteur tonale\* du son : *aigu*, *grave*, *haut* et *bas*. Pourtant, ce ne sont pas toujours les mêmes sujets qui utilisent ces échelles de manière très resserrée. En effet, si pour l'échelle liée à *aigu*, de même que pour l'échelle liée à *haut*, ce sont les mêmes sujets (a4, a15m et n8) qui l'utilisent de manière très resserrée, seuls 2 de ces 3 sujets (a4 et a15m) utilisent aussi les échelles liées à *bas* et à *grave* de manière très resserrée, le sujet n8 les utilisant part contre de façon respectivement dilatée et normale.

Par contre, sur les 4 échelles considérées ici, le sujet a5 n'utilise que l'échelle liée à *grave* de manière resserrée. Dans ses remarques, le sujet a5 (qui est un vrai non-musicien : sans aucune formation musicale) note d'ailleurs pour l'échelle liée à *grave* : "grave = "octave" ici pas différent, alors que pour l'échelle liée à *aigu* qu'il a utilisée de manière dilatée, il notait "pas adapté sauf si subjectif", il n'a pas fait de remarque pour l'échelle liée à *haut* (utilisée aussi de manière dilatée) mais pour l'échelle liée à *bas*, il note : "= étouffé". On remarque donc que, pour les sujets n8 et a5, les échelles liées à *aigu* et *haut* ne sont pas utilisées comme des antonymes des échelles liées à *grave* et *bas*. Ceci est un argument supplémentaire en faveur de protocoles avec des échelles unipolaires plutôt que bipolaires. Enfin, on peut remarquer que l'utilisation resserrée de ces échelles n'est pas propre aux musiciens.

Pour l'échelle liée à *soufflé*, 3 sujets l'utilisent de manière très resserrée (on retrouve les sujets a15m, n8 auxquels s'ajoute le sujet n12m) et 4 de manière resserrée (les sujets a3, a10m, n2m et n16).

Il est là aussi intéressant de remarquer que les sujets ne se comportent pas de la même façon avec cette échelle qu'avec celle liée à son opposé en matière de mode de fabrication des sons : *pincé*. En effet, l'échelle liée à *pincé* est celle qui recueille le plus d'utilisations de façon dilatée. L'asymétrie vient peut-être aussi d'une asymétrie dans la synthèse des sons : les sons imitant des sons produits en pinçant (ou frappant) paraissent beaucoup plus vraisemblables que les sons imitant des sons émis en soufflant, qui ont un caractère beaucoup plus synthétique.

L'échelle liée à *fort*, avec celle liée à *haut*, présente le plus de diversité au niveau de son utilisation. Parmi les sujets, 2 l'utilisent de manière très serrée, 6 serrée, 13 normale et, 11 dilatée. C'est une échelle qui a priori concerne une qualité non pertinente pour discriminer les sons qui ont été égalisés en sonie\*. Il est donc assez normal qu'elle fasse partie des échelles utilisées de manière peu discriminante. Pourtant ce mot faisait partie des plus souvent utilisés pour comparer les sons de manière spontanée, malgré le fait que la corrélation inter-sujets était la plus mauvaise justement pour cette échelle (cf. §II.5) Restera à comprendre ce que les sujets ont bien pu juger sur cette échelle.

L'interprétation pour les autres échelles est sans doute à nuancer vu que les utilisations avec un écart-type inférieur à 1 correspondent chaque fois à moins de 5 sujets.

Les échelles utilisées de façon normale ou dilatée permettent de bien répartir les sons présentés et doivent donc être considérées comme particulièrement pertinentes, d'autant que tous les sujets les utilisent ainsi. C'est le cas des échelles liées à *pincé*, *attaque*, *métallique*, *net*, *doux*, *clair*, *nasillard*, *diffus* et *brillant*. Nous verrons chapitre 7 le lien que cela peut avoir avec la pertinence de ces échelles pour les sujets qui les utilisent ainsi.

### II.7.3. Répartitions en fonction des sujets

Il y a très souvent plus de non-musiciens que de musiciens qui utilisent les échelles de manière très large (sur 22 échelles utilisées de façon large, 18 le sont par des non-musiciens). Par contre, pour l'utilisation très resserrée : sur 11 échelles concernées, 5 sont plus utilisées de cette façon par les non-musiciens et 6 par les musiciens. Ce comportement n'est donc pas spécifique d'un type de compétence musicale. Il nous a paru intéressant de vérifier si les différentes manières d'utiliser les échelles étaient une particularité de certains sujets ou si ce comportement était partagé par un grand nombre d'entre eux. Nous avons donc calculé, pour chaque sujet, le nombre d'échelles qu'il avait utilisées de façon très resserrée à dilatée. Les résultats sont rassemblés dans le tableau de l'annexe A.4.14 .

Sur les 32 sujets, 9 d'entre eux n'utilisent les échelles que de manière normale ( $.1 < e < .2$ ) ou dilatée ( $e > .2$ ), ce qui correspond *a priori* à la consigne implicite. En effet, si on demande à des sujets de positionner des sons sur des échelles sémantiques, il est implicite qu'on leur demande de les "répartir" sur l'échelle. Cela reste d'ailleurs le comportement le plus fréquent quant à l'utilisation des échelles sémantiques. En effet, 30 des 32 sujets utiliseront au moins une fois une échelle de manière normale et seul un sujet n'utilisera aucune échelle de manière dilatée. Et même pour les sujets qui utilisent au moins une fois une échelle de manière très resserrée ou resserrée, ce sera pour un nombre faible



d'échelles (en moyenne 2 sur 23), alors qu'en moyenne chaque sujet utilise 21 échelles de manière normale ou dilatée. Seul un sujet (a15m, une musicienne professionnelle) utilisera 10 des 23 échelles de manière resserrée ou très resserrée ( $e < .1$ ). En fait seuls 12 sujets (soit 37% d'entre eux) utiliseront au moins une fois une échelle de manière très resserrée (et en général, sauf pour le sujet a15m pour uniquement 1 ou 2 échelles). Ceci peut paraître un pourcentage faible, c'est pourtant un comportement qui n'était ni attendu ni proposé aux sujets et dont le sens est donc à prendre en compte.

#### II.7.4. Comportement de la moyenne

Pour chaque échelle, nous avons aussi regardé la manière dont la moyenne, calculée sur l'ensemble des 32 sujets, répartissait les 12 sons (voir le tableau de l'annexe A.4.14). En effet, sur les 23 échelles, la répartition des sons résultant d'un calcul de leur positionnement moyen pour les 32 sujets, la grande majorité (17 sur 23 soit 74%) se retrouve utilisée de manière normale, aucune ne se retrouve utilisée de manière très resserrée, 5 seulement de manière resserrée ( $.05 < e < .1$ ), et 1 de manière dilatée. Il est donc intéressant de se pencher sur les échelles qui se répartissent différemment des autres.

L'échelle dont les positions moyennes des sons se répartissent sur l'échelle de manière dilatée est celle liée à *pincé* (moyenne  $m = .41$ , écart-type  $e = .22$ ). Ceci n'est pas très étonnant, vu que c'est aussi l'échelle que le plus grand nombre de sujets (78%) a utilisée de manière dilatée, aucun ne l'ayant utilisée de manière très resserrée ou resserrée. C'est aussi une des échelles ayant les meilleurs corrélations inter-sujets (cf. §II.5). On peut donc considérer que c'est une échelle qui permet facilement de bien discriminer nos 12 sons, de façon cohérente pour les sujets.

Parmi les échelles dont les positionnements moyens des sons se répartissent sur l'échelle de manière resserrée, on trouve celles liées à *riche* ( $m = .45$ ,  $e = .06$ ), *long* ( $m = .49$ ,  $e = .06$ ), *large* ( $m = .45$ ,  $e = .07$ ), *grave* ( $m = .40$ ,  $e = .09$ ) et *fort* ( $m = .46$ ,  $e = .09$ ). Toutes ces échelles étaient parmi celles qui faisaient le moins consensus entre les sujets : moins de 10% des corrélations inter-sujets significatives à  $p < .1$ , sauf pour *long* pour laquelle 11% des corrélations inter-sujets sont significatives à  $p < .1$  (cf. II.5). On retrouve aussi parmi elles les échelles liées à *fort* et *grave* qui sont considérées comme globalement pertinentes par peu de sujets (voir chapitre 7) (respectivement 23 et 40%), les échelles liées à *large*, *long* et *riche* n'étant elles aussi pertinentes globalement que pour respectivement 52, 52 et 55% des sujets.

#### II.7.5. Différents calculs de moyennes

Ces différences dans l'utilisation des échelles sémantiques nous ont conduits à calculer plusieurs types de moyennes sur ces données (voir chapitre 5 II.). En effet, en s'inspirant de calculs de moyennes effectués uniquement sur les jugements considérés comme pertinents par leurs auteurs, il nous a paru intéressant d'envisager, en plus d'une moyenne de l'ensemble des jugements calculée "normalement", une moyenne ne prenant pas en compte les jugements des échelles utilisées comme très resserrées et calculée après avoir normalisé la plage de variation des échelles utilisées de manière resserrée, normale et dilatée par les sujets. De plus, nous testerons dans le chapitre 7 l'hypothèse que les échelles ayant

des utilisations très resserrées correspondent à des mots qui ne sont pas pertinents dans le contexte pour différencier les sons.

## II.8. Analyse du nombre de re-écoute des sons

Lors de l'expérience, les sujets pouvaient re-écouter les sons librement, autant de fois qu'ils le souhaitent, pour chacun des jugements qui leur étaient demandés. Le nombre de répétitions pour chaque essai (un sujet, un son, une échelle, une séance) a été enregistré grâce à l'interface PsiExp\*.

### **II.8.1. Une mesure des difficultés rencontrées ?**

Si les sujets sont libres de re-écouter les sons proposés comme ils le souhaitent, on peut supposer qu'ils le font quand ils en ont besoin et que ces besoins sont liés à une difficulté pour émettre le jugement qui leur est demandé. Ces difficultés peuvent être liées :

- soit à des causes non intentionnelles, par exemple à une baisse passagère d'attention, qui ne nécessite pas d'interprétation particulière ;
- soit à une réelle augmentation de la difficulté, à la nécessité de s'assurer ou de confirmer le jugement. Ceci peut au contraire être interprété en terme de difficulté liée à une échelle sémantique particulière (retrouve-t-on dans les échelles qui ont engendré plus de répétitions des sons celles qui ont été jugées comme pertinentes, ou non pertinentes ? celles qui ont bénéficié d'une forte ou d'une faible cohérence inter-sujets ?), ou à un type de son particulier (est-ce que les hybrides, sons absolument non familiers sont écoutés plus souvent ? dans quelles situations ?).

De plus, une analyse du nombre d'écoutes des sons peut nous donner des informations permettant de comparer les comportements de différents groupes de sujets : est-ce que les non-musiciens écoutent plus les sons que les musiciens avant d'effectuer leur jugement ? Même si nous ne pensons pas que ces données soient à interpréter en tant que telles, elles peuvent être reliées et comparées à d'autres pour confirmer ou infirmer certaines tendances.

### **II.8.2. Description générale**

En moyenne sur l'ensemble des sujets et l'ensemble de l'expérience, il y a 0.44 répétitions par jugement, soit en moyenne pour chaque échelle sémantique jugée (12 jugements), 5 sons sont répétés. En moyenne, les sujets re-écoutent donc peu les sons (tous les sons ne sont pas re-écoutés). Il faut dire que même s'ils ne les re-écoutent pas volontairement, ils auront à écouter 46 fois chacun des 12 sons (puisque'ils ont 46 jugements à faire pour chacun de ces 12 sons), ce qui porte à 552 le nombre d'écoutes obligatoires, même si aucun son n'est re-écouté ! Il serait intéressant de vérifier si ce comportement de re-écoute des sons ne varie pas au cours du temps lors de la passation, les sons et les jugements demandés devenant de plus en plus familiers.

### **II.8.3. Différences inter-séances**

Notre hypothèse est que pour chaque échelle, les sujets sont plus familiarisés avec la tâche après la séance 1 et donc devraient moins re-écouter les sons lors de la séance 2. On remarque une différence de comportement entre les deux séances qui va dans le sens inverse de notre hypothèse : en moyenne sur l'ensemble des sujets, il y a 0.36 répétitions lors de la passation d'entraînement (séance 1), soit 4 des 12 sons répétés en moyenne, et 0.53 répétitions lors de la passation de l'expérience (séance 2), soit 6 des 12 sons répétés en moyenne. Ces différences sont très significatives ( $\chi^2(1)=323$ ,  $p<.0001$ ) et pour 19% des sujets l'augmentation observée à la séance 2 est supérieure à l'écart-type lié à la moyenne de la séance 1.

Contrairement à l'hypothèse d'une plus grande familiarité avec les sons et la consigne, on observe donc en moyenne une augmentation du nombre de répétitions lors de la séance 2. Ceci est vrai individuellement pour 61% des sujets qui écoutent les sons plus souvent lors de la séance 2 que lors de la séance 1. Ceci pourrait être lié au fait que la passation de la séance 1 est présentée comme un entraînement alors que les jugements effectués lors de la séance 2 sont définitifs : les sujets pourraient re-écouter les sons plus souvent afin d'être plus sûr de leur jugement lors de la deuxième passation.

#### **II.8.4. Différences inter-sujets**

De manière générale, les différences inter-sujets sont très grandes et très significatives. Le nombre total de re-écoutes sur l'ensemble de l'expérience varie en effet entre 3 pour le sujet a13m (un musicien professionnel ayant participé à notre première expérience) et 1395 pour le sujet n6m (lui aussi musicien et acousticien). Une analyse de cluster sur les sujets ne fait pas ressortir de groupes très saillants, en particulier elle ne permet pas de départager les musiciens des non-musiciens. Pourtant, en moyenne, les musiciens écoutent plus les sons que les non-musiciens ( $\chi^2(1)=311$ ,  $p<.0001$ ). En moyenne, par échelle, les musiciens re-écoutent 6 sons alors que les non-musiciens en re-écoutent 4. Cette différence est encore plus marquée lors de la séance 2 où les musiciens re-écoutent en moyenne 8 sons pour chaque échelle alors que les non-musiciens n'en re-écoutent en moyenne que 5.

Cela va à l'encontre de notre intuition initiale que plus l'expérience paraîtrait facile, moins on observerait de répétitions, considérant qu'*a priori* les jugements proposés seraient plus faciles à réaliser par des musiciens.

#### **II.8.5. Différences inter-échelles**

On peut supposer que s'il y a des différences de comportement en fonction des échelles sémantiques, c'est en liaison avec la pertinence d'une échelle : est-ce que les échelles plus pertinentes (plus évidentes ?) sont jugées en écoutant moins souvent les sons ? Est-ce qu'au contraire, plus une échelle est pertinente pour comparer les sons, plus la comparaison sera faite en vérifiant chacun des jugements, donc avec un plus grand nombre de re-écoutes ?

Les différences en terme de nombre de re-écoutes entre les échelles, sont significatives entre les deux séances, et ce, quel que soit le groupe de sujet considéré : tous les sujets, les musiciens ou les non-musiciens (pour tous les sujets :  $\chi^2(22)=55.27, p=.0001$ ). On a toujours un plus grand nombre de re-écoutes lors de la deuxième passation mais les répartitions sur les différentes échelles sont différentes (voir tableau en annexe A.4.15). Les différences de comportement en matière de re-écoute des sons sont significativement différentes selon les échelles sémantiques considérées. En effet, les  $\chi^2$  calculés sur le nombre total de re-écoutes effectuées sur tous les groupes de sujets considérés et pour chacune des échelles sémantiques sont tous très élevés et très significatifs (on a toujours  $\chi^2(1)>111$  et  $p<0.0001$ ). Nous analyserons donc quelles sont les échelles qui ont motivé plus ou moins de re-écoutes que la moyenne, d'abord pour la séance 2 (celui qui était considéré comme l'expérience proprement dite) et nous comparerons ensuite avec les résultats pour la séance 1.

Pour la séance 2, si on considère l'ensemble des sujets, les échelles qui ont motivé un plus grand nombre de re-écoutes que la moyenne sont celles liées à *brillant*, *vibré*, *clair* et *pincé*. On retrouve parmi celles-ci *vibré* qui est une des échelles ayant la plus mauvaise corrélation inter-sujets et *pincé* qui est une des échelles ayant la plus forte corrélation inter-sujets ! Il est aussi intéressant de remarquer que l'échelle liée à *brillant* est celle qui a engendré le plus de re-écoutes, quel que soit le groupe de sujets considéré, alors que c'est un terme utilisé fréquemment par les musiciens pour définir le timbre de sons d'instruments de musique, et c'était le terme choisi par les experts pour désigner la dimension 1 de l'espace de Krumhansl (1989). La question là aussi est de savoir si ce comportement est dû à une difficulté ressentie par les sujets ou à leur volonté de le rendre le plus précis et fiable possible. Les échelles ayant au contraire engendré le moins de re-écoutes des sons sont celles liées à *haut*, *bas* et *aigu*. Ces échelles font toutes parties de celles liées à des termes renvoyant habituellement à la notion de hauteur tonale et souvent considérées comme non-pertinentes dans notre contexte où ce paramètre a été égalisé.

Toujours pour la séance 2, on considère les différences selon le groupe de sujets : musiciens ou non-musiciens. On retrouve, pour le groupe des musiciens, les échelles liées à *brillant* et *clair* (respectivement *haut* et *bas*) parmi celles qui ont engendré le plus (respectivement le moins) de re-écoute. La nouveauté pour le groupe des musiciens est de re-écouter beaucoup de sons lors des jugements de l'échelle liée à *riche* mais pas pour les échelles liées à *vibré* et *pincé*. On peut noter que *riche* est aussi un terme fréquemment employé par les musiciens pour qualifier un son. Par contre dans les échelles qui ont engendré moins de répétitions que la moyenne, on trouve aussi les échelles liées à *rond* et *fort*, mais pas celle liée à *aigu*, qui engendre en moyenne plus de re-écoutes chez les musiciens que chez les non-musiciens. Dans le cas de l'échelle liée à *aigu*, ceci est sans doute à mettre en relation avec le fait que le sens de ce mot a pour les musiciens un sens précis lié conventionnellement à la notion de tonie\* non pertinente ici. Il se peut que de demander à des musiciens de juger des sons égalisés en hauteur sur une échelle peu aigu - très aigu, les ait laissés perplexes et leur ait demandé de répéter beaucoup de sons.

Pour le groupe des non-musiciens, on retrouve ici aussi les échelles liées à *vibré*, *brillant* et *pincé* (respectivement à *aigu*) parmi celles qui ont engendré le plus (respectivement le moins) de re-écoutes. On trouve en sus les échelles liées à *métallique* (respectivement à *doux*, *nasillard* et *grave*).

Pour la séance 1, on retrouve le même comportement de l'ensemble du groupe de sujets pour les échelles liées à *brillant*, *vibré* et *clair* (engendrant plus de re-écoutes) et celles liées à *aigu* et *bas* (engendrant moins de re-écoutes dans les séances 1 et 2). Le fait d'avoir à refaire le même jugement deux fois de suite ne modifie donc pas leur particularité dans le nombre de re-écoutes des sons. On observe par contre des différences pour les échelles suivantes :

- L'échelle liée à *riche* engendre plus de répétitions à la séance 1 mais pas à la séance 2 (sauf pour le groupe des musiciens qui maintiennent un grand nombre de re-écoutes).
- Au contraire l'échelle liée à *pincé* engendre beaucoup de re-écoutes seulement lors de la séance 2.
- De même, les échelles liées à *métallique*, *sec* et *long* engendrent peu de re-écoutes lors de la séance 1 alors que ce nombre de re-écoutes rejoint la moyenne lors de la séance 2.
- A l'inverse, le nombre de re-écoutes de l'échelle liée à *haut*, dans la moyenne lors de la séance 1 devient faible lors de la séance 2, peut être en liaison avec le comportement de l'échelle liée à *aigu*.

Il y a donc des différences très grandes dans le nombre de re-écoutes en fonction des échelles jugées, même si ces différences sont parfois difficiles à interpréter.

#### II.8.6. Différences inter-sons

Comme dans cette expérience, nous avons proposé aux sujets soit des sons imitant des sons d'instruments de musique, soit des hybrides, construits comme des mélanges de deux sons d'instruments de musique différents. On s'attend à ce que les sons hybrides, moins familiers, soient plus ré-écoutés que les sons imitant des instruments de musique. Et ceci d'autant plus que l'échelle jugée touchera à une des qualités sonores modifiée lors de l'hybridation. Nous avons totalisé le nombre de re-écoutes de chacun des sons pour chaque séance et sur l'ensemble des 32 sujets. Afin de voir si certains sons étaient plus souvent répétés que d'autres, nous avons calculé les moyennes et écart-types de ces nombres de répétitions ainsi que les  $\chi^2$  pour tester la significativité de ces différences inter-sons. Les résultats sont rassemblés dans le tableau 4.6.

Tableau 4.6. Nombre de re-écoutes pour chacun des 12 sons proposés pour chaque séance et au total.

	m	e	$\chi^2$	p	TPT	TBN	VBR	SNO	PNE	CVC	CRA	CLA	VBN	OBC	CFT	PNO
Run 1	259	19.6	2.32	0.13	245	<b>231</b>	251	<b>308</b>	241	259	272	252	275	260	258	259
Run 2	392	43.4	52.94	2.10 <sup>-7</sup>	425	408	<b>325</b>	<b>491</b>	357	396	394	<b>339</b>	417	389	372	388
Total	651	57.9	56.57	4.10 <sup>-8</sup>	670	639	<b>576</b>	<b>799</b>	598	655	666	<b>591</b>	692	649	630	647

Avec m : moyenne, e : écart type, en gras : les valeurs s'écartant de la moyenne de plus d'un écart-type

On remarque que pour la séance 1 les différences inter-sons dans les nombres de re-écoutes ne sont pas significatives ( $\chi^2(1)=2.32$ ,  $p=.13$ ) alors qu'elles sont au contraire très significatives pour la séance 2 ( $\chi^2(1)=52.94$ ,  $p=2.10^{-7}$ ). On peut expliquer cette différence en considérant que lors de la

séance 1 d'entraînement (où les sujets re-écoutent en général moins les sons) le jugement est fait de façon plus "automatique" : les re-écoutes se font au hasard, à l'occasion d'une baisse d'attention momentanée par exemple. Au contraire lors de la séance 2, les re-écoutes ont lieu sur des sons qui posent plus de difficulté de jugement. On remarque ainsi que le son qui est le plus souvent ré-écouté (nombre de re-écoutes supérieur à la moyenne de plus d'un écart-type) est le son SNO qui est l'hybride appelé *striano* et imitant un mélange de corde frottée et de piano.

Notre hypothèse que parmi les sons les plus ré-écoutés on trouverait des hybrides est bien confirmée. Un son hybride étant moins familier, plus étrange, n'induisant *a priori* pas de référence à un instrument de musique précis sera jugé uniquement à partir de critères perceptifs (et non pas en y ajoutant des connaissances sur le type d'instrument de musique auquel il renvoie par exemple). Les sujets n'ayant pas de comportement pré-appris, ils devraient être obligés de se poser la question juste à partir de leur perception. Les deux sons les moins souvent ré-écoutés sont les sons VBR et CLA. Il est plus difficile d'interpréter ce résultat. En effet, alors que le son VBR est un de ceux les plus facilement identifiables, le plus *doux* et *agréable*, et fait partie des sons les mieux imités comme la plupart des sons percussifs, le son CLA est un son souvent considéré comme très artificiel, comme la plupart des sons imitant des sons soufflés.

Les différences entre les nombres de répétitions des différents sons peuvent aussi dépendre de l'échelle sémantique jugée. En effet, si la qualité liée à une échelle donnée concerne quelque chose de délicat pour un ou deux sons, ceci pourrait justifier une re-écoute de ces sons plus que d'autres. Les répartitions du nombre de répétitions des sons sur les 12 sons proposés varient donc beaucoup selon l'échelle sémantique considérée (voir tableau en annexe A.4.16). En effet, pour les 5 échelles liées à *attaque rapide*, *bas*, *haut*, *long* et *riche*, on observe en moyenne sur l'ensemble des 32 sujets une équipartition du nombre de répétitions des sons sur les 12 sons proposés. En effet, les différences entre les sons ne sont significatives pour aucune des 2 séances et aucun des groupes de sujets considérés. On retrouve, parmi ces échelles, à la fois celles bénéficiant de la plus forte et de la plus faible cohérence de jugements des sujets (respectivement l'échelle liée à *attaquerapide* et à *riche*, voir §II.5).

A l'inverse, pour aucune échelle les différences entre les nombres de répétitions de chaque son sont significatives dans tous les cas considérés. Par contre, les cas où cette différence est significative sur le groupe constitué par l'ensemble de tous les sujets et pour les deux séances (1 et 2) concernent 4 échelles, celles liées à *soufflé*, *diffus*, *clair* et *grave*, sachant que cette différence n'est très significative que dans le cas de l'échelle liée à *soufflé*. Cette différence est due essentiellement aux non-musiciens pour l'échelle liée à *clair*, aux musiciens pour l'échelle liée à *diffus*, et cela dépend des séances pour les échelles liées à *grave* et *soufflé* : dans les deux cas, on a des différences dues aux non-musiciens dans la séance 1 et aux musiciens dans la séance 2.

Tableau 4.7. Sons les plus souvent ré-écoutés (+) et moins souvent ré-écoutés(-) par l'ensemble des sujets (T), les musiciens (M) ou les non-musiciens (NM), pour les échelles où les différences de répartition des re-écoutes sont significatives.

	run 1						run 2					
	T		M		NM		T		M		NM	
sons ré-écoutés	+	-	+	-	+	-	+	-	+	-	+	-
Soufflé	VBN	PNO			VBN		TPT	VBR, OBC	TPT, CFT			
diffus	OBC	CLA, TPT, VBR	OBC	CLA			TBN	OBC, TPT	TBN	OBC, TPT		
clair	VBR, TPT	TBN, PNE			VBR, VBR		TPT, SNO	VBR			TPT, PNO	
grave	CFT, CLA	TBN			CFT		TPT		TPT			
net	PNO		PNO, CRA	PNE								
aigu	SNO, CFT		SNO, CFT	VBR								
fort	TPT, OBC				OBC, TPT							
résonnant	PNE, SNO	VBR			PNE, TBN							
sourd	VBR, VBR	CVC	VBR		VBR, PNE, SNO	TBN						
nasillard							VBR	CFT	VBR, SNO			
vibré							CVC, VBR	TBN, CRA			CVC, PNE	
pincé							VBR, OBC	CVC, CRA			VBR, OBC	CRA, CLA
ùmétallique											CRA, TPT	TBN, CFT

Pour l'échelle liée à *soufflé*, les sons les plus souvent ré-écoutés sont le son VBR à la séance 1, du fait des non-musiciens, et le son TPT à la séance 2, du fait des musiciens. Ceci pourrait s'expliquer par la notion d'ambiguïté : le son VBR comme hybride mélangeant un son d'émission soufflée et d'émission frappée serait difficile à juger par rapport à l'échelle *soufflé*, le son TPT, par son caractère très synthétique pourrait engendrer des doutes quant à sa qualification de *très soufflé*. En effet, il semble impossible à partir de la seule écoute du son de considérer qu'il a pu être émis en soufflant. Ce n'est qu'en passant par son identification comme imitation de trompette ou de cuivre, qu'on peut lui attribuer cette qualité. Les sons les moins souvent ré-écoutés sont les sons PNO à la séance 1, VBR et OBC à la séance 2. Les deux premiers sons font parti des imitations de sons d'instruments percussifs, plutôt vraisemblables, et qui peuvent donc être jugés *peu soufflé* très rapidement. Le son OBC, en dépit du fait que c'est un hybride, ressemble beaucoup plus au son de clavecin (dont il a le profil temporel) qu'au son de hautbois, et a été classé très souvent comme un son d'instrument percussif.

Pour l'échelle liée à *diffus*, on retrouve le son TPT comme un des moins ré-écoutés lors de la séance 1 comme de la séance 2. On peut relier cela encore une fois à son caractère particulièrement synthétique, il présente donc très peu d'évolutions temporelles et paraît "droit". Par contre les sons CLA et VBR ne sont peu ré-écoutés que lors de la séance 1. On pourrait donner le même genre d'explication (très

synthétique donc peu d'évolution temporelle) pour le son CLA mais cette explication ne peut pas correspondre au son VBN qui est un des sons présentant le plus d'évolutions dans le temps (comme une double attaque au début). Le son OBC est le plus ré-écouté lors de la séance 1 mais un des moins ré-écoutés lors de la séance 2 (6 fois seulement à la séance 2 contre 24 fois à la séance 1 pour l'ensemble des 32 sujets). On est donc dans un cas inverse de la tendance générale de re-écouter plus souvent les sons lors de la séance 2. Peut être que dans le cas de la qualité "diffus" pour ce son OBC, les hésitations ont été levées lors de la séance 1.

Pour l'échelle liée à clair, que le son VBR, jugé parmi les plus clairs fait partie des plus souvent ré-écoutés, notamment par les non-musiciens, mais uniquement lors de la séance 1. Pour l'échelle liée à grave, le son CFT est plus ré-écouté (notamment par les non-musiciens) lors de la séance 1, alors que c'est le son TPT qui est le plus ré-écouté lors de la séance 2, cette fois-ci par les musiciens. Ceci nous paraît difficile à interpréter.

On trouve aussi 5 échelles pour lesquelles des différences dans la répartition des répétitions sur les sons sont significatives uniquement dans la séance 1. C'est le cas des échelles liées à *net*, *aigu*, *fort*, *résonnant*, et *sourd*. Pour l'échelle liée à *sourd*, cette différence est significative pour les 3 groupes de sujets considérés (tous, musiciens et non-musiciens), même si c'est beaucoup plus significatif pour les non-musiciens. Pour les 4 autres échelles de ce groupe, les différences sont toujours significatives pour l'ensemble de tous les sujets mais ceci est dû au groupe des musiciens pour les échelles liées à *aigu* et *sec*, alors que cela est dû au groupe des non-musiciens pour les échelles liées à *fort* et *résonnant*.

On trouve enfin 4 échelles pour lesquelles des différences dans la répartition des répétitions sur les sons sont significatives uniquement dans la séance 2. Il s'agit des échelles liées à *nasillard*, *vibré*, *pincé* et *métallique*.

### II.8.7. Résumé

Les données de nombre de re-écoutes de chaque son, en fonction de la séance, des sujets, des échelles et des sons considérés ne sont pas faciles à interpréter. Ceci semble être dû au fait que la re-écoute d'un son peut être motivée par diverses raisons contradictoires. Est-ce qu'on re-écoute un son parce qu'on a eu un moment d'inattention, parce qu'on trouve ce jugement difficile, parce qu'on hésite, qu'on ne sait pas, ou au contraire parce qu'on trouve ce jugement important, qu'on veut être précis ?

Il semble que contrairement à nos hypothèses de départ, les sujets re-écoutent plus souvent les sons quand ils trouvent le jugement important. C'est ainsi qu'on trouve beaucoup plus de re-écoutes lors de la séance 2 que de la séance 1, plus de re-écoutes chez les musiciens que chez les non-musiciens, et les plus fort taux de re-écoutes pour les échelles liées à des termes habituellement utilisés pour décrire les sons comme *brillant* ou *riche*.



### III. Expérience 3 : Reconnaissance de portraits verbaux

Cette expérience a pour principal objectif de tester la possibilité de reconnaître un son à partir de sa description en langue. Pour cela, nous avons construit plusieurs ensembles de types de descriptions différents de nos 12 sons à partir du matériel verbal récolté lors de notre première expérience. Nous avons appelé chacune de ces descriptions des portraits verbaux des 12 sons.

Les sujets devaient lire ces descriptions, écouter les 12 sons proposés et associer ceux qui leur semblaient décrits par le portrait donné.

Nous allons exposer ci-dessous la façon dont nous avons construit les portraits à tester, les sujets, stimuli, matériel et procédure utilisés, ainsi que quelques analyses préliminaires. L'essentiel des résultats récoltés ici seront analysés dans les chapitres 6§II, 7§II et surtout dans le chapitre 8.

#### III.1. Construction de portraits des sons

L'ensemble des portraits testés dans cette expérience est reproduit en annexe A.4.17. Nous avons vu chapitre 6 que l'essentiel des verbalisations récoltées quand on demande à des auditeurs de décrire les sons qu'ils entendent, se répartit en deux grandes catégories : des descriptions sur le son lui-même, en décrivant ses qualités par exemple, et des descriptions de la source sonore qui aurait pu produire ce son. Lors de la construction de portraits verbaux à reconnaître, il nous a paru nécessaire de tester des portraits appartenant à ces deux façons de parler des sons, afin de pouvoir éventuellement comparer notre facilité à les reconnaître.

Lorsque les sujets parlent des sources sonores, ils peuvent nommer un instrument de musique, une famille d'instruments de musique, ou faire une description en terme de matériau, mode d'excitation ou description plus complète d'un dispositif associant un mode d'excitation et un matériau pour construire une source sonore.

Nous avons extrait de l'ensemble des verbalisations libres obtenues dans l'expérience 1, tous les noms d'instruments produits pour décrire les sons (37 différents, par exemple : accordéon, clarinette, flûte...), et tous les noms de familles d'instruments produits (6 en tout, par exemple : instrument de la famille des bois, des claviers, des cordes...). Nous avons ajouté à cet ensemble, centré autour de la notion d'instrument de musique l'ensemble des 12 instruments visés par leur concepteur : soit les noms des instruments imités (par exemple trompette), soit les "parents" des hybrides (par exemple "mélange de vibraphone et de trombone" pour le son de *vibrone*).

Pour ce qui est des autres descriptions de sources sonores mais ne faisant pas appel directement à un nom d'instrument ou de famille d'instruments, nous avons rassemblé tous les noms de mode d'excitation produits spontanément (12 en tout, par exemple : on frappe, on froisse, on gratte...), ainsi que tous les noms de matériaux évoqués par ces sons (17 en tout, par exemple : de l'air, de

l'eau, de la peau et du métal...). Enfin, nous avons testé l'ensemble de toutes les descriptions combinant une action sur une matière (26 en tout, par exemple : du vent dans un tuyau, on casse du verre, on frappe du bois...). D'autre part, pour tester la possibilité de reconnaître un son décrit à travers ses qualités sonores, nous avons testé deux autres ensembles de portraits.

Pour tester dans un autre contexte le matériel verbal utilisé dans l'expérience 2 sous forme d'échelles sémantiques, nous avons repris les 23 adjectifs testés, auxquels nous avons enlevé haut et bas, qui étaient trop redondants par rapport à aigu et grave et nous les avons présentés, ainsi que leur négations sous forme de portrait verbal (2\*21 en tout, par exemple : un son aigu, un son clair, un son pas résonnant...). Puis afin de voir s'il était plus facile de reconnaître un son si on le décrivait à l'aide de plusieurs adjectifs différents, nous avons construit un ensemble de 12 portraits composés chacun de l'association de 3 adjectifs destinés à décrire chacun de nos 12 sons de façon spécifique. Pour choisir les trois adjectifs correspondant à chacun de nos 12 sons, nous avons utilisé les résultats des positions moyennes des sons sur les 23 échelles sémantiques dans l'expérience 2. Pour chaque son, nous avons choisi les 3 adjectifs parmi les 23 échelles qui étaient utilisées de la manière la plus extrême vis à vis de ce son. Par exemple, le portrait de 3 adjectifs sensé décrire le son de trompette (TPT) a été construit comme : "un son nasillard, peu pincé et peu doux", ce son étant positionné de façon extrême sur les échelles liées à nasillard, pincé et doux. Dans les cas où un même son était positionné de manière extrême sur plus de 3 échelles, le choix des 3 adjectifs s'est fait de manière à pouvoir distinguer les 12 portraits (qu'ils aient le minimum d'adjectifs en commun). Afin de rendre la compréhension de ces portraits plus facile, le nom des sons visés par chacun d'eux est donné en annexe A.4.17, sachant que les portraits proposés aux sujets ne comportaient pas ces noms.

Enfin, nous avons souhaité tester d'autres adjectifs décrivant les qualités des sons que ceux testés dans l'expérience 2. En effet, ces derniers étaient les plus fréquemment utilisés spontanément pour comparer les sons deux à deux. Mais si on considérait les adjectifs de qualité utilisés le plus fréquemment pour classier les sons, on pouvait ajouter 16 autres adjectifs (par exemple : un son agressif, un son chaleureux, un son cuivré...). Enfin, nous n'avons pas souhaité tester toutes les combinaisons possibles de 2 adjectifs, nous avons donc choisi toutes les paires d'adjectifs qui contenaient au moins un des adjectifs testés en solo et qui avaient été verbalisés ensemble spontanément (21 paires en tout, par exemple : un son aigu et métallique, un son court et ramassé, un son doux et froid...).

### III.2. Sujets

Quatre groupes de 20 sujets différents ont passé les 4 parties de cette expérience. Chaque groupe était composé pour moitié de musiciens (professionnels ou amateurs ayant plus de 10 ans de pratique musicale, codés pro, am1 et am2) et pour moitié de non-musiciens (vrais néophytes et amateurs ayant moins de 10 ans de pratique musicale, codés am3, am4 et néo). Nous avons essayé d'équilibrer les hommes et les femmes (voir annexes A.4.18 à A.4.21).

Les sujets qui ont passé la partie "instrument" étaient composés de 9 femmes et 11 hommes d'âge moyen  $M=33.9 \pm 10.4$  dont : 2 musiciens professionnels (pro), 3 musiciens amateurs ayant plus de 10 ans de pratique et un métier lié à la musique ou une pratique semi-professionnelle (am1), 5 musiciens amateurs ayant plus de 10 ans de pratique (am2), 6 musiciens amateurs ayant pratiqué autrefois ou ayant moins de 10 ans de pratique (am3) et 4 néophytes (néo). Ils ont été numérotés de mi1 à mi10 pour les musiciens et de ni1 à ni10 pour les non-musiciens.

Les sujets qui ont passé la partie "source, matériaux, excitation" étaient composés de 8 femmes et 12 hommes d'âge moyen  $M=28.9 \pm 5.3$  dont : 2 pro, 5 am1, 3 am2, 7 am3 et 3 néo. Ils ont été numérotés de ms1 à ms10 pour les musiciens et de ns1 à ns10 pour les non-musiciens.

Les sujets qui ont passé la partie qualité issue de l'expérience 2 ("qualités simple et triple"), étaient composés de 7 femmes et 13 hommes d'âge moyen  $M=29.6 \pm 9.0$  dont : 3 pro, 3 am1, 4 am2, 3 am3 et 7 néo. Ils ont été numérotés de mq1 à mq10 pour les musiciens et de nq1 à nq10 pour les non-musiciens.

Enfin, les sujets qui ont passé la partie "autres qualités simple et double" étaient composés de 9 femmes et 11 hommes d'âge moyen  $M=26.3 \pm 4.6$  dont : 2 pro, 4 am1, 4 am2, 3 am3 et 7 néo. Ils ont été numérotés de md1 à md10 pour les musiciens et de nd1 à nd10 pour les non-musiciens.

### III.3. Stimuli

Les 12 sons présentés ici sont les mêmes que ceux utilisés dans les deux expériences précédentes (voir annexe A.4.3.).

### III.4. Procédure

Les sujets commençaient par lire l'ensemble des descriptions proposées d'un des sous-groupes de portraits qui les concernait (dans un ordre aléatoire) puis ils entendaient l'ensemble des 12 sons. Ensuite, chacun des "portraits verbaux" du sous-groupe leur était présenté sur l'écran au dessus de 12 boutons numérotés. Ils pouvaient entendre chacun des sons en cliquant sur un des boutons (l'affectation des sons aux boutons était tirée au sort à chaque nouveau portrait). Les sujets devaient lire chaque portrait, écouter au moins une fois chacun des sons et sélectionner pour chaque portrait le ou les sons qui lui paraissaient correspondre au portrait proposé. Ils étaient obligés de choisir au moins un son pour chaque portrait. En fin de passation d'un sous-groupe, ils devaient dire pour chaque portrait proposé si, pour eux, celui-ci avait un sens précis, vague ou pas de sens du tout dans notre contexte. Puis ils passaient à un autre sous-groupe de portraits après une pause.

### III.5. Matériel

Les conditions d'expérience étaient identiques à celles de l'expérience 2. Les auditeurs étaient assis dans une cabine audiométrique en face de l'écran d'ordinateur. L'expérience était contrôlée par le programme PsiExp version 2.5 développé à l'Ircam par Bennett Smith (1995) qui contrôlait l'interface entre une console NeXT et les sons enregistrés sur le disque dur. Les stimuli étaient présentés diatiquement à travers un casque Sennheiser HD 520 II.

### III.6. Analyses préliminaires : utilisation de la consigne

Nous souhaitions tout d'abord analyser la manière dont les sujets ont utilisé la consigne proposée en analysant le nombre de sons qu'ils ont associé à chacun des portraits proposés. En effet, à partir du moment où les sujets étaient obligés d'associer au moins un son mais pouvaient en associer le nombre qu'ils jugeaient bon, nous voulions savoir s'ils avaient utilisé cette possibilité et de quelle manière. Est-ce que le nombre de sons associés à chaque portrait dépend du type de portrait proposé (décrivant le son vs décrivant la source), du sous-groupe de portrait, d'un des portraits en particulier ou plutôt de la compétence musicale ou de la personnalité des sujets ?

En moyenne, sur l'ensemble des 201 portraits proposés aux différents groupes de sujets, ceux-ci ont choisi d'associer 2.3 sons à chacun des portraits. On peut donc déjà dire que l'invitation à associer un ou plusieurs sons à chacun des portraits proposés a été suivie par ceux-ci.

Nous allons analyser ci-dessous les éventuelles variations de ce nombre de sons associés à chacun des portraits tout d'abord en fonction des portraits proposés, puis en fonction des auditeurs. (Les tableaux d'analyses de comparaison de moyenne par t de Student sont en annexe A.4.22).

#### **III.6.1. Nombre de sons associés en fonction du portrait proposé**

En effet, parmi les portraits proposés aux différents groupes de sujets, 6 sous-groupes décrivaient la source sonore des sons proposés (sous forme de nom d'instrument ou de famille d'instruments ou sous forme de mode d'excitation, de matériau ou de combinaison d'un mode d'excitation d'un matériau), et 4 sous-groupes décrivaient des qualités des sons proposés (sous forme d'adjectifs simples ou associés par 2 ou par 3). Nous nous attendons à d'éventuelles différences de comportement des sujets en fonctions de ces types de portraits.

##### Dépendance du type de portrait

En moyenne, les 110 portraits décrivant la source sonore ont été associés à 1.7 sons alors que les 91 portraits décrivant les qualités des sons ont été associés à 3.0 sons. Cette différence est significative ( $t(199)=9.72, p<.0001$ ). Les auditeurs ont donc tendance à choisir d'associer plus de sons différents aux portraits formés à partir d'adjectifs décrivant des qualités des sons qu'aux portraits décrivant une source sonore possible de ces sons.

##### Dépendance du sous-groupe de portrait

En moyenne, les 55 portraits de noms de mode d'excitation, de matériaux ou de combinaison des deux ont été associés à 2.0 sons alors que les 55 portraits de noms d'instruments ou de famille d'instruments ont été associés à 1.4 sons. Cette différence est significative ( $t(108)=4.20$ ,  $p<.0001$ ). Les auditeurs ont donc tendance à choisir d'associer un peu plus de sons différents aux portraits formés de noms de matière ou de mode d'excitation qu'aux portraits décrivant un instrument ou une famille d'instruments. Ceci peut s'expliquer par le caractère plus précis de la désignation de la source sonore par un nom d'instrument, par exemple "un son d'accordéon", par rapport à une désignation sous forme de mode d'excitation par exemple "on frappe".

En fait, parmi les portraits décrivant un mode d'excitation, un matériau ou une combinaison des deux, les 12 portraits de noms de mode d'excitation ont été associés à 2.2 sons, les 17 portraits de noms de matériaux à 2.1 sons et les 26 portraits de combinaison des deux à 1.8 sons. Bien qu'on note une diminution du nombre moyen de sons associés aux portraits quand on passe des composantes : excitation et matériau à la combinaison des deux, ces différences ne sont pas significatives ( $t<1.9$ ,  $p>.06$ , dl entre 27 et 41). Par contre les différences entre les sous-groupes de portraits décrivant les sons à l'aide de nom de famille d'instruments, d'instrument ou des instruments visés par leur concepteur, les différences observées sont significatives. En effet, les 6 portraits de noms de famille d'instruments ont été associés en moyenne à 2.9 sons, alors que les 37 noms d'instruments ont été associés à 1.3 sons et les 12 instruments visés à 1.1 sons (ces deux dernières moyennes sont tout juste significativement différentes entre elles :  $t(47)=2.00$ ,  $p=.052$ ). Les auditeurs ont donc tendance à associer plus de sons aux portraits de noms de famille d'instruments qu'aux noms d'instruments ( $t(41)=7.04$ ,  $p<.0001$ ) ou qu'aux noms de fichiers ( $t(16)=6.94$ ,  $p<.0001$ ). Ceci s'explique facilement, compte tenu que le nom de sons attendus pour chacun des portraits de noms de famille d'instruments était en général supérieur à 1. Par exemple, "un son de la famille des bois" pouvait être associé aux 2 sons CLA et CRA, "un son de la famille des vents" aux 4 sons TPT, TBN, CLA et CRA.

Pour les portraits décrivant les qualités des sons, en moyenne, les 54 portraits construits à partir de l'expérience 2 (21 adjectifs simples utilisés spontanément lors des comparaisons graduelles entre les sons, leurs 21 négations ainsi que les associations de 3 adjectifs visant chacun des 12 sons) ont été associés à 3.5 sons alors que les 37 autres portraits de qualités (adjectifs simples utilisés spontanément pour classer les sons et associations spontanées de 2 adjectifs) ont été associés à 2.3 sons. Cette différence est significative ( $t(89)=5.28$ ,  $p<.0001$ ). Les auditeurs ont donc tendance à choisir d'associer presque 2 fois plus de sons différents aux portraits formés à partir de l'expérience 2 qu'aux autres portraits de qualité sonore.

En fait, parmi les différents sous-groupes de portraits décrivant des qualités sonores des sons, les différences entre le nombre moyen de sons associés aux portraits associant 2 et 3 qualités, respectivement de 2.0 et 1.9 ne sont pas significatives ( $t(31)=0.97$ ,  $p=.34$ ).

Par contre, les 21 adjectifs simples (présentés de manière positive) issus de l'expérience 2 ont été associés en moyenne à 3.3 sons, alors que les 16 autres adjectifs simples (utilisés spontanément pour classer les sons plutôt que pour les comparer 2 à 2) n'ont été associés qu'à 2.8 sons. Cette différence est significative ( $t(36)=3.07$ ,  $p=.004$ ). Cela peut être dû à l'utilisation de ces adjectifs. En effet, pour pouvoir dire de 2 sons que "le premier est plus métallique que le deuxième" par exemple, nécessite que les deux soient tout de même "un peu métallique". On peut donc s'attendre à ce que plus de sons soient associés aux portraits construits à partir des adjectifs les plus souvent utilisés pour comparer deux sons (par exemple "un son métallique") que ceux construits à partir des adjectifs les plus souvent utilisés pour classer les sons.

Toutes les autres différences entre sous-groupes de portraits construits à base d'adjectifs de qualités sont très significatives ( $t$  allant de 4.40 à 13.51,  $p<.0001$ ,  $dl$  entre 26 et 41). En effet, le sous-groupe de portraits pour lequel le plus grand nombre de sons a été associé en moyenne à chaque portrait et celui construit à partir des adjectifs simples présentés de manière négative (par exemple "un son pas métallique"). Ce sous-groupe a en effet été associé en moyenne à 4.6 sons, ce qui est significativement différent de tous les autres nombres moyens. Ces adjectifs ayant le plus souvent été produits de manière positive, on associerait plus de sons à leur présentation négative sous forme de portrait qu'à leur présentation positive. On obtient donc le plus de sons associés aux portraits construits à partir d'un seul adjectif présenté de manière négative (4.6), puis aux adjectifs utilisés pour comparer les sons 2 à 2 (3.3) puis aux autres adjectifs simples (2.8) et enfin, aux adjectifs associés par groupes de 2 (2.0) et par groupes de 3 (1.9).

#### Dépendance du portrait

Pour ce qui est des portraits de noms de **mode d'excitation**, on peut considérer 3 groupes de portraits : ceux qui sont associés à plus de 2.7 sons en moyenne : "on frappe", "on percute", "on tape", "on pince" et "on souffle", ceux qui sont associés à moins de 1.5 sons en moyenne : "on racle", "on gratte", "on froisse" et des portraits plus intermédiaires : "on glisse", "on frotte", "on effleure" et "on tire". On peut remarquer que les portraits des deux derniers groupes rassemblent tous les modes de production utilisés pour décrire un des 12 sons, celui de CFT. Sur les 66 différences analysées, 15 sont significatives ( $t(19)>4.05$ ,  $p<0.05$  corrigé pour 66 analyses), et mettent toutes en jeu un des portraits du premier groupe décrit avec un de ceux du deuxième groupe.

Pour ce qui est des portraits de noms de **matériau**, on peut aussi considérer trois groupes : ceux associés à plus de 2.7 sons : "un matériau électronique", "du métal" un matériau naturel", "un matériau synthétique", "un matériau électrique" et "une corde métallique", ceux associés à moins de 1.5 sons : "de la peau", "du verre", "du cristal", "de l'eau", "de la peau et du métal". On peut remarquer que ces noms de matériaux ont été utilisés spontanément pour décrire soit le son VBR soit le son VBN. Enfin, ici aussi, on peut considérer dans un groupe intermédiaire : "du vent", "de l'air", "du cuivre", "une corde en boyau", "de la tôle" et "du bois". Sur les 136 différences analysées, 15

sont significatives ( $t(19) > 4.36$ ,  $p < 0.04$  corrigé pour 136 analyses), et mettent ici aussi toutes en jeu un des portraits du premier groupe décrit avec un de ceux du deuxième groupe.

Pour ce qui est des portraits de noms de **source**, formés comme association d'un mode d'excitation et d'un matériau, on ne trouve que 2 portraits associés à plus de 2.7 sons. Il s'agit de "un son produit par une touche d'orgue électronique" et de "on bouge une plaque métallique". Pourtant, sur les 47 différences significatives après correction selon Bonféroni pour 325 analyses, 14 contiennent le deuxième portrait mais aucune le premier. Le portrait "on bouge une plaque métallique" ayant été produit spontanément essentiellement pour décrire le son VBN, il est étonnant qu'il ait été si peu spécifique lors de la reconnaissance. Par contre, on s'attendait à trouver le portrait "un son produit par une touche d'orgue électronique" largement réparti sur l'ensemble des sons puisque tous peuvent répondre à cette description. A l'autre extrémité, on trouve 11 portraits associés à moins de 1.5 sons. On trouve parmi eux "une pièce de monnaie sur des cordes" ou "une double percussion sur une peau", produits spontanément pour décrire le son VBN, donc *a priori* assez spécifique, mais aussi "on frappe une corde" ou "on souffle dans du cuivre" pour lesquels on attendait *a priori* plus de sons. (L'ensemble de ces portraits ainsi que leurs nombres de sons moyens sont reproduits en annexe A.4.23).

Pour ce qui est des portraits de noms de **familles d'instruments**, on obtient de notables différences de nombre de sons associés à chacun des 6 portraits proposés. En effet, le portrait associé au plus grand nombre de sons différents est celui de "famille des cordes" (4.3 sons), puis "famille des vents", puis "famille des cuivres", "famille des claviers" et "famille des bois" et enfin "famille des percussions" (1.9 sons). Si on corrige les seuils de significativité selon Bonféroni pour 15 analyses, on obtient des différences significatives entre les portraits "famille des cordes" et "famille des percussions" ( $t(19) = 5.45$ ,  $p < 0.001$ ), de même entre les portraits "famille des percussions" et "famille des vents" ( $t(19) = 4.34$ ,  $p = 0.004$ ), entre les portraits "famille des bois" et "famille des cordes" ( $t(19) = 4.26$ ,  $p < 0.006$ ), entre les portraits "famille des bois" et "famille des vents" ( $t(19) = 3.87$ ,  $p < 0.01$ ) et entre les portraits "famille des claviers" et "famille des cordes" ( $t(19) = 3.72$ ,  $p < 0.02$ ).

Pour ce qui est des portraits de noms d'**instruments**, la grande majorité de ces portraits est associée en moyenne à moins de 1.5 sons (89% d'entre eux). En fait, seuls 4 de ces portraits sont associés à plus de 1.5 sons. Il s'agit des portraits "un son de synthétiseur" (3.5 sons), "un son d'orgue", "un son de trompe" et "un son de cithare" (1.6 sons). Les 4 portraits associés toujours à un seul son par tous les auditeurs concernent les instruments suivants : flûte, marimba, tambour et pipeau. Les différences entre ces deux groupes de sons sont donc notables ( $t$  environ 3.0), pourtant, si on corrige les seuils de significativité selon Bonféroni pour le nombre d'analyses (666 en tout mais même seulement par 36 pour chacun des portraits), aucune de ces différences n'est significative.

Pour ce qui est des portraits de **noms d'instruments visés par le concepteur**, les différences sont encore plus minimes, tous ces portraits ayant été associés au maximum à 1.2 sons. Nous ne

sommes donc pas surpris de ne trouver aucune de ces différences significatives ( $t(19) < 1.83$ ,  $p > 0.08$ ). Si on considère l'ensemble des sous-groupes de portraits construits à partir d'adjectifs décrivant des qualités des sons, on trouve aussi plusieurs comportements.

Pour ce qui est des portraits utilisant des adjectifs **simples** parmi ceux utilisés spontanément pour **comparer les sons 2 à 2**, 3 portraits se distinguent : "un son avec une attaque rapide", associé à 4.4 sons en moyenne d'une part, et d'autre part "un son fort" et "un son doux" associés respectivement à 2.3 et 2.5 sons, qui sont les portraits de ce sous-groupe associés au moins de sons différents. Il est intéressant de remarquer déjà ici que lors de l'expérience 2, *doux* était une échelle considérée comme pertinente pour beaucoup de sujets et de sons alors que *fort* était l'échelle considérée comme la moins pertinente et pour le moins de sons : les utilisations de ces adjectifs diffèrent donc grandement selon qu'ils sont utilisés pour positionner tous les sons ("peu doux - très doux") ou pour les caractériser ("un son doux"). Les deux différences significatives (après correction pour 231 analyses) concernent ces portraits ( $t(19) > 4.50$ ,  $p < 0.05$ ). Aux autres adjectifs sont associés 2.7 à 3.8 sons.

Pour ce qui est des portraits utilisant ces mêmes adjectifs **simples** mais de manière **négative** ("un son pas métallique"), on trouve parmi les portraits associés au plus grand nombre de sons différents "un son pas pincé" et "un son pas soufflé". On peut remarquer qu'ils ont tous les deux été construits à partir de modes de productions. A l'autre extrême, on trouve 3 portraits associés à moins de 4 sons : "un son pas fort", "un sons pas riche" et "un son pas net". Parmi les 231 analyses, seules 3 sont significatives après correction selon Bonféroni ( $t(19) > 4.57$ ,  $p < 0.05$ ) et elles concernent des portraits de ces deux groupes. Aux autres portraits de ce sous-groupe sont associés 4.0 à 5.3 sons.

Enfin, les autres portraits utilisant des mêmes adjectifs **simples** parmi ceux utilisés spontanément pour **catégoriser les sons**, on trouve 4 portraits associés à plus de 3.2 sons : "un son continu", et "un son constant" d'une part, qui peuvent être considérés comme synonymes et "un son terminé" et "un son étouffé" qui contiennent eux aussi la même idée. A l'autre extrême, on trouve 2 portraits associés à environ 2 sons : "un son pur" et "un son chaleureux". Parmi les 120 analyses, seules 3 sont significatives après correction selon Bonféroni ( $t(19) > 4.52$ ,  $p < 0.02$ ) et elles concernent des portraits de ces deux groupes. Les autres portraits de ce sous-groupe sont associés à entre 2.2 et 3.2 sons.

Pour ce qui est des portraits utilisant des associations de **deux adjectifs**, on trouve 5 portraits associés à plus de 2.5 sons et 2 associés à moins de 1.5 sons. Il s'agit de "un son ni mou, ni agressif", "un son net et franc", "un son pur et net", "un son rapide et précis" et "un son pincé et résonnant" et de "un son doux et froid" et "un son froid et agréable". Pourtant parmi les 210 différences testées une seule est significative avec correction selon Bonféroni, il s'agit de la différence entre les portraits : "un son pincé et résonnant" et "un son doux et froid" ( $t(19) = 4.72$ ,  $p = 0.02$ ).



Enfin, aux portraits utilisant des associations de **trois adjectifs**, sont associés de 1.5 sons pour "un son doux, sourd et soufflé" à 2.2 sons pour "un son assez long, assez doux et peu sec". Aucune des différences n'est significative après correction pour 66 analyses. On peut tout de même remarquer que le portrait associé au plus grand nombre de sons différents est le seul à avoir été construit à partir de deux adverbes "assez", il était censé décrire le son SNO, positionné de manière peu discriminante sur les 23 échelles sémantiques de l'expérience 2.

### III.6.2. Nombre de sons associés en fonction des auditeurs

Il peut être aussi intéressant de considérer si le nombre de sons associés à chacun des portraits dépend de la compétence musicale des auditeurs ou de leur personnalité.

#### Dépendance de la compétence musicale des auditeurs

Analyses faites sur les nombres de sons moyens associés à chacun des portraits par les deux groupes de sujets : musiciens et non-musiciens. Pour l'ensemble des 201 portraits testés, les différences entre les nombres moyens de sons associés aux portraits par les auditeurs de chacun des deux groupes (respectivement 2.3 pour les musiciens et 2.2 pour les non-musiciens) ne sont pas significatives ( $t(400)=0.88$ ,  $p=0.38$ ). Ces différences restent non significatives si on considère seulement les portraits décrivant la source sonore ou les portraits décrivant des qualités des sons (respectivement  $t(218)=0.52$  avec  $p=0.60$ , et  $t(180)=0.91$  avec  $p=0.36$ ).

Si on regarde les différences pour chacun des sous-groupes de portraits, les différences entre les musiciens et les non-musiciens ne sont pas significatives pour la plupart de ces sous-groupes ( $t < 1.75$ ,  $p > 0.08$ , dl variant de 10 à 72). Pourtant, on obtient une différence très significative pour le sous-groupe de portraits construits à partir d'adjectifs simples présentés de manière négative (par exemple "un son pas métallique") ( $t(38)=4.59$ ,  $p < .0001$ ). En effet, pour les portraits de ce sous-groupe, les musiciens associent en moyenne 5.3 sons alors que les non-musiciens y associent seulement 4.0 sons. Nous avons déjà noté la tendance des auditeurs à associer beaucoup de sons différents aux portraits de ce sous-groupe, cette tendance est en fait d'autant plus marquée que les sujets sont musiciens.

#### Dépendance de la personnalité des auditeurs

Les différents sous-groupes de portraits ayant mené à des différences significatives en terme de nombre de sons associés à chacun des portraits, nous ne comparerons les sujets qu'au sein de chacun des 4 groupes qui ont eu à associer les sons aux mêmes portraits.

Ainsi, parmi les 20 sujets qui ont associé les sons aux portraits de **sources sonores** (modes d'excitation, matériau et association des deux), le nombre de sons qu'ils ont associés en moyenne à chacun des portraits qui leur ont été proposés varie entre 1.0 pour le sujet ms8 et 2.8 pour le sujet ms3. En fait, sur les 190 analyses testées, seules 30 (soit 16%) sont significatives après correction selon Bonféroni ( $t(108) > 3.88$ ,  $p < 0.04$ ). Or, 18 de ces 30 différences (soit 60%) correspondent au

sujet ms8, qui est le seul à avoir associé quasiment systématiquement un seul son à chacun des portraits qui lui ont été proposés.

Parmi les 20 sujets qui ont associé les sons aux portraits de noms d'**instruments et de famille d'instruments**, le nombre de sons qu'ils ont associés en moyenne à chacun des portraits qui leur ont été proposés varie entre 1.0 pour le sujet ni5 et 1.9 pour le sujet mi3. Cette fourchette est plus resserrée que pour le groupe précédent. Seules 8 des 190 différences (soit 5%) sont significatives après corrections selon Bonféroni ( $t(108) > 3.80$ ,  $p < 0.04$ ). Or 7 de ces 8 différences (soit 87%) correspondent au sujet ni5. Il s'agit ici aussi d'un sujet qui a associé systématiquement un seul son à chacun des portraits qui lui ont été proposés.

Par contre, dans le groupe des 20 sujets qui ont associé les sons aux portraits construits à bases d'**adjectifs utilisés dans l'expérience 2**, les différences entre les sujets sont beaucoup plus marquées. En effet, le nombre de sons qu'ils ont associés en moyenne à chacun des portraits qui leur ont été proposés varie entre 1.1 pour le sujet nq2 et 5.0 pour le sujet mq10. Sur les 190 différences, 88 (soit 46%) sont significatives après corrections selon Bonféroni ( $t(106) > 3.96$ ,  $p < 0.02$ ). Les 19 différences que le sujet nq2 entretient avec chacun des autres sujets sont toutes significatives. C'est aussi le cas de 11 des différences qui concernent le sujet nq8 qui a associé en moyenne 2.2 sons à chacun des portraits qui lui ont été proposés. On trouve aussi un groupe de 3 sujets qui ont associé environ 2.5 sons à chacun des portrait, les sujets nq1, mq2 et mq5. A l'autre extrême, sur les trois sujets qui ont associés en moyenne plus de 4.5 sons à chacun des portraits qui leur ont été proposés, on trouve les sujets mq10, nq5 et nq4. Au moins 8 des différences qui les concernent aboutissent à des différences significatives.

Enfin, parmi les 20 sujets qui ont associé les sons aux portraits construits à bases d'**adjectifs simples et doubles**, le nombre de sons qu'ils ont associés en moyenne à chacun des portraits qui leur ont été proposés, varie entre 1.5 pour le sujet md9 et 3.2 pour le sujet nd6. Sur les 190 différences testées, 22 sont significatives après corrections selon Bonféroni ( $t(72) > 4.05$ ,  $p < 0.02$ ). En fait, on peut former 3 groupes de sujets : ceux qui associent plus de 2.9 sons aux portraits (les sujets nd6, nd3, nd8 et nd5, qui sont tous non-musiciens), ceux qui ont associé moins de 2 sons aux portraits (les sujets md9, nd4, md6, nd7 et md2) et les autres qui ont associé entre 2 et 2.7 sons aux portraits.

### III.6.3. Résumé

Les portraits décrivant des qualités des sons sont donc associés à plus de sons différents que ceux décrivant des sources sonores possibles. C'est d'autant plus le cas pour les adjectifs simples proposés de manière négative ("un son pas métallique"), puis pour les adjectifs simples ayant été utilisés spontanément pour comparer les sons 2 à 2. Pour ce qui est des portraits décrivant une source sonore possible, ceux qui sont associés au plus grand nombre de sons différents sont les portraits utilisant des noms de famille d'instruments et certains des noms de mode d'excitation.

Les seules différences de comportement liées à une différence de compétence musicale sont observées dans le cas des adjectifs simples proposés de manière négative : les musiciens les associent à plus de sons que les non-musiciens.

On remarque enfin des différences entre les sujets qui ne semblent dépendre que de leur personnalité, en particulier quand l'un d'entre eux n'associe qu'un seul son aux portraits proposés de manière systématique. Ces différences de comportement ne semblent pas liées à une compétence musicale. Il semble donc que la possibilité d'associer un ou plusieurs sons à chacun des portraits proposés ai été en général utilisée par les sujets et d'une manière qui a du sens.

## Chapitre 5 : Corrélats verbaux des dimensions perceptives

L'objectif général de ce chapitre est de dégager des corrélats verbaux des dimensions perceptives du timbre, mises en évidence par des analyses multidimensionnelles (MDS) des jugements de dissemblance perceptive entre paires de sons recueillis dans l'expérience 1 et analysés dans le chapitre 4, §I.5.

Depuis Plomp (1970), une technique utilisée pour modéliser notre représentation perceptive du timbre musical a été de construire des espaces multidimensionnels à partir de jugements de dissemblance ou de similitude donnés par les sujets sur des paires de sons. On peut établir une relation monotone entre des jugements de dissemblance et des distances dans un espace perceptif multidimensionnel. La plupart des études depuis 1977 (Grey, 1977 ; Krumhansl, 1989 ; McAdams et al., 1995) tendent vers un espace en trois dimensions partagé par tous les sons, avec pour certains sons des caractéristiques spécifiques. On a cherché des corrélats acoustiques de ces dimensions perceptives. Grey & Gordon (1978) montrent qu'une des dimensions est liée au "Centre de Gravité Spectral" (CGS) du son. Ce centre de gravité est calculé à partir des moyennes temporelles des amplitudes des harmoniques du son. Tous les sons de cette étude ayant la même fréquence fondamentale, 
$$CGS = \left( \sum_{k=1}^N k \cdot A_k \right) / \left( \sum_{k=1}^N A_k \right)$$
 avec  $A_k$  = Amplitude du  $k^{\text{ème}}$  partiel. Sa valeur est d'autant plus élevée que le spectre comporte plus d'harmoniques aigus importants. L'étude de Krimphoff et al. (1994) retrouve ce résultat et le complète par des corrélats acoustiques des deux autres dimensions perceptives de l'espace de Krumhansl (1989). La première dimension est corrélée au "Logarithme du Temps de Montée" (LTM) calculé à partir de l'enveloppe temporelle ( $LTM = \log_{10}(t_{\max} - t_{\text{seuil}})$  avec  $t_{\max}$  = l'instant où l'enveloppe atteint sa valeur maximale et  $t_{\text{seuil}}$  = l'instant où l'enveloppe passe le seuil perceptif, pris comme 2% du maximum par Krumhansl). La troisième dimension de l'espace de Krumhansl (1989) est corrélée à une mesure de la "Déviation" (DEV) du spectre par rapport à son enveloppe spectrale globale (structure fine), c'est à dire l'écart entre le spectre réel et son enveloppe fréquentielle ( $DEV = \sum_{k=1}^N \left| 20\log(A_k) - \left\{ 20\log(A_{k-1}) + 20\log(A_k) + 20\log(A_{k+1}) \right\} / 3 \right|$ ). La troisième dimension de l'espace de McAdams et al. (1995) est corrélée à une mesure du "Flux Spectral" (FL), c'est-à-dire le degré d'évolution de l'enveloppe spectrale au cours de la durée du son.

Dans le chapitre 4, §I.5., nous avons comparé l'espace perceptif que l'on peut calculer à partir des jugements de dissemblance recueillis dans l'expérience 1 à ceux dégagés lors des études précédentes qui ont utilisé les mêmes sons (Krumhansl, 1989 ; McAdams et al., 1995). Nous avons aussi testé les paramètres acoustiques calculés sur ces sons par Krimphoff (1993) et Krimphoff et al. (1994). Les deux premières dimensions, liées au temps d'attaque et à la richesse spectrale, ne sont pas modifiées par la tâche co occurrente de verbalisation. Nous allons essayer de compléter cette approche en recherchant, de la même manière, d'éventuels corrélats verbaux de ces dimensions perceptives, à partir des données de nos deux premières expériences.

Dans l'expérience 1, nous avons recueilli en même temps que les jugements de dissemblance perceptive, permettant de construire un modèle de la représentation perceptive du timbre sous forme d'espace multidimensionnel, des verbalisations libres sur toutes les différences et similarités que les sujets perçoivent entre les deux sons de la paire. Nous avons sélectionné une partie de ces verbalisations libres pour comparer leur production lors de la comparaison des deux sons de chaque paire avec la position relative de ces deux sons dans l'espace perceptif. Nous attendons que certaines unités verbales employées par les sujets correspondront directement à une dimension perceptive du timbre. On s'attend ainsi à trouver des corrélats verbaux des dimensions perceptives.

La deuxième expérience reprend une partie du vocabulaire dégagé dans la première expérience et le teste de manière plus quantitative et systématique au moyen d'une procédure de type VAME (*Verbal Attribute Magnitude Estimation*, Kendall & Carterette, 1992a). Cette fois-ci, c'est la position des sons sur les échelles sémantiques proposées qui a été comparée avec la position de ces sons sur les dimensions perceptives afin de déterminer si certaines échelles peuvent être considérées comme de bons corrélats verbaux de certaines dimensions perceptives.

## I. Expérience 1 : recherche de corrélats dans des verbalisations libres

Pour tester l'hypothèse de trouver dans les verbalisations libres des corrélats verbaux des dimensions perceptives, nous avons choisi d'extraire de l'ensemble des verbalisations libres recueillies une partie seulement des unités verbales. Par unité verbale, on entend la plus petite unité ayant un sens, cela peut donc aller d'un mot à une portion de phrase (Nosulenko & Samoylenko, 1997).

Nous avons choisi de considérer les unités verbales utilisées lors d'une comparaison graduelle entre les 2 timbres de chacune des paires écoutées, du type "*le timbre A est plus (ou moins) XXX que le timbre B*". En effet, cette syntaxe pourrait refléter une organisation de ces 2 timbres sur une dimension sémantique liée au sens de l'unité verbale considérée. En retenant les mots les plus souvent utilisés au centre d'unités verbales ayant cette construction syntaxique, nous pourrions comparer leur utilisation sur l'ensemble des paires avec chacune des dimensions perceptives obtenues lors de l'analyse multidimensionnelle. Si l'ordre des timbres obtenu par rapport à un des mots retenus est bien corrélé avec l'ordre relatif sur une des dimensions de l'espace perceptif, ce mot sera considéré comme un corrélat verbal de cette dimension.

### I.1. Choix du matériel analysé

En moyenne, les sujets ont verbalisé lors de cette expérience 4518 mots dont seulement 3.1% dans le contexte de comparaisons graduées entre les timbres. Globalement, les musiciens ont produit un peu moins de mots que les non-musiciens (respectivement 4518 contre 5704) mais avec un pourcentage de comparaisons graduées légèrement supérieur (3.7% contre 2.5%). Pourtant, une analyse de clusters effectuée sur ces données n'a pas révélé des groupes correspondant à l'expérience musicale des sujets.

Les verbalisations des sujets étant spontanées, aucun mot n'a été utilisé systématiquement par tous les sujets pour caractériser chaque paire de timbres. En contrepartie, lorsqu'un mot a été utilisé dans un cas particulier (un sujet et une paire donnée), nous le considérons comme saillant et pertinent dans ce contexte. Parmi les 596 mots différents utilisés au centre de comparaisons graduées, nous avons retenu pour les comparer aux dimensions perceptives les 28 attributs verbaux rassemblés dans le tableau 5.1. en utilisant les critères suivants :

A partir des unités verbales du type "*le timbre A est plus (ou moins) XXX que le timbre B*", nous avons calculé le nombre d'occurrences de chacun des mots positionnés au centre de ce type de comparaison. En considérant les mots ayant un nombre d'occurrences total, sur l'ensemble des sujets, supérieur à 32 dans ce contexte (soit supérieur ou égal à la moitié du nombre de paires écoutées : 66), nous obtenons les 16 attributs verbaux suivants : *aigu, attaque (rapide ou lente), grave, long, net, sourd, rond, clair, résonnant, nasillard, métallique, vibré, fort, sec, doux, et riche* (groupe C1). En gardant un nombre total d'occurrences supérieur à 32, nous avons regroupé des mots de sens voisin

reflétant des qualités remarquées assez systématiquement par les sujets. Ceci nous permet d'ajouter 5 ensembles de mots de sens proches dans notre contexte : le premier rassemble les notions de *percussif/frappé/impact rapide*, le deuxième celles de *chaud/chaleureux*, le troisième d'*homogène/constant/continu/plat/régulier/horizontal*, le quatrième celles d'*agréable/sympathique* et le dernier celles de *naturel/artificiel/synthétique* (groupe C2). Nous avons ajouté à cette liste des mots ayant un nombre d'occurrences plus faible mais utilisés habituellement dans la littérature comme des corrélats du timbre : *pincé*, *soufflé* et *brillant* (groupe C3). Enfin, nous voulions tester deux mots utilisés chacun de façon systématique par un seul sujet : *large* et *diffus*, ainsi que des synonymes courants de *aigu* et *grave*, respectivement, *haut* et *bas* (groupe C4).

**Tableau 5.1.** Attributs verbaux testés, C = Critère de sélection, NT : nombre total d'utilisations graduelles pour les 32 sujets, %M %N = utilisation par des musiciens et non-musiciens, %S = pourcentage de sujets ayant utilisé ce mot, %P = pourcentage de paires de sons pour lesquelles ce mot a été utilisé spontanément.

C	Attributs verbaux	NT	%M	%N	%S	%P
C1	<i>aigu</i>	119	40	60	95	86
	<i>attaque rapide</i>	101	63	37	55	71
	<i>grave</i>	98	54	46	65	76
	<i>long</i>	74	50	50	85	58
	<i>net</i>	73	67	33	35	76
	<i>sourd</i>	72	71	29	50	62
	<i>rond</i>	62	69	31	50	59
	<i>clair</i>	57	58	42	65	58
	<i>résonnant</i>	56	70	30	75	45
	<i>nasillard</i>	51	61	39	50	41
	<i>métallique</i>	49	84	16	45	50
	<i>vibré</i>	47	26	74	55	44
	<i>fort</i>	43	74	26	55	50
	<i>sec</i>	37	76	24	55	45
	<i>doux</i>	34	74	26	45	36
<i>riche</i>	32	34	66	40	36	
C2	<i>percussif/frappé/impact rapide</i>	46	74	26	50	50
	<i>chaud/chaleureux</i>	45	24	76	45	52
	<i>homogène/constant/continu/plat/régulier/horizontal</i>	41	54	46	55	45
	<i>agréable/sympathique</i>	36	17	83	40	38
	<i>naturel/artificiel/synthétique</i>	25	84	16	25	33
C3	<i>pincé</i>	21	67	33	40	27
	<i>brillant</i>	14	50	50	15	21
	<i>soufflé</i>	14	71	29	40	18
C4	<i>large</i>	24	96	4	20	33
	<i>diffus</i>	19	16	84	15	20
	<i>bas</i>	19	26	74	30	21
	<i>haut</i>	8	75	25	30	15

C1: NT > 32 après avoir rassemblé les formes au féminin, masculin, singulier et pluriel. C2: groupes de mots ayant des sens proches. C3: corrélats verbaux du timbre utilisés dans la littérature. C4: *Large* et *diffus* ont été utilisés systématiquement par un sujet et *bas* et *haut* sont des synonymes de *aigu* et *grave*.

Si on regarde la répartition des occurrences de chacun de ces mots en fonction de l'expertise des sujets, 64% des différences sont significatives ( $\chi^2(1) > 4.2$ ,  $p < .05$ ) et concernent pour 64% d'entre

elles des mots qui ont été plus souvent produits spontanément par des musiciens. Ceci pourrait refléter une plus grande homogénéité de vocabulaire parmi les musiciens (le principal critère de sélection de ces mots étant le nombre d'occurrences). Les 7 différences les plus significatives ( $\chi^2(1) > 11$ ,  $p < .001$ ) concernent les mots suivants : *métallique*, *naturel/artificiel/synthétique* et *sourd*, qui sont respectivement produits à 84, 84 et 71% par des musiciens, *agréable/sympathique*, *chaud/chaoureux* et *vibré*, qui sont respectivement produits à 84, 76 et 74% par des non-musiciens (*large* étant essentiellement dû à un sujet musicien, cf. groupe C4). On peut remarquer que les mots utilisés principalement par des sujets non-musiciens se rapportent à des qualités hédoniques des sons.

Il est d'autre part intéressant de remarquer que les mots les plus utilisés spontanément par les sujets (*aigu*, *grave*, *long* et *fort*) se rapportent principalement à des paramètres acoustiques qui ont été égalisés sur l'ensemble des timbres (hauteur, durée et sonie) et ne devraient donc pas être pertinents dans notre contexte. Pourtant, sur ces 4 adjectifs, seul *aigu* a été principalement utilisé par des non-musiciens (60% des utilisations) sachant que 95% des sujets l'ont utilisé au moins une fois. *Fort* a même été principalement utilisé par des sujets musiciens (74%), peut-être dans le sens de "puissant", "présent" ou de "force".

## I.2. Analyses

Nous allons présenter ici trois types de comparaisons entre l'utilisation des attributs verbaux sélectionnés ci-dessus pour comparer les sons des paires écoutées et les positions des sons de ces paires sur les dimensions perceptives, telles que l'analyse multidimensionnelle nous les donne. Tout d'abord, nous avons calculé des corrélations de rang de Kendall pour chacun des attributs verbaux et chacune des 4 dimensions perceptives. Puis nous avons effectué des régressions multiples pour chacun des attributs verbaux avec l'ensemble des 4 dimensions perceptives. Enfin, nous avons imaginé une représentation graphique qui permette de visualiser la manière dont chaque attribut verbal est utilisé pour comparer les sons positionnés dans l'espace perceptif.

### **I.2.1. Corrélations de rang de Kendall**

Nous avons calculé des corrélations de rang de Kendall entre la proportion  $\bar{Y}$  de sujets ayant jugé que "*le premier timbre d'une paire est plus XXX que le deuxième*" et le vecteur-coordonnées de la paire de timbres sur la dimension de notre espace perceptif en 4 dimensions sans spécificité. Les 4 meilleurs attributs verbaux de chacune des dimensions perceptives sont représentés en gras dans le tableau 5.2.

Parmi les 28 attributs verbaux testés, 11 sont fortement corrélés à la dimension 1 de notre espace perceptif : ***net***, ***attaque rapide***, ***clair***, ***métallique***, puis *pas sourd*, *percussif*, *pas homogène*, *pas long*, *pas large*, *pas grave* et *résonnant* ( $p < .01$ ). On trouve aussi corrélé à cette première dimension *sec*, *pas soufflé*, *aigu* et *naturel* ( $p < .05$ ). A part *pincé* (très faiblement corrélé), corrélat habituellement donné à cette première dimension dans la littérature, on retrouve tous les attributs verbaux ayant un caractère temporel, qu'ils décrivent le début du son (*attaque rapide*, *percussif*, *pas soufflé*, *net*, *sec*)



ou de manière plus générale le déroulement temporel du son (*homogène, pas long, pas large, résonnant*). Ceci est en accord avec la corrélation de cette dimension perceptive avec le logarithme du temps de montée de l'enveloppe temporelle (LTM), qui est en général fortement corrélé avec la durée perçue du son (Krimphoff, 1993). La présence de *clair, métallique, sourd, grave et aigu* est plus étonnante, car ils représentent plutôt des concepts qu'on aurait associés à l'idée de hauteur spectrale donc à la dimension 2, à laquelle d'ailleurs ces attributs verbaux sont aussi corrélés, à part *clair*. Le fait que l'utilisation de *naturel* soit corrélée à cette première dimension vient peut être du mode de synthèse de ces sons : les imitations d'instruments pincés ou frappés paraissent plus naturelles que celles des instruments soufflés, présentant peu de variations temporelles après le début du son (corrélation négative avec *homogène*). Ce type d'explication peut nous permettre de comprendre aussi la présence de *clair*: les sons plus percussifs étant mieux imités donnent une impression de son de corde frappée ou pincée plus naturelle et donc sans doute plus *claire* que les sons qui imitent ceux produits par des instruments à vent, d'aspect beaucoup plus synthétiques.

**Tableau 5.2.** Corrélations de rang (df=10) entre les proportions de verbalisations "plus XXX" pour chaque paire et les positions relatives des sons le long de chacune des 4 dimensions de l'espace perceptif.

(les valeurs sont les z de Kendall, NB = nombre de dimensions corrélées significativement).

Attributs Verbaux	Dim 1	Dim 2	Dim 3	Dim 4	NB
<i>aigu</i>	3.20 †	<b>4.99 *</b>	1.72	2.43	2
<i>attaque rapide</i>	<b>6.74 *</b>	-0.09	1.98	1.32	1
<i>grave</i>	-3.59 *	<b>-5.00 *</b>	-0.46	-3.14 †	3
<i>long</i>	-3.72 *	0.08	<b>-3.50 †</b>	-3.32 †	3
<i>net</i>	<b>6.92 *</b>	-1.79	0.67	0.27	1
<i>sourd</i>	-5.13 *	-4.16 *	0.54	-1.69	2
<i>rond</i>	-1.68	<b>-5.00 *</b>	0.56	<b>-3.60 *</b>	2
<i>clair</i>	<b>5.35 *</b>	2.06	-0.88	-0.20	1
<i>résonnant</i>	3.59 *	-1.42	-1.66	-2.24	1
<i>nasillard</i>	-2.97	4.10 *	-0.48	3.19 †	2
<i>métallique</i>	<b>5.30 *</b>	3.52 †	-0.66	1.16	2
<i>vibré</i>	-0.68	0.18	-2.27	1.45	0
<i>fort</i>	0.50	-0.28	-1.73	0.95	0
<i>sec</i>	3.39 †	1.34	1.34	2.32	1
<i>doux</i>	-1.05	-2.44	1.59	<b>-4.00 *</b>	1
<i>riche</i>	-0.71	0.78	<b>-3.10 †</b>	1.18	1
<i>percussif/frappé/impact rapide</i>	4.90 *	-2.02	-0.78	1.94	1
<i>chaud/chaleureux</i>	-2.42	<b>-4.70 *</b>	-1.52	-2.66	1
<i>homogène/constant/continu/plat/régulier</i>	-4.27 *	2.17	-1.00	<b>-3.30 †</b>	2
<i>agréable/sympathique</i>	1.94	-4.21 *	-0.15	<b>-3.40 †</b>	2
<i>naturel/artificiel/synthétique</i>	3.18 †	-1.54	-0.19	-2.88	1
<i>pincé</i>	1.26	2.51	0.42	1.57	0
<i>brillant</i>	2.86	1.96	-1.96	1.21	0
<i>soufflé</i>	-3.30 †	-0.62	0.62	-0.09	1
<i>large</i>	-3.65 *	-1.38	-0.12	-0.28	1
<i>diffus</i>	-0.83	-1.66	-0.83	0.83	0
<i>bas</i>	-1.32	-2.57	-0.08	0.70	0
<i>haut</i>	2.45	2.69	0.98	0.73	0

\* : p<.01, † p<.05 (Fisher's r-to-z avec correction de Bonférioni pour 28 analyses). Les 4 meilleurs corrélats de chacune des dimensions sont en gras.

Pour la deuxième dimension perceptive, corrélée au centre de gravité spectral (CGS), on trouve 7 attributs verbaux fortement corrélés : *pas grave*, *aigu*, *pas rond* et *pas chaud* puis *pas agréable*, *pas sourd* et *nasillard* ( $p < .01$ ). On trouve aussi corrélé à cette deuxième dimension le mot *métallique* ( $p < .05$ ). On retrouve donc ici *aigu* et *grave*, habituellement utilisés pour décrire la hauteur tonale d'un son. Apparemment, ces notions s'appliquent également aux aspects fréquentiels du timbre : un son *aigu* pouvant donc être, suivant le contexte, un son de fréquence fondamentale élevée ou un son ayant un centre de gravité spectral élevé. Cette confusion entre hauteur tonale et hauteur spectrale pourrait expliquer la difficulté que rencontrent beaucoup d'étudiants de conservatoire lors de dictées à plusieurs voix jouées par des instruments différents. Il est aussi intéressant de remarquer que les variations de centre de gravité spectral d'un son sont liées aux notions de *rond*, *d'agréable* et de *chaleureux* : plus un son a un centre de gravité spectral bas, plus il est perçu comme rond, agréable et chaleureux. Un son rond ou sourd serait plus agréable qu'un son nasillard ou métallique.

On ne trouve que deux attributs verbaux corrélés à la troisième dimension perceptive, qui est corrélée à la déviation spectrale (DEV) entre la structure fine du spectre et son enveloppe spectrale générale : *pas long* et *pas riche* ( $p < .05$ ). La notion de richesse pourrait être liée à la perception d'une structure fine très accidentée par rapport à son enveloppe spectrale. *Long* est aussi significativement corrélé à la dimension 1 "temporelle". Le son perçu comme le plus *long* de l'ensemble testé (bien que les sons aient été égalisés en longueur perçue) est l'imitation de corde frottée dont le timbre se trouve à une des extrémités de la dimension 3. Sa corrélation avec cette dimension peut être due au fait que les sons de cordes frottées présentent beaucoup de variations de leur enveloppe spectrale et donc une très forte structure fine.

Notre quatrième dimension perceptive, bien qu'ayant aucun corrélat acoustique dans l'ensemble de paramètre de Krimphoff (1993), présente 7 corrélats verbaux dont deux fortement corrélés : *pas doux* et *pas rond* ( $p < .01$ ). Viennent ensuite *pas agréable*, *pas homogène*, *pas long*, *nasillard* et *pas grave* ( $p < .05$ ). À part *doux*, corrélé uniquement avec cette quatrième dimension, les 6 autres attributs verbaux sont tous corrélés avec au moins une autre dimension, la dimension 2 pour 4 d'entre eux : *rond*, *agréable*, *grave* et *nasillard*, et la dimension 1 pour *long* et *homogène*.

On peut remarquer que parmi les sons de coordonnées négatives sur cette dimension, qui correspondent aux imitations de sons de clarinette, corde frottée, piano et vibraphone, le son d'imitation de clarinette est souvent décrit comme "pas d'évolution, plat, homogène", et celui d'imitation de vibraphone comme "rond et doux" (Faure et al., en préparation).

Enfin, parmi ces 28 attributs verbaux testés, 7 ne sont significativement corrélés à aucune de nos 4 dimensions perceptives : *fort* (sons égalisés en sonie), *diffus*, *haut* et *bas* (ceci peut être dû au faible nombre d'occurrences de ces mots dans le contexte testé) mais aussi *vibré*, *pincé* et *brillant* qui font pourtant partie du vocabulaire habituellement utilisé par les musiciens pour parler du timbre.

### I.2.2. Régressions multiples

On peut remarquer que, sur les 28 attributs verbaux testés, seuls 12 d'entre eux sont significativement corrélés à une seule dimension perceptive. En effet, on trouve 9 attributs verbaux significativement corrélés à plusieurs dimensions (2 ou 3), ce qui suggère qu'ils soient l'expression d'une variation "complexe" à travers plusieurs dimensions perceptives. Afin d'estimer les contributions relatives des différentes dimensions perceptives à cette expression, nous avons effectué une série de régressions multiples incrémentales. Les résultats sont résumés dans le tableau 5.3.

Tableau 5.3. Coefficients de régression incrémentale entre l'utilisation de chaque attribut verbal et les 4 dimensions perceptives (F pour entrer = 4). ( $\bar{Y} = c_0 \cdot \mathbf{1} + \sum_{i=1}^4 c_i \cdot \overline{\text{Dim}_i}$ ) et variance expliquée (R<sup>2</sup>).

Attributs Verbaux	R <sup>2</sup>	C <sub>0</sub>	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>
<i>nasillard</i>	0.75 *	0.55	-0.48	0.71	-0.28	
<i>long</i>	0.57 *	0.62	-0.40		-0.46	-0.41
<i>rond</i>	0.55 *	0.52	-0.26	-0.56		-0.34
<i>grave</i>	0.55 *	0.48	-0.39	-0.49		-0.26
<i>aigu</i>	0.50 *	0.56	0.34	0.57		0.22
<i>sourd</i>	0.69 *	0.53	-0.63	-0.51		
<i>agréable/sympathique</i>	0.69 *	0.36	0.38	-0.77		
<i>métallique</i>	0.62 *	0.58	0.65	0.28		
<i>large</i>	0.59 *	0.54	-0.73	-0.30		
<i>attaque rapide</i>	0.69 *	0.58	0.80		0.19	
<i>brillant</i>	0.54	0.55	0.58		-0.48	
<i>percussif/frappé...</i>	0.67 *	0.49	0.77			0.29
<i>homogène/constant...</i>	0.63 *	0.67	-0.62			-0.44
<i>natural/artificiel...</i>	0.41	0.61	0.48			-0.37
<i>chaud/chaleureux</i>	0.59 *	0.55		-0.69		-0.27
<i>soufflé</i>	0.76 *	0.53	-0.88			
<i>net</i>	0.60 *	0.61	0.78			
<i>clair</i>	0.55 *	0.61	0.75			
<i>résonant</i>	0.31 †	0.47	0.58			
<i>sec</i>	0.24	0.44	0.51			
<i>haut</i>	0.54	0.53		0.77		
<i>bas</i>	0.31	0.52		-0.61		
<i>pincé</i>	0.16	0.55		0.45		
<i>riche</i>	0.30	0.51			-0.57	
<i>vibré</i>	0.13	0.31			-0.41	
<i>doux</i>	0.49 *	0.37				-0.72

\* : p<.01, † p<.05 (avec corrections de Bonféroni pour 28 analyses)

On obtient 5 attributs verbaux tridimensionnels : *nasillard* est corrélé aux trois premières dimensions, *long* aux dimensions 1, 3 et 4, *rond*, *grave* et *aigu* aux dimensions 1, 2 et 4. Ainsi, un son est perçu comme plus *nasillard* si son temps d'attaque est plus lent, son centre de gravité spectral plus haut et sa déviation spectrale moins importante. *Aigu*, au contraire de *grave*, correspond à un son avec à la fois une attaque rapide et un centre de gravité spectral élevé. D'autre part, 10 autres attributs verbaux sont bidimensionnels : *sourd*, *agréable*, *métallique* et *large* sont corrélés à la fois aux dimensions 1 et 2 ; *attaque rapide* et *brillant* aux dimensions 1 et 3 ; *percussif*, *homogène* et *naturel* aux dimensions 1 et 4,

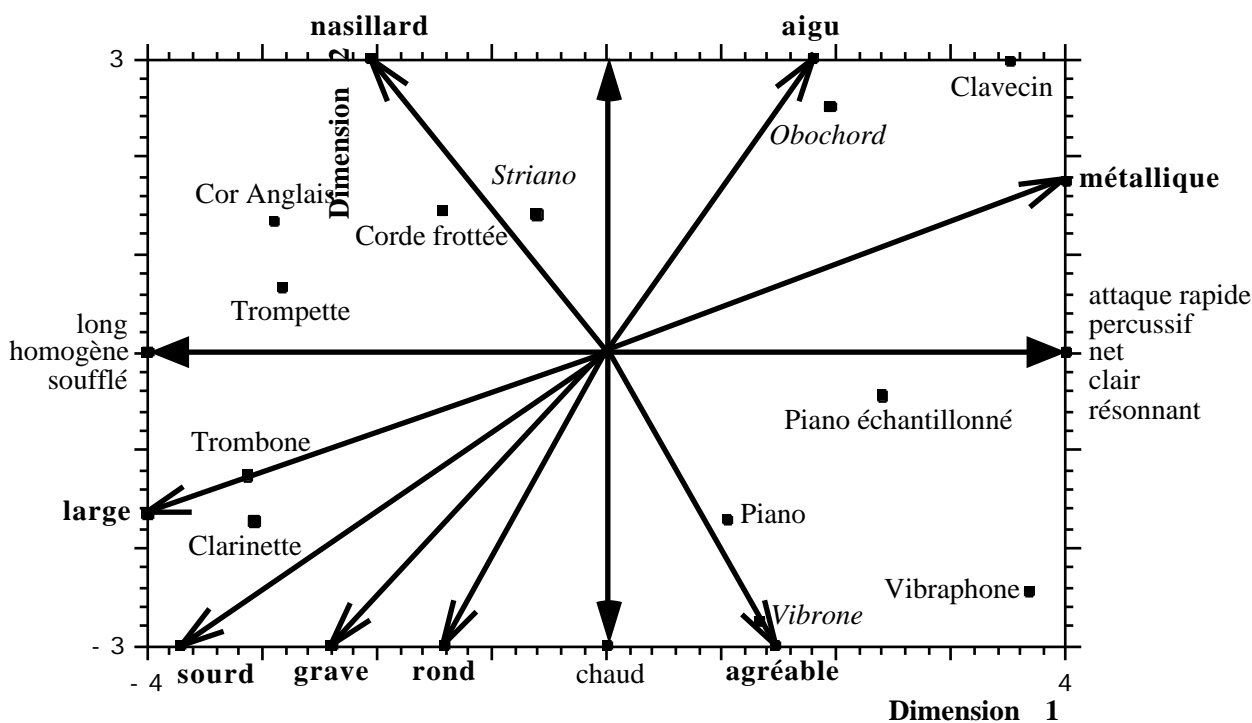
*chaud* aux dimensions 2 et 4. Ainsi, par exemple, un son est perçu comme plus *agréable* s'il a une attaque assez rapide et un centre de gravité spectral bas.

Les autres attributs verbaux ne présentent pas de caractère composite des dimensions perceptives. Seuls 5 attributs verbaux présentent des corrélations significatives avec une seule des dimensions perceptives : *soufflé*, *net*, *clair* et *résonnant* pour la dimension 1, *doux* pour la dimension 4. Il est intéressant de noter que *brillant*, couramment utilisé comme corrélat verbal d'une dimension liée au centre de gravité spectral dans la littérature apparaît ici corrélé aux dimensions 1 et 3, mais pas à la dimension 2 liée au centre de gravité spectral. Cette dimension 2 semble mieux expliquée par une sensation de "hauteur spectrale" (liée à la tonie, aspect *aigu/grave* de la hauteur, mais fortement corrélé à son contenu spectral). *Fort* et *diffus* n'apparaissent pas ici, car ils ne sont significativement corrélés à aucune des 4 dimensions perceptives.

### I.2.3. Représentation graphique

Nous avons essayé de représenter le sens d'un attribut verbal composite de plusieurs dimensions perceptives en représentant la direction suivant laquelle varie son utilisation dans chacun des plans perceptifs (Figures 5.1 à 5.6). Sur ces figures, on représente par une flèche la projection, sur le plan des deux dimensions concernées, de la direction de variation des attributs verbaux significativement corrélés à ces deux dimensions. La pente de la droite étant égale au rapport des coefficients de régression multiple avec ces deux dimensions.

Figure 5.1. Plan engendré par les dimensions 1 et 2 et ses attributs verbaux (timbres hybrides en italique, attributs composites des deux dimensions du plan en gras).



On peut par exemple remarquer sur la Figure 5.1 que, bien que *nasillard*, *métallique* et *aigu* soient tous les trois corrélés de façon positive et significative à la fois aux dimensions 1 et 2, ils reflètent des

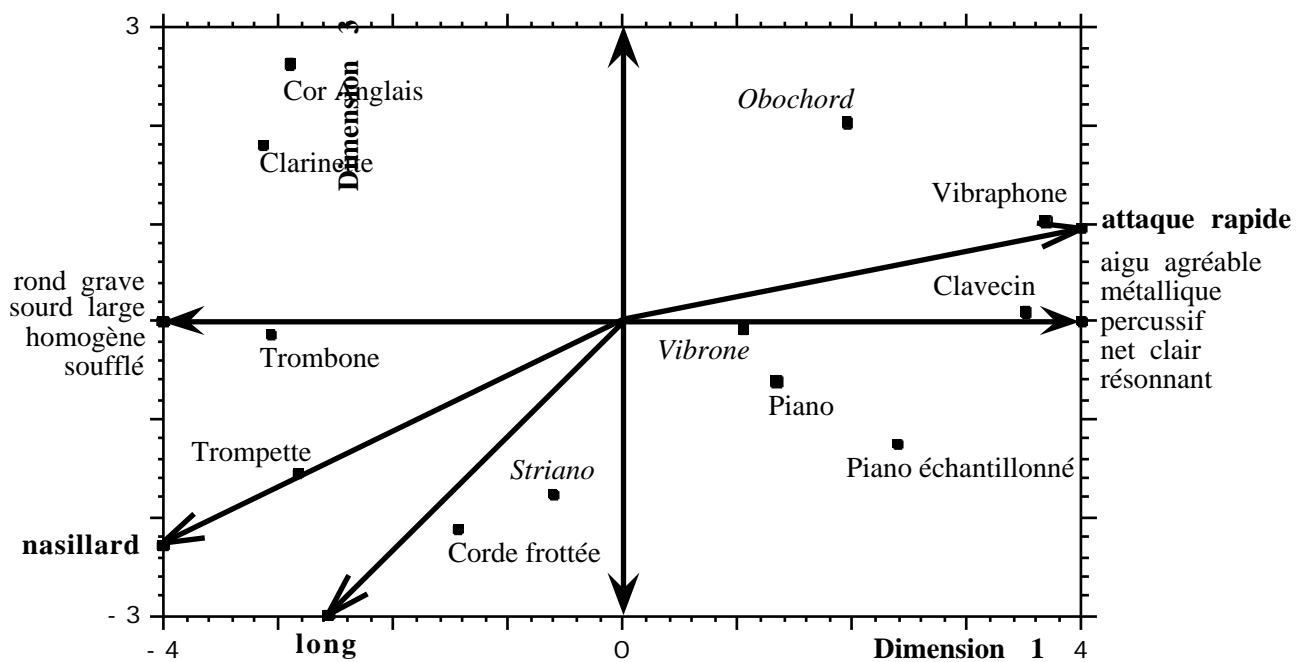
perceptions différentes : *métallique* semble correspondre à une variation entre un son de clarinette ou de trombone et un son de clavecin alors que *nasillard* correspondrait plutôt à une variation entre un son de vibraphone et un son de cor anglais, ces deux directions étant orthogonales. Cette observation va à l'encontre de l'habitude de considérer ces deux mots (auxquels on ajoute souvent *brillant*) comme des corrélats interchangeables d'une dimension perceptive liée au centre de gravité spectral des sons (Kendall & Carterette, 1996). En effet, l'idée de *métallique* semble plus fortement influencée par un contenu temporel du timbre et donc par le mode de production du son : un son frappé ou pincé paraîtrait plus facilement *métallique*.

D'autre part, le sens de mots comme *aigu* et *grave* (utilisés comme des antonymes vis-à-vis de notre perception du timbre) dénote à la fois une variation temporelle et spectrale : un son est perçu comme *plus aigu* si son centre de gravité spectral est plus haut et son attaque est plus courte. Ceci peut aussi révéler des covariations obligatoires de ces attributs acoustiques dans la facture des instruments (en effet, on ne trouve pas d'instrument ayant un CGS aussi élevé que le clavecin et une attaque très lente, du moins pour la hauteur (Mib4) utilisée dans cette étude).

Au contraire de *aigu*, *nasillard* et *métallique* dont les directions de variation d'utilisation divergent, celles correspondant à *sourd*, *grave*, *chaud* et *rond* sont plus proches si on ne considère que leur variation par rapports aux dimensions 1 et 2.

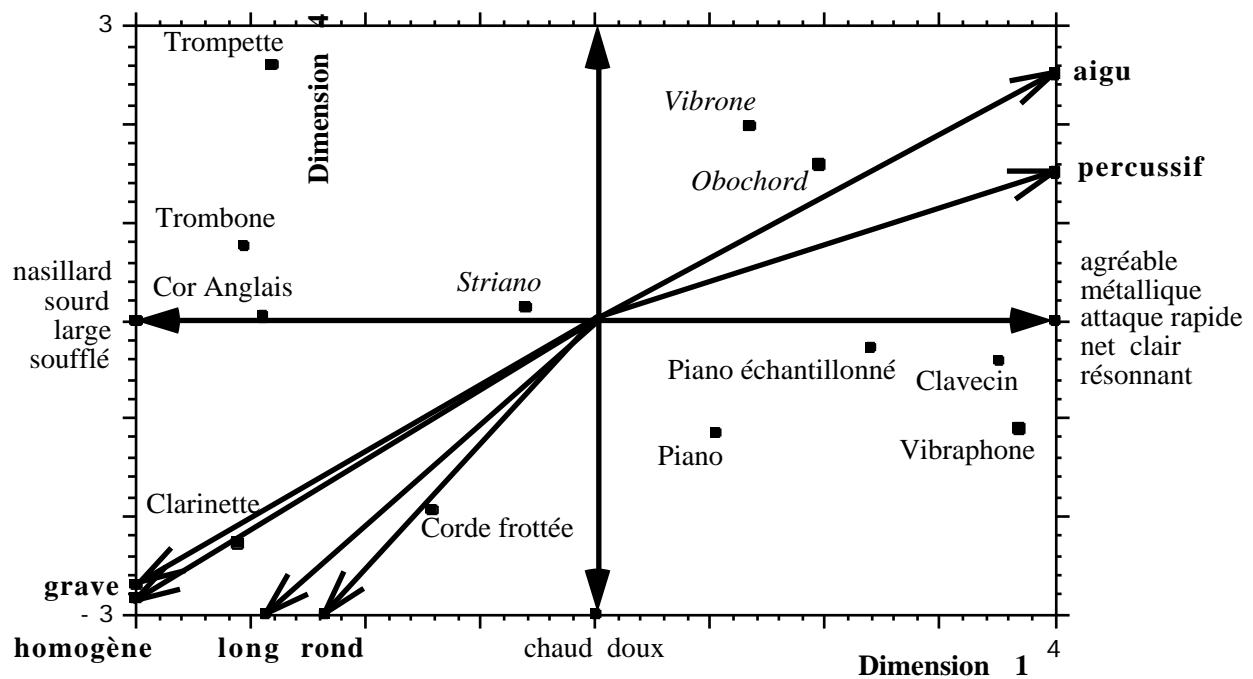
Enfin, *agréable* apparaît ici comme un bon antonyme de *nasillard*. Ils varient en effet de manière opposée. On retrouve sur le graphe l'idée qu'un son est d'autant plus *agréable* que son attaque est rapide (et qu'on peut entendre une résonance libre, puisque les deux co-varient) et que son Centre de Gravité Spectrale est bas (c'est à dire qu'il est perçu plus *grave* qu'*aigu*).

Figure 5.2. Plan engendré par les dimensions 1 et 3 et ses attributs verbaux (voir légende fig. 5.1).



De même, la figure 5.2. montre que la notion d'*attaque rapide/lente*, si elle est essentiellement liée à la perception du temps d'attaque (LTM, dimension 1), ce à quoi on s'attendait, est aussi légèrement augmentée par la perception d'un flux spectral faible. C'est le contraire pour *long* : un son serait perçu d'autant plus long que son attaque serait lente et que son flux spectral serait important, ce qui est le cas par exemple de la corde frottée. La corrélation de l'utilisation de *long* avec le LTM est à rapprocher de celle trouvée par McAdams et al. (1995) entre LTM et la "durée effective" du son, mesurée à -3dB de l'amplitude maximale dans l'enveloppe temporelle de ces sons. Par contre, le fait que *nasillard* soit lié à un flux spectral élevé est plus étonnant : le son de cor anglais, pourtant souvent considéré comme une sorte de prototype de son *nasillard*, apparaît ici comme moins *nasillard* que celui de trompette.

Figure 5.3. Plan engendré par les dimensions 1 et 4 et ses attributs verbaux (voir légende fig.5.1)

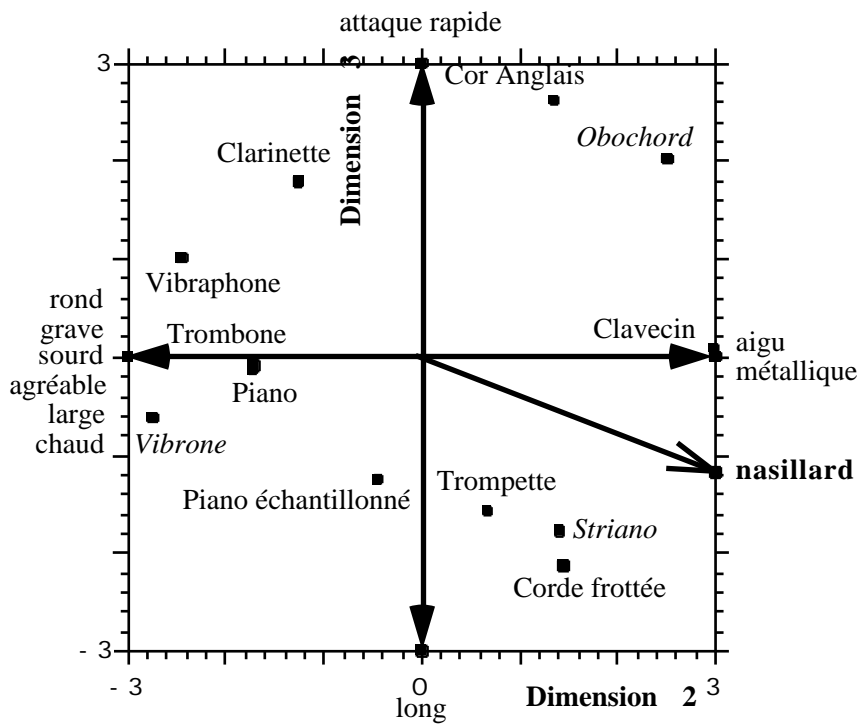


Sur la figure 5.3., on observe une variation proche des notions d'*homogène*, *long*, *rond* et *grave*, qui varient de la région où se trouve le *vibrone* et l'*obochord*, deux sons hybrides qui présentent tous deux un caractère double dans l'attaque, à la région où se trouve la clarinette, dont l'imitation est perçue comme particulièrement "plate" et "sans évolution". Cette figure nous montre aussi que la notion de *percussif* n'est pas liée qu'à la vitesse d'attaque, elle se colore d'autre chose liée à cette dimension 4 dont on n'a pas trouvé de corrélat acoustique dans l'ensemble développé par Krimphoff et al. (1994). Ceci nous permet de suggérer la recherche d'un corrélat acoustique qui refléterait des variations temporelles dans la structure du son.

Comme dans la figure 5.1., on retrouve dans ce plan *aigu* comme antonyme de *grave*. Ces deux mots étant composites des trois dimensions 1, 2 et 4, on peut les considérer comme de vrais antonymes dans la manière dont on les utilise pour comparer le timbre des sons.

La figure 5.4. présente essentiellement le fait que la notion de son *agréable* semble dépendre, outre du temps d'attaque (cf. fig. 5.1.), à la fois d'un centre de gravité spectral assez bas et d'un flux spectral assez élevé.

Figure 5.4. Plan engendré par les dimensions 2 et 3 et ses attributs verbaux (voir légende fig.5.1)



La figure 5.4 reprend une dernière fois l'opposition entre *aigu* et *grave*. S'y ajoute le fait que les notions de *doux* et de *chaud/chaoureux* se rapprochent ici de celle de *grave* : apparaissant à la fois liée à la dimension 4 (absence de variations dans l'enveloppe temporelle ?) et à un centre de gravité spectral bas.

Figure 5.5. Plan engendré par les dimensions 2 et 4 et ses attributs verbaux (voir légende fig.5.5)

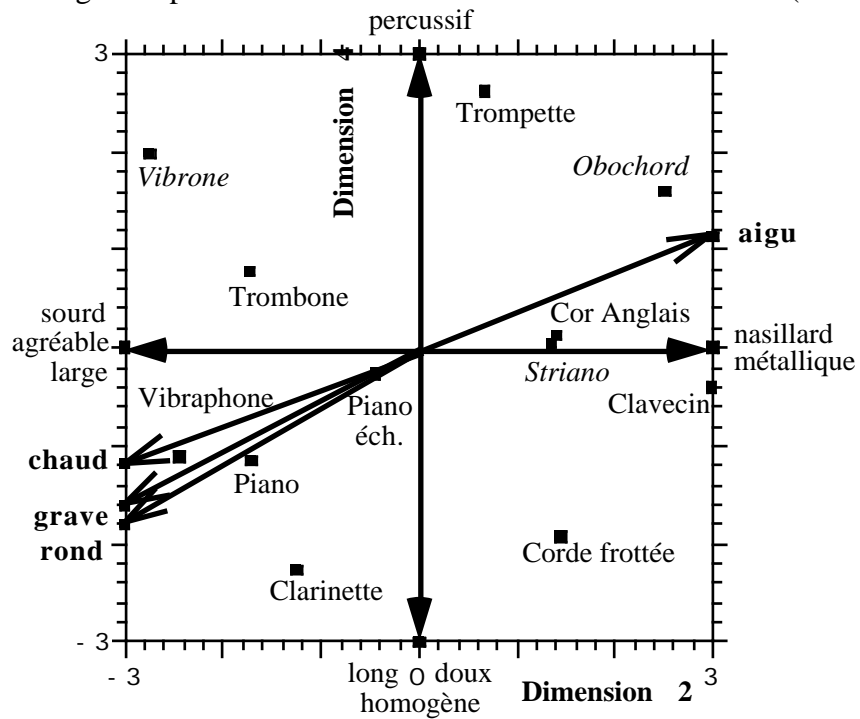
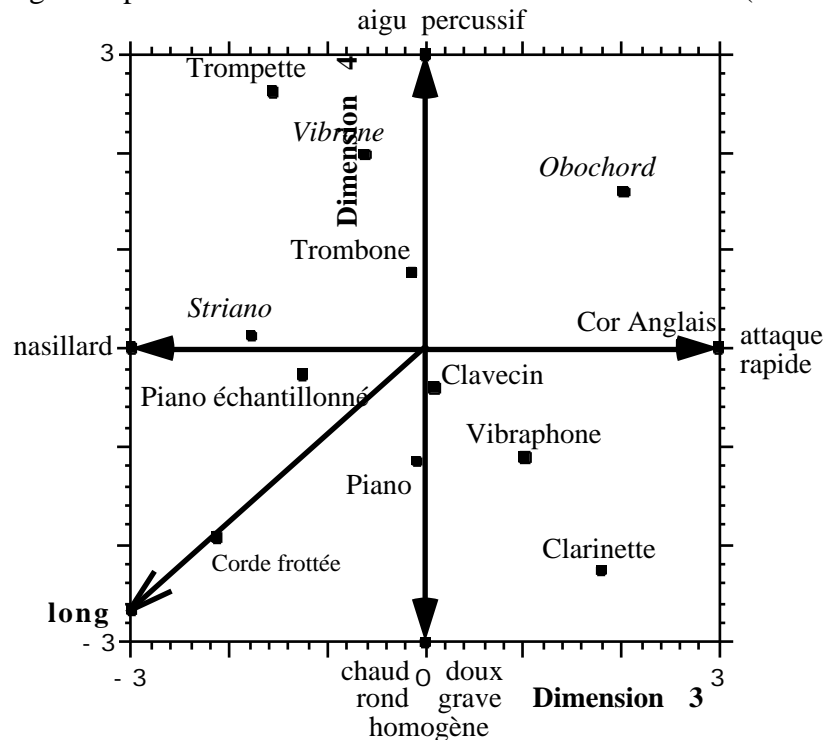




Figure 5.6. Plan engendré par les dimensions 3 et 4 et ses attributs verbaux (voir légende fig.5.1)



Enfin, la figure 6 reprend le fait que la notion de *long*, liée principalement à la perception du temps d'attaque des sons (dimension 1), est aussi liée aux variations des dimensions 3 et 4. Un son étant perçu comme d'autant plus *long* que sa déviation spectrale et que ses variations temporelles sont faibles ?

### I.3. Résumé

L'analyse multidimensionnelle des jugements de dissemblance obtenus dans cette étude, comparés à ceux obtenus dans la littérature sans verbalisations (McAdams et al., 1995), montre que les verbalisations n'affectent ni la structure, ni la répartition des timbres dans l'espace perceptif multidimensionnel, et ce même si on laisse aux sujets la possibilité de modifier leurs jugements après verbalisation. En effet, l'espace perceptif obtenu spontanément est le même que celui obtenu après réflexion, ce qui renforce sa crédibilité et justifie la possibilité d'une comparaison avec l'espace sémantique de ces sons.

Les deux premières dimensions apparaissent comme extrêmement robustes puisqu'on les retrouve quel que soit le sous-ensemble de timbres utilisés et les sujets interrogés (McAdams et al., 1995). Il ne serait donc pas étonnant que ces deux dimensions soient les plus saillantes perceptivement pour tout sujet. La perception d'une troisième et/ou d'une quatrième dimension dépend plus de l'ensemble de timbres étudiés et des sujets interrogés. Il est important de noter que l'espace obtenu par Krumhansl (1989) portait sur 9 sujets, tous musiciens ou psychoacousticiens professionnels habitués à travailler avec des sons de synthèse du type des stimuli utilisés, soit une population extrêmement

ciblée. Le nombre et la nature des éventuelles dimensions 3, 4, ... qui varient selon l'expérience peuvent donc être propre à une population donnée et à l'ensemble de timbres étudiés.

L'analyse d'une partie des verbalisations libres a permis de dégager quatre corrélats principaux de chacune des dimensions perceptives. Ceci est résumé dans le tableau 5.4.

Tableau 5.4. Principaux corrélats verbaux de chacune des 4 dimensions perceptives.

D1	D2	D3	D4
<i><b>net</b></i>	<i>pas rond</i>	<i>pas long</i>	<i><b>doux</b></i>
<i>attaque rapide</i>	<i>pas grave</i>	<i><b>pas riche</b></i>	<i>rond</i>
<i><b>clair</b></i>	<i>aigu</i>		<i>agréable</i>
<i>métallique</i>	<i>pas chaud / chaleureux</i>		<i>homogène</i>

Les attributs verbaux corrélés avec une seule dimension perceptive sont en gras.

La grande majorité des mots obtenus lors des verbalisations libres présentait un caractère composite des dimensions perceptives. Ceci peut être dû au mélange des utilisations partielles d'un même attribut par divers sujets et refléter des variations sémantiques de chaque mot suivant le locuteur. Il se peut aussi que l'espace perceptif et l'espace sémantique ne coïncident pas directement, ce qui reste à tester de manière plus systématique. Notre deuxième expérience a été construite dans ce but : pouvoir réunir, à propos de chacun des attributs verbaux sélectionnés ici, les réponses de chaque sujet pour chaque son. Ceci afin de permettre une comparaison plus systématique de l'utilisation de ces attributs verbaux avec l'espace perceptif de ces sons.

#### I.4. Comparaison avec le vocabulaire de la littérature utilisant des échelles sémantiques

Nous avons souhaité comparer le vocabulaire recueilli ici spontanément avec celui utilisé dans la littérature sur des échelles sémantiques, d'une part la liste initiale élaborée par Osgood et al. (1957) et utilisée de façon générale sur des stimuli variés, d'autre part, le vocabulaire utilisé dans d'autres études s'intéressant aux sons : sons de sonar pour Solomon (1957), sons complexes de musique, de voitures et des sons artificiels pour Kerrick et al. (1969), un ensemble de timbres musicaux et de bruits pour Von Bismarck (1974a), et enfin sons d'instruments à vent joués en duos pour Kendall & Carterette (1992a, b). Ce dernier vocabulaire étant extrait du traité d'orchestration de Piston (1955), nous comparerons l'ensemble de ce vocabulaire à celui que nous avons recueilli. Enfin, notre vocabulaire sera aussi comparé à celui utilisé par Wedin (1972) dans une procédure de choix d'adjectifs dans une liste lors de l'écoute d'extraits musicaux.

Pour finir, et dans le but de comparer notre vocabulaire à celui utilisé pour des stimuli appartenant à d'autres modalités, nous allons considérer les travaux de Tucker (1955) sur des tableaux, abstraits ou non, jugés par des artistes peintres et des néophytes ainsi que le vocabulaire standard utilisé pour décrire les odeurs rassemblé par Dravnieks (1978, 1982). La plupart de ce vocabulaire étant composé de noms de sources odorantes, seuls les adjectifs ont été comparés à notre vocabulaire.

L'ensemble des verbalisations spontanées de notre expérience 1 fait apparaître une liste de 1124 mots différents, pertinents dans le contexte de la description de sons complexes dont 514 adjectifs qui vont être comparés aux échelles utilisées dans la littérature. Ces comparaisons sont rassemblées dans le tableau 5.5.

Tableau 5.5. Comparaison de notre vocabulaire avec celui de la littérature. Pour chaque auteur, nombre de mots testés (NB), pourcentage (%) pour lequel un équivalent en français a été produit spontanément par nos sujets, et  $r^2$  entre les listes de mots, 2 à 2, ( $dl=1$ ).

Auteurs	NB	%	N.E.	1	2	3	4	5	6	7	8
1. Osgood et al. (1957)	100	65	34 *								
2. Solomon (1957)	100	59	5 †	26 *							
3. Kerrick et al. (1969)	30	60	5 †	47 *	20 *						
4. von Bismarck (1974)	60	67	23 *	31 *	36 *	19 *					
5. Kendall & Carterette(1992)	21	67	6 †	3	0	2	16 *				
6. Piston (1955)	177	48	8 ‡	3	1	0	1	32 *			
7. Wedin (1972)	101	36	1	8 ‡	1	3	2	0	19 *		
8. Tucker (1955)	80	51	5 †	28 *	17 *	3	9 ‡	3	2	1	
9. Drawnieck (1982)	25	48	1	1	1	0	1	1	2	2	0

Avec N.E. : notre étude (514 adjectifs), \*:  $p<.001$ , ‡:  $p<.01$ , †:  $p<.05$

De manière générale, les verbalisations libres permettent de dégager cinq fois plus d'adjectifs utilisés pour décrire le timbre que les autres études de la littérature. En moyenne, pour 57% de ces mots, un équivalent en Français du vocabulaire utilisé dans ces travaux a été spontanément produit par nos sujets. Ce qui veut dire que globalement 43% du vocabulaire utilisé dans ces études n'est pas spontanément utilisé par un groupe de sujets d'expériences musicales variées pour parler du timbre musical, et ce malgré le fait que nos sujets produisent beaucoup plus de vocabulaire que celui testé. Le vocabulaire testé dans les protocoles de sémantique différentielle appliqués aux sons n'est donc souvent pas naturel pour les sujets.

Les mesures du lien entre les listes de mots 2 à 2, obtenues par calcul de  $r^2$  entre les mots communs et spécifiques à chacune des listes, font apparaître deux groupes de travaux : ceux dans la lignée d'Osgood et al. (1957) utilisant des techniques de sémantique différentielle et partant d'une base d'échelles sémantiques bipolaires communes (c'est-à-dire ceux de Solomon, Kerrick et al., von Bismarck et Tucker) ; ceux de Kendall & Carterette (1992a ; 1992b) qui se rapprochent de ceux de Piston (1955) et de von Bismarck (1974a), leur étude se voulant un test du vocabulaire utilisé par von Bismarck et une amélioration de ce vocabulaire en considérant le traité d'orchestration de Piston. Il est étonnant de remarquer que la liste de mots la plus liée au vocabulaire spontanément produit par nos sujets est celle d'Osgood et al. (1957) qui est pourtant celle regroupant des mots peu spécifiques de la modalité auditive, vu qu'ils sont censés pouvoir s'appliquer à tout domaine d'étude. Ceci rejoint notre remarque sur le fait que le vocabulaire employé spontanément par les sujets pour parler du timbre des sons est très peu spécifique du domaine sonore (cf. chapitre 6).

## **II. Expérience II : Recherche de corrélats parmi les échelles sémantiques**

Nous allons reprendre ici la recherche d'éventuels corrélats verbaux des dimensions perceptives, cette fois-ci parmi les échelles sémantiques proposées dans l'expérience 2.

Il s'agit ici d'analyser les corrélations entre les jugements effectués par les sujets sur chacune des échelles sémantiques proposées dans cette expérience d'une part, et la position des sons dans l'espace perceptif multidimensionnel déduit lors de notre première expérience des jugements de dissemblance perceptive entre les paires de ces mêmes sons.

Les analyses sont effectuées tout d'abord à partir des résultats moyennés sur des groupes de sujets différents, puis on effectue les mêmes analyses sujet par sujet, ce qui est rendu possible grâce au caractère systématique du recueil de données. Enfin, nous comparons ces résultats à ceux obtenus sur l'utilisation des différentes échelles sémantiques chapitre 4 §II.6 ainsi qu'aux jugements de pertinence effectués par les sujets sur ces dimensions (voir chapitre 7, §II). Enfin, nous comparons les résultats obtenus de manière analogue dans le cadre de notre première expérience.

### II.1. Analyses à partir des résultats moyennés sur des groupes de sujets

Nous avons calculé 4 types de moyennes, à chaque fois sur l'ensemble des 32 sujets et sur les deux groupes de sujets correspondant à des niveaux d'expertise musicale différents (musiciens et non-musiciens), ce qui nous amène à  $4 \times 3 = 12$  moyennes différentes calculées.

#### **II.1.1 Différentes moyennes calculées**

Nous avons tout d'abord calculé des moyennes à partir des jugements bruts des sujets, après avoir effectué les corrections mentionnées dans les analyses préliminaires (chapitre 4 §II.6), dans les cas où les sujets exprimaient une erreur de leur part et les cas d'inversion nette du sens d'une échelle. Ces moyennes sont notées : Moyennes, M-musiciens, M-non-musiciens.

Ensuite, nous avons calculé d'autres moyennes uniquement à partir des réponses jugées pertinentes par le sujet qui les donnait, les jugements non pertinents étant enlevés du calcul de la moyenne (Attention : chaque moyenne n'est donc plus calculée à partir du même nombre de jugements !!). Cette méthode utilise le jugement de pertinence donné après chaque positionnement sur chaque échelle sémantique lors de la séance 2 : "ce mot me semble (ou ne me semble pas) approprié pour décrire ce son". Ces moyennes sont notées PM (pour Pertinentes), PM-musiciens, PM-non-musiciens.

Après, considérant l'observation que les échelles sont utilisées avec une amplitude de jugement très différente selon les sujets et les mots considérés, nous avons normalisés les jugements en pleine échelle, sauf pour les échelles volontairement utilisées de façon très resserrées (chapitre 4 §II.6). Ainsi, les jugements des sujets qui répartissaient leur jugements sur moins de 30% de l'échelle ont été

éliminés de ces moyennes (en considérant que ce type d'utilisation est un synonyme de dire que l'échelle ne convient pas pour différencier les sons). Ces moyennes sont notées NM (pour Normées), NM-musiciens, NM-non-musiciens.

Enfin, nous avons combiné les deux derniers types de moyennes en calculant une nouvelle moyenne en normant les échelles, mais uniquement sur les jugements considérés pertinents par leur auteur. Ces moyennes sont notées NPM (pour Normées et Pertinentes), NPM-musiciens et NPM-non-musiciens.

Nous avons calculé les corrélations entre ces 12 moyennes différentes pour chacune des 23 échelles proposées, ce qui fait 66 coefficients de corrélation inter-moyennes. Les résultats sont rassemblés dans le tableau de l'annexe A.5.1. Globalement, les corrélations entre les diverses moyennes calculées sont très fortes. En effet, pour 14 échelles (soit 61% d'entre elles), toutes les corrélations inter-moyennes sont très significatives ( $r > .83$ ,  $p < .01$ ,  $dl=10$ ). Donc pour les échelles liées à *aigu*, *attaque rapide*, *clair*, *doux*, *grave*, *haut*, *métallique*, *nasillard*, *net*, *pincé*, *résonnant*, *sec*, *soufflé* et *sourd*, le mode de calcul de la moyenne - que ce soit sur des groupes de sujets différents ou en tenant compte ou non des jugements de pertinence ou des échelles resserrées - n'intervient pas de manière significative. Pour les 9 échelles restantes, qui présentent donc quelques corrélations non significatives entre les différentes moyennes calculées, 2 échelles présentent moins de 10% des corrélations inter-moyennes non significatives : *rond* et *brillant*.

Pour *rond*, on remarque que les corrélations les plus faibles correspondent toutes à des corrélations faisant intervenir une moyenne calculée sur le groupe des musiciens, comparés à une moyenne calculée sur le groupe des non-musiciens. C'est le cas notamment des deux corrélations non significatives qui correspondent aux comparaisons : PM-musiciens / PM-non-musiciens et PM-musiciens / NM-non-musiciens. C'est aussi le cas des corrélations correspondant aux 4 corrélations significatives avec seulement  $p < .05$ , et aux 10 dernières corrélations comparant un groupe musicien et un autre non-musicien qui sont significatives à  $p < .01$ , mais avec un coefficient de corrélation  $r < .9$  (alors que les autres corrélations significatives à  $p < .01$  ont un  $r > .9$ ). *Rond* serait donc une échelle qui serait jugée de manière significativement différente par les musiciens et les non-musiciens...

Pour *brillant*, les 4 corrélations non significatives contiennent toutes, dans un des deux éléments de la comparaison, la moyenne PM-musiciens, calculée sur le groupe des musiciens à partir des jugements pertinents, et dans l'autre élément de la comparaison une des 4 autres moyennes effectuées sur le groupe de non-musiciens. On trouve donc, là aussi, des différences entre les jugements des musiciens et des non-musiciens pour l'échelle liée à *brillant*, avec une particularité des jugements pertinents des musiciens. Il est intéressant de remarquer que les 6 autres comparaisons moins significatives (à  $p < .05$ ) comportent elles aussi un terme calculé sur le groupe des musiciens et un autre sur le groupe des non-musiciens. On peut aussi remarquer que l'échelle liée à *brillant* est considérée comme globalement pertinente par 53% des musiciens contre 71% des non-musiciens (voir chapitre 7), alors que si on regarde pour quels sons ces deux groupes de sujets ont considéré que cette échelle est

adaptée, on trouve 6 sons pour les musiciens (dont deux pour plus de 75% des sujets musiciens) alors que les non-musiciens en plébiscitent seulement 4 d'entre eux, et pour des pourcentages de sujets plus faibles. Cela pourrait peut-être expliquer la particularité des données pertinentes des musiciens.

Si nous considérons maintenant les 3 échelles qui présentent 11% des corrélations inter-moyennes non significatives : *diffus*, *fort* et *long*, on retrouve ici aussi dans toutes les corrélations les moins significatives des comparaisons entre une moyenne calculée sur le groupe des musiciens et une autre sur le groupe des non-musiciens. Ce sont toutes trois des échelles jugées globalement peu pertinentes par les sujets, qu'ils soient musiciens ou non, (respectivement 57, 23 et 52 % des sujets, voir chapitre 7). Par contre si on regarde les sons pour lesquels chacun des groupes de sujets, musiciens et non-musiciens, considère que ces échelles sont adaptées, on trouve des différences. En effet, par exemple l'échelle liée à *diffus* est considérée comme adaptée par plus de 50% des non-musiciens pour 9 des 12 sons alors que ce n'est le cas que pour 6 des 12 sons pour les musiciens. De même pour l'échelle liée à *fort*, les non-musiciens la considèrent comme adaptée pour les sons PNO et VBN alors que les musiciens considèrent seulement PNO. C'est encore le même cas pour *long*, considéré comme adapté pour 7 sons par les non-musiciens et seulement pour 5 sons par les musiciens, deux sons n'appartenant pas aux deux groupes : 56% des musiciens considèrent que l'échelle liée à *long* est adaptée pour le son CLA alors que c'est le cas pour seulement 50% des non-musiciens ; inversement pour le son CRA pour lesquels l'échelle est adaptée pour 44% des musiciens et 56% des non-musiciens.

Puis viennent les 4 échelles qui présentent le plus de corrélations inter-moyennes non significatives (entre 18 et 36%). Il s'agit des échelles liées à *bas*, *vibré*, *large* et *riche*. On retrouve ici aussi le même principe : les corrélations les moins significatives sont liées à une différence de compétence entre les groupes de musiciens et de non-musiciens, même si cela apparaît de manière moins systématique.

Il est important de noter que toutes ces échelles (les 4 dernières plus les 3 précédentes) font partie de celles qui ont la plus mauvaise corrélation inter-sujets. En effet, ces 7 échelles totalisent toutes moins de 5% des corrélations inter-sujets significatives (et ce, même si on considère le seuil de significativité  $p < .1$ ). Ceci pourrait expliquer le fait que les données de chacun des sujets étant peu corrélées entre elles, dès qu'on calcule une moyenne sur un groupe de sujets différents ou en éliminant les données de certains sujets (par exemple ceux qui utilisent l'échelle de manière très resserrée ou ceux qui la considèrent comme pas adaptée à certains sons), on modifie significativement les résultats.

Les corrélations entre les différents types de moyennes que nous avons calculées étant beaucoup plus fortes que ce à quoi nous nous attendions, nous poursuivrons les analyses uniquement sur les moyennes calculées de manière classique. Les résultats obtenus à partir des autres moyennes (P, N, et NP) ne seront indiqués que dans la mesure où ils diffèrent des résultats obtenus à partir de la moyenne calculée de manière classique.

### II.1.2. Corrélations avec les dimensions perceptives

Nous avons effectué les corrélations pour chaque échelle sémantique entre les moyennes des jugements des 32 sujets et les positions sur chacune des dimensions perceptives. Les résultats (corrigés pour 23 analyses selon Bonféroni) sont rassemblés dans le tableau 5.6.

Tableau 5.6. Coefficients de corrélation (dl=10) des moyennes des jugements des 32 sujets avec les dimensions perceptives.

Échelles	D1.4	D2.4	D3.4	D4.4
aigu	0.51	0.73	0.17	0.20
attaque	0.87 *	0.03	0.15	0.25
bas	-0.50	-0.75	-0.10	-0.19
brillant	0.68	0.57	0.07	0.12
clair	0.76	0.31	0.08	-0.11
diffus	-0.37	-0.43	-0.37	-0.17
doux	-0.12	-0.62	-0.04	-0.65
fort	0.19	0.11	0.01	0.19
grave	-0.42	-0.83 *	-0.08	-0.28
haut	0.43	0.77	0.14	0.26
large	-0.04	-0.65	-0.36	-0.56
long	-0.02	-0.25	-0.35	-0.37
métallique	0.74	0.58	-0.04	0.31
nasillard	-0.46	0.60	0.06	0.62
net	0.67	0.35	0.31	-0.07
pincé	0.88 *	0.36	0.17	0.14
résonnant	0.81 †	-0.30	-0.19	-0.09
riche	0.49	0.00	-0.55	0.19
rond	-0.29	-0.83 *	-0.08	-0.44
sec	0.66	0.45	0.21	0.33
soufflé	-0.96 *	-0.15	-0.03	-0.01
sourd	-0.66	-0.55	-0.04	0.06
vibré	0.71	0.10	-0.26	0.21

avec \*:  $p < .01$  ; †:  $p < .05$  (ajustés selon Bonféroni pour 23 analyses)

On remarque qu'au contraire des résultats obtenus à partir des verbalisations libres (voir §5.I), très peu d'échelles sémantiques sont corrélées à chacune des dimensions perceptives et jamais à plusieurs en même temps. La notion d'échelle composite de plusieurs dimensions perceptives disparaît donc dans le cadre du paradigme de la sémantique différentielle. Les échelles pouvant être considérées comme des corrélats verbaux des dimensions perceptives sont donc rares : 6 seulement sur les 23 proposées (soit 26% d'entre elles).

Quatre échelles sont corrélées à la dimension 1 : celles liées à *soufflé*, *pincé*, *attaquerapide* suivi de *résonnant*. Ces 4 termes renvoient tous directement à l'idée d'un paramètre temporel, que ce soit le début du son, l'*attaque*, ou sa fin, la *résonance*. L'*attaque* pouvant être caractérisée par son mode d'émission : *pincé* (souvent confondu perceptivement avec frappé) ou *soufflé* (souvent confondu avec frotté) qui permettent de catégoriser les temps d'*attaque* en *attaque rapide* (pour *pincé*) ou *lente* (pour *soufflé*).

Deux échelles sont utilisées de manière corrélée à la dimension 2 : celles liées à *grave* et *rond*. Aucune des échelles sémantiques proposées n'est significativement corrélée aux dimensions 3 et 4. Sur les 23 échelles sémantiques, 17 ne sont corrélées à aucune dimension.

Remarque : Si on ne corrigeait pas les seuils selon Bonféroni pour les 23 analyses effectuées en parallèle, on obtiendrait un bien plus grand nombre d'échelles corrélées à des dimensions perceptives : 19 sur 23 (soit 82% contre 23% sinon). En effet, la dimension 1 aurait trois nouveaux corrélats verbaux *clair*, *métallique* et *vibré* (avec  $r > .71$  et  $p < .01$ ) plus 4 autres (avec  $r > .66$  et  $p < .05$ ) : *brillant*, *nasillard*, *sec* et *sourd*. Pour la deuxième dimension, de même, on obtiendrait 3 nouveaux corrélats verbaux (avec  $r > .73$  et  $p < .01$ ) : *haut*, *bas* et *aigu* plus 5 autres (avec  $r > .57$  et  $p < .05$ ) : *large*, *doux*, *nasillard*, *métallique* et *brillant*. Enfin la dimension 4 aurait un corrélât verbal : *doux* (avec  $r = -.65$  et  $p < .05$ ). On se retrouverait alors avec 3 attributs verbaux composites de deux dimensions perceptives : *brillant* et *métallique* pour D1 et D2 et *doux* pour D2 et D4.)

### II.1.3. Différences entre musiciens et non-musiciens

Toujours à partir des moyennes sur les jugements bruts corrigés des sujets, nous avons comparé les résultats obtenus en calculant les moyennes correspondant aux groupes des musiciens (M) et des non-musiciens (nM). Par rapport aux résultats obtenus à partir de l'ensemble des sujets, il y a quelques différences de comportement entre le groupe des musiciens et celui des non-musiciens (voir tableau A5.2 en annexe).

Pour la dimension 1, les échelles liées à *soufflé*, *pincé* et *attaquerapide* restent corrélées à cette dimension quel que soit le groupe de sujets considéré (bien que la corrélation de *pincé* soit moins forte pour le groupe des musiciens ( $p < .05$  au lieu de  $p < .01$ ). Il y a des différences entre les groupes de sujets considérés pour les échelles liées à *résonnant*, *clair* et *métallique*. En effet, l'échelle liée à *résonnant*, qui est globalement corrélée à la dimension 1 ne l'est en fait que pour le groupe des non-musiciens. Ceux-ci semblent utiliser aussi l'échelle *métallique* de manière corrélée à la dimension 1. Par contre pour les musiciens, c'est l'échelle liée à *clair* qui est corrélée à cette dimension. Il est difficile d'interpréter cette différence, *clair* n'étant pas un terme spécifiquement employé dans le monde des musiciens. On peut par contre comprendre que la notion de *métallique* soit liée à un caractère temporel : les sons de notre ensemble imitant des modes de production pincés ou frappés (de temps d'attaque rapide) imitant tous des sons produits sur des pièces de métal (corde ou plaques). Pourtant deux des sons imitaient des instruments de la famille des cuivres (trompette et trombone) de temps d'attaque lent, donc situés à l'autre extrémité de la dimension 1 et qui peuvent eux aussi être qualifiés de métalliques (surtout le son TPT).

Pour la dimension 2 les différences entre les groupes sont encore plus marquées. En effet, seul *rond* est corrélé à la dimension 2 quel que soit le groupe de sujets considérés. C'est d'ailleurs la seule échelle corrélée à cette dimension pour le groupe des non-musiciens. Les musiciens y ajoutent les échelles liées à *grave*, *aigu*, *haut* et *bas*, soit les 4 termes qui semblent liés à la notion de hauteur tonale



dans leur utilisation habituelle. Ces termes sont considérés comme les moins pertinents pour décrire ces sons (qui sont égalisés en hauteur), et en général d'autant plus par les musiciens pour qui *aigu* a un emploi quasiment conventionnel pour désigner la hauteur tonale. Il est intéressant de remarquer que si on force les musiciens à utiliser tout de même cette échelle pour qualifier nos sons, ils utilisent leur perception du Centre de Gravité Spectral (CGS\*) correspondant à la deuxième dimension pour positionner les sons.

On pourrait donc expliquer le changement de comportement entre une utilisation spontanée et une utilisation systématisée sous forme d'échelles sémantiques des mêmes mots pour décrire les mêmes sons ainsi :

- Quand il s'agit de décrire spontanément un son, on emploie le terme qui nous vient à l'esprit à ce moment, donc de façon spécifique à chaque fois par rapport à ce qu'on a perçu. On se sent peut-être moins obligé de "normaliser" une utilisation sur l'ensemble des sons écoutés ; on décrit l'un d'eux, c'est tout. Lorsqu'on rassemble de telles utilisations, on peut s'attendre à ce que le même terme soit utilisé avec des nuances légèrement différentes suivant le sujet considéré et le timbre décrit. On obtient donc une utilisation qui paraît composite de plusieurs dimensions perceptives. (Pourquoi est-ce beaucoup corrélé aux dimensions ??).
- Lorsqu'on demande à chacun des sujets d'utiliser chacun des mots de façon systématique pour chacun des sons proposés, ceux-ci sont obligés de considérer le sens du mot globalement par rapport à l'ensemble des sons, ce qui va souvent lui faire penser qu'il n'est pas très adapté. Et dans son "désarroi", il va "aligner" l'utilisation de certains mots sur une des perceptions les plus saillantes, celles qui correspondent aux axes de l'espace perceptif.

Pourquoi moins d'échelles corrélées dans cette expérience que dans la première ?

#### **II.1.4. Différences inter-mode de calcul des moyennes**

Lors de l'analyse des données et en fonction des remarques sur les données brutes faites lors des analyses préliminaires sur l'utilisation des échelles sémantiques et sur les jugements de pertinence effectués par les sujets, nous avons jugé bon de calculer plusieurs types de moyennes. Les différences obtenues dans les corrélations des échelles avec les dimensions perceptives entre les 4 types de moyennes calculées sont présentées dans le tableau A5.3 en annexe. Les calculs ont été effectués pour l'ensemble des 32 sujets.

On observe très peu de différences entre les différentes moyennes.

L'échelle liée à *résonnant* est plus fortement corrélée à la dimension 1, quand les moyennes sont calculées à partir des jugements normalisés. Cette échelle n'avait pourtant été utilisée de manière resserrée que par un seul des sujets, la plupart se répartissant entre une utilisation normale et dilatée de cette échelle. Peut-être que le fait d'enlever les données du sujet qui a utilisé cette échelle de manière très resserrée et de dilater les données de ceux qui l'ont utilisée de manière normale, suffit à expliquer l'augmentation de la corrélation de cette échelle avec la première dimension perceptive.

L'échelle liée à *haut* n'est corrélée à la dimension 2 que pour les moyennes calculées à partir des jugements pertinents, qu'ils soient normalisés ou non. Ces jugements pertinents étant très rares pour cette échelle (en moyenne, 38% des sujets seulement considèrent cette échelle comme adaptée pour chacun des 12 sons proposés), on peut remarquer que les sons pour lesquels cette échelle est considérée comme la plus pertinente sont les sons CVC et OBC, qui sont les sons dont le CGS\* est le plus élevé, positionnés à un des extrêmes de la dimension 2.

On retrouve donc globalement le fait que les différents types de calculs de moyenne sur ces données sont très fortement corrélés les uns aux autres.

### II.1.5. Résumé

Les corrélations entre les positions des 12 sons sur les dimensions perceptives et les moyennes des jugements des sujets sur les 23 échelles sémantiques proposées sont résumées dans le tableau 5.7.

Tableau 5.7. Corrélats des dimensions perceptives. Données moyennées.

D1.4	D2.4	D3.4	D4.4
<b>pas soufflé</b>	<b>pas rond</b>		
<b>pincé</b>	<b>pas grave</b>		
<b>attaque rapide</b>	haut	∅	∅
résonnant	pas bas		
clair	aigu		
métallique			

En gras, les corrélats obtenus quel que soit le groupe de sujets considéré et le mode de calcul de la moyenne.

### II.2. Analyses sujet par sujet

L'intérêt du paradigme de sémantique différentielle, c'est justement d'avoir l'utilisation de chacune des échelles par chacun des sujets (pour l'ensemble des 12 sons). Ceci nous permet d'observer comment le même mot pourrait être utilisé de manière différente par différents sujets. En effet, lorsque nous avons analysé les corrélats verbaux des dimensions perceptives dans le cadre de notre première expérience, nous obtenions beaucoup d'attributs verbaux composites de plusieurs dimensions perceptives et nous avons envisagé comme une explication possible le fait d'avoir mélangé des utilisations partielles du même mot par des sujets différents. Nous avons donc essayé de chercher, pour chacun des sujets si l'utilisation qu'il faisait des échelles sémantiques proposées était corrélée à une de nos 4 dimensions perceptives.

Nous avons analysé, pour chaque sujet, les corrélations (sur les jugements bruts non corrigés de la séance 2) entre l'utilisation d'une échelle sémantique et la répartition des sons sur chacune des 4 dimensions de l'espace perceptif (voir tableau 5.8). Ces résultats n'ont pas été corrigés selon Bonféroni car alors les résultats chutent dramatiquement ! Nous comparerons les deux types de résultats en fin de paragraphe.

**Tableau 5.8.** Pourcentage de sujets ayant des corrélations entre le positionnement des sons sur une échelle sémantique et sur chacune des dimensions perceptives avec  $r > .58$  et  $p < .05$

(dont pourcentage avec  $r > .88$ ,  $p < .0001$ )

	D 1.4	D 2.4	D 3.4	D 4.4
aigu	28	44	0	6
attaque	75 (19)	0	0	3
bas	9	5	0	3
brillant	31	22	0	6
clair	5	19	0	0
diffus	28	13	6	16
doux	6	28	0	38
fort	13	0	0	0
grave	13	44 (3)	0	0
haut	22	5 (6)	0	0
large	47 (6)	19	6	0
long	31 (6)	19	6 (3)	6
métallique	66 (13)	19	6	9
nasillard	9	44	0	22
net	47	3	6	6
pincé	69 (28)	22	3	0
résonnant	53 (9)	6	0	3
riche	19	16	6	6
rond	25 (9)	47 (9)	0	9
sec	53 (3)	16	6	3
soufflé	66 (34)	3	0	0
sourd	34	19	3	9
vibré	38 (3)	13	3	6

Attention ces valeurs n'ont PAS été ajustées selon Bonféroni.

De manière générale, on remarque que pour certaines échelles, les sujets semblent utiliser les échelles de manière corrélée à plusieurs dimensions perceptives. Nous analyserons plus loin de manière systématique s'il s'agit des mêmes sujets, qui jugent donc une échelle de manière composite de plusieurs dimensions perceptives, ou s'il s'agit de sujets différents qui utilisent les échelles de manière différente.

Globalement, la première dimension remporte le plus vif succès.

En effet, les pourcentages de sujets dont l'utilisation des échelles sémantiques est significativement corrélée avec la dimension 1 sont plus forts qu'avec la dimension 2, eux-mêmes plus forts qu'avec les dimensions 3 et 4. Il y a pourtant quelques exceptions : l'échelle liée à *doux* est corrélée par le plus fort pourcentage de sujets avec la dimension 4 (pour 38% des sujets). Celles liées à *aigu*, *grave*, *nasillard* et *clair* sont corrélées par le plus fort pourcentage de sujets avec la dimension 2 (respectivement par 44% des sujets pour les trois premières échelles et par 19% des sujets pour l'échelle liée à *clair*).

Quand on trouve des sujets qui utilisent une échelle sémantique de manière significativement corrélée avec la dimension 2, cette échelle est toujours aussi corrélée à la dimension 1 pour un certain nombre de sujets. De plus, toutes les échelles sémantiques sont significativement corrélées avec la dimension

1 pour au moins 5% des sujets. Ceci est peut-être dû au choix du vocabulaire pour construire les échelles sémantiques. En effet, comme les mots choisis sont les plus fréquemment utilisés pour comparer les sons 2 à 2, ce sont peut-être justement les mots qui correspondent tous à la dimension perceptive la plus saillante du timbre musical, c'est-à-dire la sensation d'attaque révélée par la dimension perceptive n°1.

Les échelles sémantiques dont l'utilisation est corrélée avec la dimension 3 sont peu nombreuses (10 sur 23) et à chaque fois c'est pour au maximum 6% des sujets (soit deux sujets différents). De plus, elles sont toutes aussi corrélées aux dimensions 1 et 2 et parfois aussi à la dimension 4 (8 sur 10). Cette dimension semble donc difficile à atteindre spécifiquement à l'aide des 23 échelles sémantiques testées.

### II.2.1. Attributs verbaux de chaque dimension perceptive

#### Dimension 1

On retrouve parmi les 4 échelles sémantiques corrélées par le plus grand nombre de sujets à la première dimension les 4 échelles obtenues lors de l'analyse sur les données moyennes des sujets, à savoir celles liées à *attaque rapide*, *pincé*, *soufflé* et *métallique*.

L'échelle sémantique *attaque rapide / attaque lente* est celle dont l'utilisation est corrélée à une dimension par le plus grand nombre de sujets. En effet, 75% de ceux-ci l'utilisent de manière corrélée à la position des sons sur la dimension 1 (liée au logarithme du temps de montée de l'enveloppe spectrale LTM\*). De plus, 19% de ces corrélations sont significatives avec  $p < .0001$ . Ces sujets utilisent donc directement l'échelle sémantique de façon analogue à la position sur la dimension perceptive n°1 déduite des jugements de dissemblance perceptive entre les paires de sons.

Ensuite, vient l'échelle *peu pincé - très pincé*, dont l'utilisation est corrélée à la dimension 1 pour 69% des sujets et à la dimension 2 pour 22% des sujets (3% des sujets l'utilisent de manière corrélée à la dimension 3). Il est intéressant de remarquer que 28% des sujets utilisent cette échelle de manière corrélée à la position sur la dimension 1 avec  $r > .88$  et  $p < .0001$  alors qu'aucune corrélation avec une des autres dimensions perceptives n'atteint ce niveau de significativité.

Puis, viennent les échelles liées à *métallique* et à *soufflé* (avec 66% des sujets). Pour l'échelle *peu soufflé - très soufflé*, déjà 34% des sujets utilisent cette échelle de manière corrélée à la position sur la dimension 1 avec  $r > .88$  et  $p < .0001$ . Ce qui est le plus fort pourcentage à ce niveau de significativité. Cette échelle est aussi corrélée pour 3% des sujets à la dimension 2. Pour l'échelle *peu métallique - très métallique*, son utilisation est déjà corrélée à la dimension 1 par 19% des sujets avec  $p < .0001$ . Cette échelle est aussi corrélée pour 19% des sujets avec  $p < .05$  à la dimension 2.

Ensuite, on a 53% des sujets qui utilisent les échelles *peu résonnant - très résonnant* et *peu sec - très sec* de manière corrélée à la dimension 1. Ces deux échelles sont aussi corrélées à la dimension 2

(respectivement par 6 et 16% des sujets) et à la dimension 4 (3% des sujets), l'échelle liée à *sec* étant aussi corrélée à la dimension 3 (par 6% des sujets).

Les autres corrélations avec la dimension 1, sont toutes au-dessous du seuil de 50% des sujets. On trouve ainsi les échelles liées à *net* et *large* (pour 47% des sujets), l'échelle liée à *large* étant plus souvent associée à d'autres dimensions que celle liée à *net*. Viennent ensuite les échelles liées à *vibré* (38%), puis *sourd* (34%), *brillant* et *long* (31%).

### Dimension 2

On retrouve ici les résultats globaux obtenus à partir des données moyennées. En effet, parmi les échelles corrélées à la dimension 2 par le plus grand nombre de sujets, on retrouve celles liées à *rond*, *aigu* et *grave*, mais aussi celle liée à *nasillard*. L'échelle sémantique utilisée de manière corrélée à la deuxième dimension par le plus grand nombre de sujets est celle liée à *rond* (47% des sujets). Viennent ensuite les échelles liées à *aigu*, *grave* et *nasillard* avec 44% des sujets. Parmi celles-ci, l'échelle liée à *aigu* est celle qui est le plus souvent associée à la dimension 1, l'échelle liée à *nasillard* est aussi associée à la dimension 4 par 22% des sujets.

On tombe après sous la barre des 30% de sujets avec l'échelle liée à *doux* : 28% des sujets utilisent cette échelle de manière corrélée à la dimension 2 mais aussi 38% d'entre eux l'utilisent de manière corrélée à la dimension 4, ce qui en fait le corrélat le plus fort de cette dernière dimension.

### Dimension 3

Toute corrélation de l'utilisation d'une échelle sémantique avec cette dimension perceptive l'est pour moins de 6% de sujets. Si on considère le seuil de significativité de  $p < .0001$ , seule l'échelle liée à *long* est corrélée pour 3% des sujets avec la dimension 3. Mais cette échelle est surtout fortement corrélée à la dimension 1. Cette dimension semble donc difficile à considérer à partir de l'ensemble de nos échelles sémantiques. C'est aussi celle pour laquelle on ne trouve aucun corrélat acoustique dans l'ensemble de paramètres calculés par Krimphoff (1993).

### Dimension 4

Comme nous l'avons mentionné plus haut, l'échelle liée à *doux* est celle qui est le plus souvent corrélée à cette dimension (pour 38% des sujets). Il n'y a que deux autres échelles, celles liées à *nasillard* et *diffus* (respectivement utilisées de manière corrélées à la dimension 4 par 22% et 19% des sujets) qui valent la peine d'être aussi citées, bien que toutes deux aient des corrélations en-dessous du seuil de 30% des sujets.

#### **II.2.2. Des attributs verbaux composites des dimensions perceptives ?**

Nous avons observé ci-dessus que la plupart des échelles sémantiques étaient utilisées de manière corrélée à plusieurs dimensions perceptives (toutes sauf celle liée à *fort*). La question est de savoir si ce sont les mêmes sujets qui utilisent une même échelle sémantique de manière corrélée à plusieurs

dimensions perceptives, ce qui pourrait rejoindre la notion d'attributs verbaux composites de plusieurs dimensions perceptives avancée lors de l'analyse des résultats de l'expérience 1, ou si au contraire, quand il y a plusieurs dimensions perceptives corrélées à l'utilisation d'une même échelle sémantique, c'est le fait de plusieurs sujets différents. Nous avons donc calculé, pour chaque échelle sémantique, le nombre de sujets dont les données étaient corrélées à chacune des dimensions et à deux d'entre elles à la fois. Ces résultats sont rassemblés dans le tableau 4.16. Ainsi par exemple, alors que respectivement 8 et 15 sujets utilisent l'échelle sémantique *peu rond - très rond* de manière corrélée aux dimensions 1 et 2, une seule personne présente les deux corrélations simultanées sur les 8 cas où cela aurait pu être possible.

Tableau 5.9. Nombre de sujets utilisant chaque échelle de manière corrélée (avec  $r > .58$  et  $p < .05$ ) avec les dimensions suivantes.

	D1	D2	D3	D4	D1&D2	D1&D3	D1&D4	D2&D3	D2&D4	D3&D4
aigu	9	14	.	2	.	.	.	.	.	.
attaque	24	.	.	1	.	.	.	.	.	.
bas	3	13	.	1	.	.	.	.	.	.
brillant	1	7	.	2	1	.	.	.	.	.
clair	13	6	.	.	1	.	.	.	.	.
diffus	9	4	2	5	.	.	2	.	.	.
doux	2	9	.	12	1	.	.	.	2	.
fort	4	.	.	.	.	.	.	.	.	.
grave	4	14	.	.	.	.	.	.	.	.
haut	7	13	.	.	.	.	.	.	.	.
large	15	6	2	.	.	.	.	.	.	.
long	1	6	2	5	3	.	1	.	.	1
métallique	21	6	2	3	1	1	.	.	1	.
nasillard	3	14	.	7	.	.	1	.	.	.
net	15	1	2	2	.	.	.	.	.	.
pincé	22	7	1	.	2	1	.	.	.	.
résonnant	17	2	.	1	.	.	.	.	.	.
riche	6	5	2	2	.	.	.	.	.	.
rond	8	15	.	3	1	.	.	.	.	.
sec	17	5	2	1	1	.	.	.	1	.
soufflé	21	1	.	.	.	.	.	.	.	.
sourd	11	6	1	3	.	.	.	.	.	.
vibré	12	4	1	2	1	.	.	.	.	.
Total réalisé	245	158	17	52	12	2	4	0	4	1
Total possible					91	16	33	16	46	13
% réalisé					13.19	12.50	12.12	0.00	8.70	7.69

Globalement, très peu de sujets utilisent une échelle corrélée significativement avec deux dimensions perceptives différentes (en moyenne dans seulement 11% des cas où ce serait possible, c'est-à-dire où une échelle est corrélée à deux dimensions différentes simultanément), et ce malgré la fréquence du phénomène de voir une échelle utilisée par plusieurs sujets de manière corrélée avec différentes dimensions perceptives. Ce pourcentage est légèrement supérieur pour des échelles corrélées simultanément aux dimensions 1 et 2. Pourtant quand une échelle est corrélée de manière simultanée à deux dimensions différentes, c'est en général le cas pour un seul des 32 sujets et au maximum pour trois d'entre eux (pour l'échelle liée à *long*, D1&D2). Trois échelles sont associées simultanément à

deux dimensions différentes par deux sujets différents, celles liées à *pincé* pour D1&D2, à *diffus* pour D1&D4 et à *doux* pour D2&D4.

Ceci va à l'encontre de l'hypothèse de corrélats verbaux composites de dimensions perceptives. Les cas où une échelle sémantique serait corrélée simultanément à plusieurs dimensions perceptives seraient dus au fait d'avoir moyenné des données de sujets qui concevaient des sens différents d'une même échelle sémantique. On peut tout de même remarquer que ce phénomène est beaucoup moins important dans cette expérience utilisant le paradigme de la sémantique différentielle que dans l'expérience 1. Une autre hypothèse pouvant expliquer cette différence serait que face au paradigme VAME, les sujets obligés d'utiliser chacune des échelles proposées pour l'ensemble des sons "alignent" leur utilisation d'un mot sur une de leurs principales perceptions (c'est-à-dire sur une des dimensions perceptives...).

### II.2.3. Seuils corrigés selon Bonféroni pour 23 analyses

Nous avons souhaité vérifier si les corrélats verbaux obtenus en analysant les données des sujets un par un étaient suffisamment robustes pour être maintenus si on corrigeait les seuils de significativité des corrélations selon Bonféroni pour les 23 analyses effectuées en parallèle. Les résultats sont rassemblés dans le tableau A5.4. en annexe. Cette opération a pour conséquence de faire chuter le nombre de corrélations significatives de 68%. En effet, le nombre total de corrélations significatives (des données d'un sujet pour une échelle avec une dimension) passe de 472 à 149.

Sur l'ensemble des 23 échelles testées, seule l'échelle liée à *fort* n'est utilisée de manière corrélée à aucune dimension par aucun des 32 sujets interrogés. Pour les 22 autres échelles, au moins un sujet les a utilisées de manière corrélée avec une des 4 dimensions perceptives. Pourtant, en général, si on corrige les seuils de significativité pour 23 analyses d'après Bonféroni, cela représente peu de sujets. En effet, une seule échelle est utilisée de manière corrélée à une dimension perceptive par 50% des sujets, celle liée à *attaque rapide*, utilisée de manière corrélée à la dimension 1. Trois autres échelles, celles liées à *pincé*, *soufflé* et *rond* sont utilisées de manière corrélée à une dimension perceptive par au moins un tiers des sujets (respectivement par 44, 41 et 38% des sujets). On trouve aussi 5 échelles corrélées à une dimension perceptive par 20 à 28% des sujets, celles liées à *sec*, *large*, *vibré*, *résonnant* et *haut*.

On observe à nouveau que les corrélations sont beaucoup plus nombreuses avec la dimension 1 qu'avec la dimension 2, elles mêmes beaucoup plus nombreuses qu'avec la dimension 4. Un seul sujet utilise une échelle (*long*) de manière corrélée avec la dimension 3. En effet, seules 4 échelles ne sont corrélées à la dimension 1 par aucun sujet. De plus, les plus forts pourcentages de sujets qui utilisent une échelle de manière corrélée avec une dimension perceptive le sont toujours avec la dimension 1. C'est le cas de *soufflé*, *attaque rapide* et *pincé* (respectivement 41, 34 et 31% des sujets à  $p < .01$ ). Pour la dimension 2, l'échelle qui présente le maximum de sujets qui lui est corrélée est *rond* (avec 25% des sujets à  $p < .05$ ). On peut aussi noter que l'échelle liée à *haut* est corrélée avec cette

deuxième dimension pour 19% des sujets dont 16% des sujets avec  $r > .83$  et  $p < .01$ . Les échelles corrélées aux dimensions 3 ou 4 ne le sont que par un seul sujet.

Sur les 23 échelles, 10 sont corrélées à une seule dimension perceptive. 7 échelles sont corrélées uniquement à la dimension 1, c'est le cas de *attaque rapide*, *diffus*, *pincé*, *résonnant*, *sec*, *soufflé* et *sourd* et 3 échelles sont corrélées uniquement à la dimension 2 : *bas*, *grave* et *nasillard*. Quand un sujet utilise une échelle de manière corrélée à une des deux dimensions 3 ou 4, il y a toujours d'autres sujets qui corrént ces mêmes échelles avec soit la dimension 1 [c'est le cas de *long* (D1 et D3) et de *net* (D1 et D4)], soit avec les dimensions 1 et 2 (c'est le cas de *doux* et de *vibré* corrélés à D1, D2 et D4).

A propos des échelles corrélées à plusieurs dimensions perceptives, il est extrêmement intéressant de remarquer que même s'il y a 12 échelles qui ont des sujets qui les corrént à plusieurs dimensions perceptives différentes, ce sont à chaque fois le fait de sujets différents. Ainsi, la notion d'échelle composite de plusieurs dimensions perceptives apparaît lorsqu'on moyenne les données de plusieurs sujets différents mais n'existe pour aucun sujet individuel.

#### II.2.4. Résumé

Nous avons résumé les analyses faites sujet par sujets dans le tableau 5.10.

Tableau 5.10. Corrélat des dimensions perceptives. Données correspondant aux échelles utilisées de manières corrélées par le plus grand nombre de sujets (% de sujets).

D1.4	D2.4	D3.4	D4.4
<b>attaque rapide</b> (75%)	<b>rond</b> (47%)		doux (38%)
<b>pincé</b> (69%)	grave (44%)		(nasillard 22%)
<b>soufflé</b> (66%)	aigu (44%)	∅	(diffus 16%)
métallique (66%)	nasillard (44%)		
résonnant (53%)			
sec (53%)			

En gras, les corrélat maintenus après correction selon Bonféroni pour 23 analyses, entre parenthèses, données obtenues pour moins d'un tiers des sujets (38%).

Alors qu'il y a 22 échelles (soit 96% d'entre elles) qui sont concernées par au moins 2 sujets qui les utilisent de manière corrélée à deux dimensions différentes, seulement pour 11 d'entre elles (soit 48% d'entre elles), il y a une partie (en générale faible : 1 à 3 sujets au maximum) des sujets qui utilisent la même échelle de manière conjointement corrélée à deux dimensions différentes. Pour les 11 autres échelles concernées, c'est le fait de sujets différents qui utilisent la même échelle mais de manière différente.

### II.3. Régressions multiples incrémentales

#### II.3.1. Régressions



Si les seuils de significativité sont corrigés selon Bonféroni pour 23 analyses, il n'apparaît plus d'échelles sémantiques *a priori* composites de plusieurs dimensions perceptives à partir des données moyennées sur l'ensemble des sujets. De plus, on observe que si on considère les données des sujets pris séparément, très peu de jugements effectués sur une échelle sémantique sont simultanément corrélés à plusieurs dimensions perceptives. Pourtant, la plupart des échelles ont des coefficients de corrélation supérieurs à .6 sur plusieurs dimensions en même temps. Nous avons donc tout de même voulu faire les analyses de régression multiple incrémentale, ne serait ce que pour pouvoir en comparer les résultats avec ceux obtenus lors des analyses de notre première expérience. Les résultats effectués sur les moyennes classiques sont rassemblés dans le tableau 5.11.

Tableau 5.11. Coefficients de régression incrémentale entre l'utilisation des échelles sémantiques et les dimensions perceptives (F pour enter = 4).  $\bar{Y} = c_0 \cdot 1 + \sum_{i=1}^4 c_i \cdot \overline{Dim}_i$  et variance expliquée (R<sup>2</sup>).

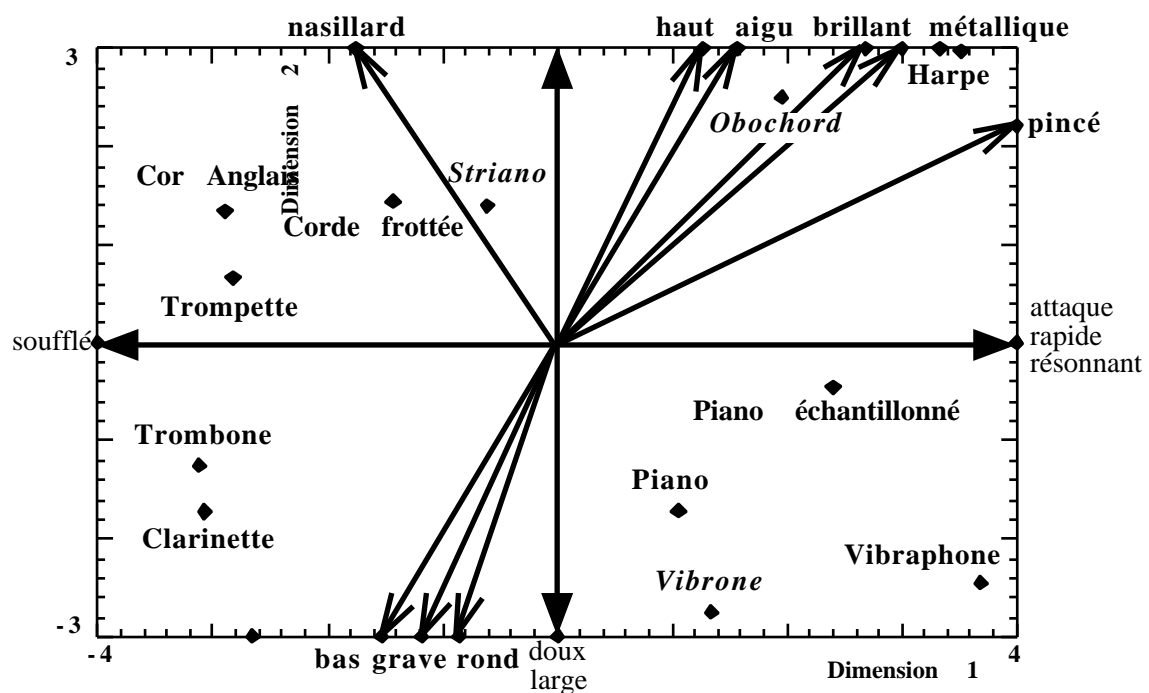
	R <sup>2</sup>	C0	C1	C2	C3	C4	Nb
métallique	0.97 *	0.466	0.041	0.041		0.028	3
rond	0.93 *	0.402	-0.014	-0.049		-0.031	3
grave	0.92 *	0.403	-0.019	-0.048		-0.017	3
nasillard	0.88 *	0.472	-0.024	0.041		0.053	3
large	0.86 †	0.451		-0.023	-0.020	-0.027	3
pincé	0.90 *	0.416	0.074	0.041			2
attaque	0.84 *	0.509	0.066			0.037	2
bas	0.82 †	0.436	-0.024	-0.047			2
aigu	0.80 †	0.492	0.025	0.048			2
brillant	0.79 †	0.470	0.035	0.039			2
haut	0.77 †	0.493	0.019	0.045			2
doux	0.75 †	0.445		-0.051		-0.069	2
sourd	0.74	0.443	-0.039	-0.044			2
sec	0.63	0.431	0.040	0.036			2
riche	0.54	0.453	0.012		-0.022		2
soufflé	0.93 *	0.332	-0.056				1
résonnant	0.66 †	0.465	0.041				1
clair	0.58	0.458	0.043				1
vibré	0.51	0.438	0.030				1
net	0.45	0.488	0.041				1
diffus	0.00	0.447					0
fort	0.00	0.460					0
long	0.00	0.493					0

Avec \*: p<.01, †: p<.05, corrigés selon Bonféroni pour 23 analyses

Douze des 23 échelles testées semblent composites de plusieurs dimensions perceptives. En effet, 5 échelles sémantiques semblent composites des trois dimensions perceptives dont 4 pour les dimensions D1, D2 et D4 (échelles liées à *métallique*, *rond*, *grave* et *nasillard*) et une échelle composite des dimensions D2, D3 et D4 (échelle liée à *large*). Ceci est significatif même après correction selon Bonféroni pour 23 analyses. De plus 7 autres échelles semblent composites de deux dimensions perceptives : dont 5 des dimensions D1 et D2 (échelles liées à *pincé*, *bas*, *aigu*, *brillant*, et *haut*) et une des dimensions D1 et D4 (échelle liée à *attaquerapide*). De plus l'échelle liée à *doux* est composite des dimensions D2 et D4.

De même que lors des analyses effectuées pour l'expérience 1, nous avons essayé de représenter le sens d'un attribut verbal composite de plusieurs dimensions perceptives en représentant la direction suivant laquelle varie son utilisation dans chacun des plans perceptifs (Figures 5.7 à 5.11 ci-dessous). Sur ces figures, on représente par une flèche la projection sur le plan des deux dimensions concernées de la direction de variation des attributs verbaux significativement corrélés à ces deux dimensions. La pente de la droite étant égale au rapport des coefficients de régression multiple avec ces deux dimensions.

Figure 5.7. Plan correspondant aux dimensions 1 et 2.

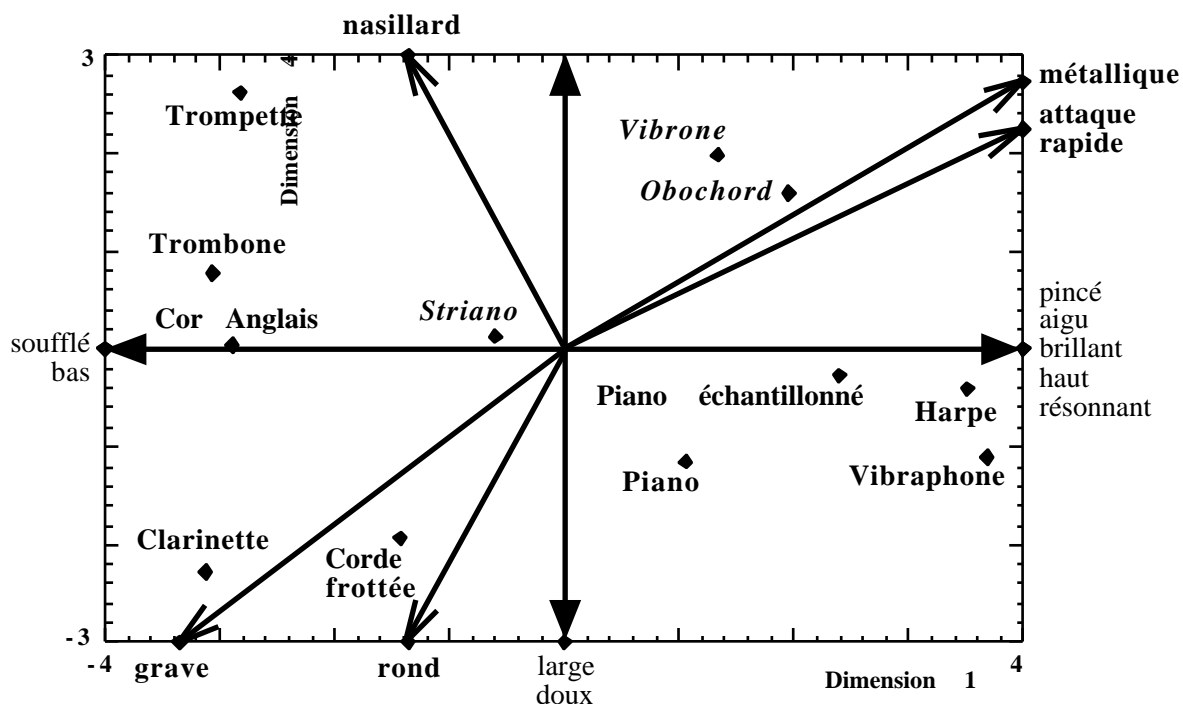


Globalement, on remarque que les échelles liées à *bas*, *grave* et *rond* pointent dans la même direction qui est très proche de la dimension D2, de manière opposée aux dimensions liées à *haut* et *aigu*. Plusieurs échelles pointent dans le même quadrant que celles liées à *haut* et *aigu* : on trouve ainsi les échelles liées à *brillant* puis *métallique* dont le caractère composite des dimensions D1 et D2 est beaucoup plus marqué. L'échelle liée à *métallique*, égalisant les influences de ces deux dimensions perceptives peut être considérée comme un concept associé à une variation vers un temps d'attaque court et des composantes harmoniques aiguës plus marquées.

Toujours dans ce même quadrant, on trouve l'échelle liée à *pincé* qui semble cette fois-ci plus influencée par la dimension 1, bien qu'elle n'y soit pas corrélée de manière exclusive, comme si ce concept associait les notions de mode d'attaque pincée et de son évoquant "nez pincé", pour suivre une remarque d'un sujet (a8m).

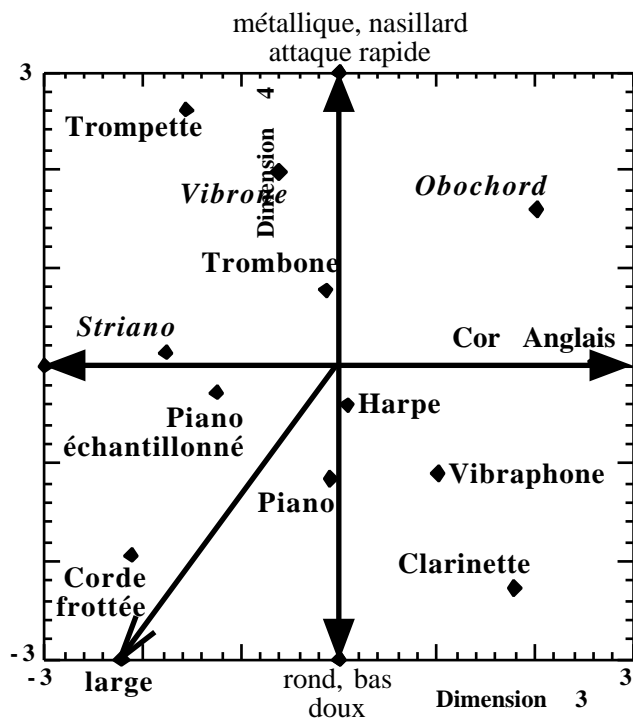
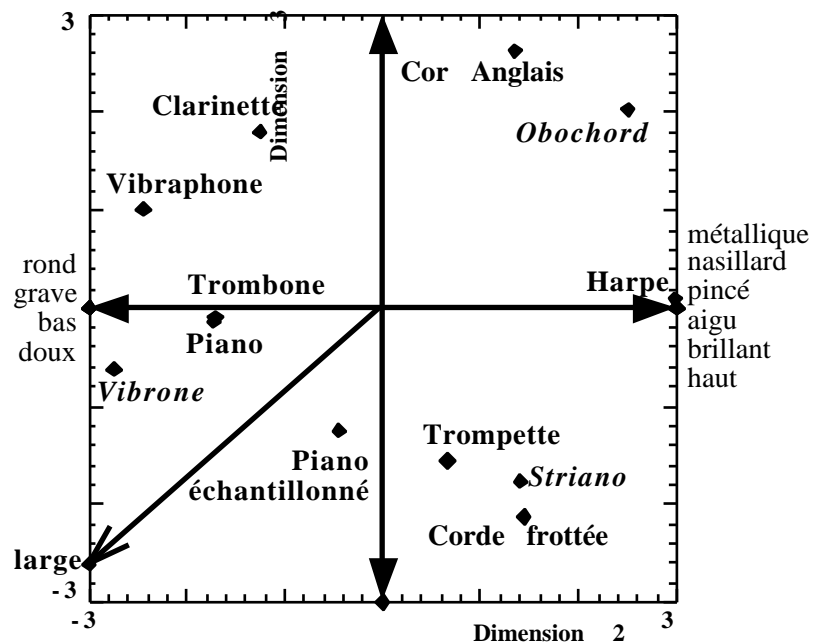
Enfin, à angle droit de cette dernière échelle, on trouve l'échelle liée à *nasillard* correspondant donc plutôt à une attaque lente, associée à la présence d'harmoniques aigus. Il est intéressant de remarquer, comme nous l'avons déjà fait dans la partie relative à l'expérience 1, que les notions de *métallique* et de *nasillard* apparaissent dans ce plan comme orthogonales, alors qu'elles sont souvent confondues dans la littérature anglophone (Kendall & Carterette, 1996).

Figure 5.8. Plan correspondant aux dimensions 1 et 4.



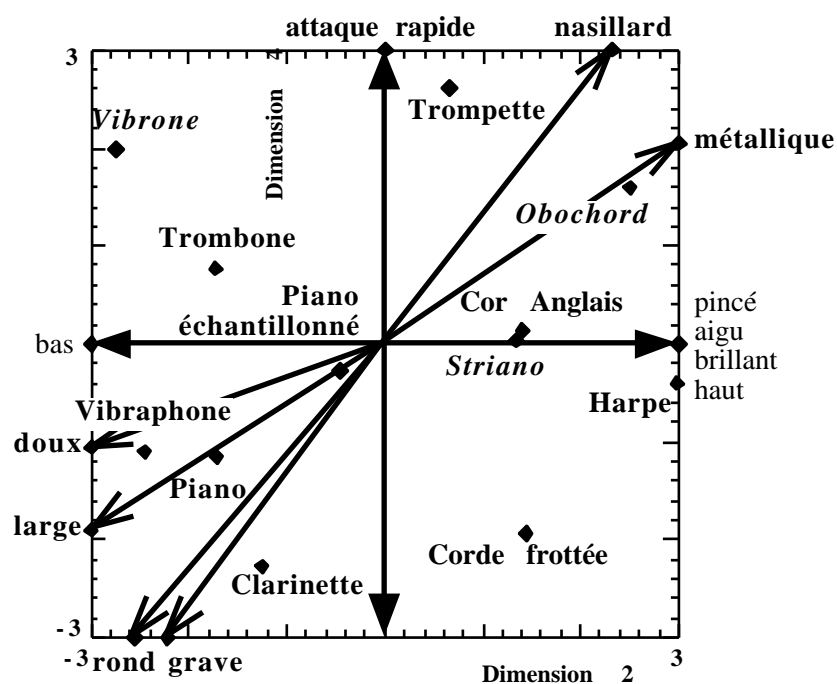
Sur la figure 5.8. les échelles liées à *métallique* et *grave* se présentent comme pointant dans des directions opposées. L'échelle liée à *métallique* pointant entre les dimensions D1 et D4 et proche de l'échelle liée à *attaque rapide*. Le fait que cette dernière apparaisse comme composite des dimensions D1 et D4 pose un problème, dans la mesure où cette échelle est censée représenter non seulement le concept habituellement associé à la première dimension, mais aussi une sorte de label de son corrélant acoustique (le logarithme de l'enveloppe temporelle étant très exactement lié à la rapidité de l'attaque du son). La quatrième dimension ne possédant pas de corrélant verbal non composite, elle est néanmoins liée à une notion temporelle (cf. chapitre 4 §I.5). Serait-il possible que lorsque les gens jugent la vitesse de l'attaque (rapide / lente) ils ne jugent pas directement une qualité perceptive liée à D1 mais un mélange avec une qualité perceptive liée à D4 ? Ou cette double corrélation est-elle le fait de la légère corrélation de D4 au LTM\* ? On retrouve aussi dans ce plan l'orthogonalité des échelles liées à *nasillard* et *métallique*, ainsi que la proximité des échelles liées à *rond* et *grave*, bien que ces deux dernières notions soient moins proches dans ce plan qu'elles ne l'étaient dans le plan des dimensions D1 et D2.

Figures 5.9 et 5.10. Plan correspondant respectivement aux dimensions D2 et D3 (haut), et D3 et D4 (bas).



Ces deux figures sont les seules où la dimension 3 apparaît, et elles sont toutes les deux dues à l'utilisation de l'échelle liée à *large* de manière corrélée à la fois aux dimensions 2, 3 et 4. Ainsi, nous n'avons pas de figure correspondant au plan des dimensions 1 et 3 car aucune échelle ne leur est significativement corrélée. Cela rejoint le questionnement de la validité perceptive de cette troisième dimension.

Figure 5.11. : plan correspondant aux dimensions 2 et 4



On trouve par contre beaucoup d'échelles sémantiques dans le plan des dimensions D2 et D4. Globalement, seule la première diagonale est occupée, comme si ces deux dimensions n'étaient pas conceptualisées séparément, mais correspondaient à un seul type de concept associé à une rotation de 45° de l'espace perceptif. En effet, alors qu'on trouve certaines échelles uniquement corrélées à la première dimension, on n'en trouve aucune corrélée uniquement à D2 ou à D4. Dans ce plan, les échelles liées à *nasillard* et *métallique* se retrouvent dans le même quadrant, bien que *métallique* pointe plus vers la dimension 2 et *nasillard* plus vers la dimension 4. Ces deux échelles se retrouvent opposées à celles liées à *rond* et *grave*, pour *nasillard*, et à celles liées à *doux* et *large* pour *métallique*. De même que les figures précédentes ont pu montrer que les échelles liées à *nasillard* et *métallique* correspondaient à des utilisations différenciées, cette figure nous permet d'observer (bien que dans une moindre mesure) que l'échelle liée à *doux* est utilisée de manière différenciée de celle liée à *rond* : le son de vibraphone est plus *doux*, celui de la clarinette plus *rond*.

### II.3.2. Différences musiciens - non-musiciens

Nous en arrivons à comparer ces mêmes analyses effectuées sur des groupes de sujets de compétences musicales différentes. Les résultats sont rassemblés dans le tableau A5.5. en annexe.

Pour 8 des 23 échelles, les corrélations entre les jugements faits sur chacune des échelles sémantiques et les différentes dimensions perceptives ne sont pas modifiées quand on considère les trois groupes de sujets différents : musiciens, non-musiciens et l'ensemble des sujets. C'est le cas des échelles utilisées de manière corrélée à aucune des dimensions perceptives, celles liées à *diffus*, *fort*, et *long*, mais aussi à certaines échelles utilisées de manière corrélée à une dimension mais de manière non significative, c'est le cas par exemple des échelles liées à *clair* et *vibré*, ainsi que les échelles liées à *métallique*, *pincé* et *soufflé*, qui restent très significativement corrélées quel que soit le groupe de

sujets considéré, respectivement à D1, D2 et D4 pour *métallique*, à D1 et D2 pour *pincé*, et à D1 pour *soufflé*.

Ensuite, pour 5 des échelles, les dimensions auxquelles elles sont liées ne sont pas modifiées, seul le niveau de significativité change avec le groupe de sujets considéré. C'est le cas des échelles liées à *aigu*, *brillant*, *doux*, *nasillard* et *sourd*, dont l'utilisation de manière corrélée respectivement aux dimensions D1 et D2 (pour *aigu*, *brillant* et *sourd*), D2 et D4 (pour *doux*), D1, D2 et D4 (pour *nasillard*) semble surtout le fait des musiciens. Aucune autre situation n'est représentée : par exemple un comportement de l'ensemble des sujets surtout dû à celui des non-musiciens, ou bien un comportement sur un des deux sous-groupes de sujets et qu'on ne retrouve pas au niveau de l'ensemble des sujets.

Enfin, viennent des cas où le nombre ou la nature des dimensions auxquelles une des échelles est corrélée varie avec le groupe de sujets considéré. C'est le cas de 3 échelles pour lesquelles ces corrélations ne sont significatives dans aucun des cas (par exemple pour les échelles liées à *net*, *riche* et *sec*) mais aussi pour 7 autres échelles (soit en tout 10 sur 23). Pour 5 de ces échelles, il n'y a pas de modification de la significativité en plus de la modification du nombre de dimensions auxquelles l'échelle est liée. C'est le cas notamment de l'échelle liée à *attaquerapide* que l'ensemble des sujets et le groupe des non-musiciens utilisent de manière corrélée à la fois aux dimensions D1 et D4, alors que les musiciens ne l'utilisent que de manière corrélée à la première dimension. La dimension 4 semblant correspondre elle aussi à un caractère temporel, peut-être que les musiciens sont mieux capables de choisir de juger un seul parmi l'ensemble des paramètres temporels globalement liés (comme par exemple le temps d'attaque mais aussi la durée subjective du plateau ou le temps de décroissance...). C'est le cas inverse pour l'échelle liée à *rond*. En effet, l'ensemble des sujets ainsi que les sujets musiciens l'utilisent de manière corrélée aux dimensions D1, D2 et D4, alors que les non-musiciens se contentent de l'utiliser de manière corrélée aux dimensions D2 et D4. C'est aussi le cas des échelles dont les corrélations avec les dimensions ne sont jamais significatives. Ainsi pour les échelles liées à *net* et *sec*, le comportement des musiciens diffère de celui des non-musiciens qui influence celui de l'ensemble des sujets (échelle corrélée à D1 et D2 pour l'ensemble des sujets et les non-musiciens, uniquement à D1 pour les musiciens pour l'échelle liée à *sec*, l'inverse pour l'échelle liée à *net*). C'est enfin le cas de l'échelle liée à *riche* : les sujets musiciens l'utilisent de manière corrélée à la troisième dimension, les non-musiciens à la première dimension, les résultats de l'ensemble des sujets étant alors corrélés aux dimensions D1 et D3 (tous ces résultats sont non significatifs après correction pour 23 analyses).

Dans les cas où les niveaux de significativité varient en même temps que les dimensions auxquelles les échelles sémantiques sont liées, on retrouve avec *large* un cas proche de celui de l'échelle *riche* : les musiciens l'utilisent de manière corrélée à la dimension D2, les non-musiciens à D4 et l'ensemble des résultats se retrouvent alors corrélés aux trois dimensions D2, D3 et D4. Il est intéressant de remarquer que les données moyennées sur l'ensemble des sujets mènent à une corrélationsignificative

alors que ce n'est pas le cas des données de chacun des sous-groupes de sujets. On retrouve un cas similaire à ceux de *net* et *sec* avec l'échelle liée à *résonnant* : l'ensemble des sujets ainsi que le groupe des non-musiciens utilisent cette échelle de manière corrélée à la dimension 1, les musiciens l'utilisant de manière conjointement corrélée aux dimensions D1 et D2.

Les échelles liées à *grave* et *haut* correspondent au même type de mécanisme : l'ensemble des sujets suit le comportement des musiciens alors que le comportement des non-musiciens correspondrait à une dimension de moins. Ainsi par exemple, il est intéressant de remarquer que l'échelle liée à *grave* n'est corrélée aux dimensions D1, D2 et D4 que du fait des jugements des musiciens, les non-musiciens ne l'utilisant que de manière corrélée aux dimensions D1 et D2 (non significatif). Pour l'échelle liée à *haut*, le raisonnement est le même : si l'ensemble des sujets utilisent cette échelle de manière corrélée aux dimensions D1 et D2, c'est le fait des musiciens, les non-musiciens l'auraient utilisée uniquement de manière corrélée à D2 (non significatif).

Pour finir, on peut considérer l'échelle liée à *bas* : l'ensemble des sujets l'utilisent de manière corrélée aux dimensions D1 et D2 (ainsi que les non-musiciens bien que de manière non significative) alors que les musiciens ajouteraient à ces deux dimensions la dimension D4 (cette corrélation étant la plus significative...). On ne peut pas tellement jouer sur les figures... quoique quelques morceaux choisis pourraient être intéressants...

#### II.4. Comparaison avec l'utilisation des échelles

Le paragraphe suivant fait appel aux analyses effectuées d'une part dans le chapitre 7 §I, et d'autre part dans le chapitre 4 §II.6.

##### **II.4.1. Comparaison avec la notion de pertinence**

Parmi les échelles dont l'utilisation est corrélée avec la dimension 1, on en retrouve certaines qui sont considérées comme pertinentes par le plus grand nombre de sujets (*attaque rapide*, *métallique*, globalement pertinentes pour respectivement 91% et 84% des sujets et pertinentes pour plus de 50% des sujets pour respectivement l'ensemble des 12 sons et 9 d'entre eux). Mais on en retrouve aussi certaines considérées comme pertinentes par un nombre de sujets beaucoup plus faible (par exemple l'échelle liée à *soufflé*, considérée comme adaptée par plus de 50% des sujets pour seulement 3 des 12 sons et globalement pertinente pour seulement 48% des sujets). Les autres échelles dont l'utilisation est corrélée à cette dimension peuvent se répartir dans ces deux groupes. Les échelles liées à *clair* et *résonnant* sont globalement considérées comme pertinentes par respectivement 71% et 79% des sujets, et considérées comme adaptées par plus de 50% des sujets pour 9 des 12 sons. Ce sont donc des échelles globalement très pertinentes. L'échelle liée à *pincé* peut s'apparenter à celle liée à *soufflé* dont elle est une sorte de complémentaire dans la façon dont les sujets ont catégorisé les sons à partir de leur temps d'attaque. En effet, cette échelle est considérée comme pertinente par plus de 50% de sujets pour seulement 5 des 12 sons, qui sont ceux jugés les plus *pincés*. Elle est par contre considérée comme globalement pertinente par 68% des sujets.

Par contre 4 des 5 échelles dont l'utilisation est corrélée à la deuxième dimension perceptive (celles liées à *aigu*, *grave*, *haut* et *bas*) font partie des échelles considérées comme globalement pertinentes par le moins de sujets (de 26 à 40% des sujets) et adaptées pour plus de 50% des sujets pour seulement 1 ou 2 des 12 sons. L'échelle liée à *rond* étant au contraire considérée comme globalement pertinente par 69% des sujets (dont 88% des sujets musiciens) et comme adaptée par plus de 50% des sujets pour 8 des 12 sons. De tous les corrélats verbaux de cette deuxième dimension, c'est donc de loin l'échelle considérée comme la plus pertinente par l'ensemble des sujets.

Il est aussi intéressant de considérer ici les autres échelles les plus pertinentes pour l'ensemble des sujets et qu'on ne trouve pas significativement corrélées à une des dimensions perceptives. Il s'agit des échelles liées à *doux* et *sourd*. Ces deux échelles ont des coefficients de corrélation égaux (dont la valeur varie de  $-.55$  à  $-.66$ ) par rapport à deux dimensions perceptives différentes. Les dimensions 2 et 4 pour *doux*, 1 et 2 pour *sourd*. L'échelle liée à *doux* étant celle dont le coefficient de corrélation avec la dimension 4 est le plus élevé ( $r=-.65$ ).

Hypothèse : ces échelles correspondent à des concepts qui ne sont pas reflétés par les dimensions perceptives (ou ils sont composites de plusieurs dimensions ?)

#### II.4.2. Comparaison avec les corrélations inter-sujets

De même, on peut remarquer que parmi les 7 échelles qui présentent la plus forte corrélation inter-sujets (chapitre 4 §II.6), on retrouve justement celles liées à *attaque*, *pincé* et *soufflé* (ayant respectivement 37, 26 et 17% des corrélations inter-sujets significatives à  $p<.1$ ) qui sont les échelles les plus fortement corrélées avec la dimension 1, quel que soit le mode de calcul des moyennes et sujet par sujet. C'est le cas aussi, dans une moindre mesure, des échelles liées à *sec* et *résonnant* (ayant respectivement 15 et 14% des corrélations inter-sujets significatives à  $p<.1$ ). En effet, si seule l'échelle liée à *résonnant*, apparaît comme corrélée à la dimension 1 lors des calculs sur les moyennes (particulièrement chez les non-musiciens ou sur les jugements normés), l'échelle liée à *sec* est significativement ( $p<.05$ ) corrélée à la dimension 1 par 9 sujets et aucun sujet ne l'utilise de manière corrélée à une autre dimension. On retrouverait donc aussi au niveau des corrélations inter-sujets, la force attractive de la dimension 1 liée au mode d'attaque des sons lors des jugements sémantiques, qui semblent d'autant plus cohérents entre eux qu'ils caractérisent mieux cette dimension perceptive.

Les cas des échelles liées à *sourd* et *net* sont plus difficiles à interpréter. En effet, ni *sourd* ni *net* n'apparaissent comme corrélats d'une dimension perceptive lors des calculs de moyennes, et lors des calculs sujet par sujet. Seuls 3 sujets utilisent ces échelles de manière corrélée à la dimension 1 (pour *net*, on a en plus un sujet qui l'utilise de manière corrélée à la dimension 4). Par contre, les échelles liées à *grave* et *rond* font partie de celles qui font le moins consensus entre les sujets (7% des corrélations inter-sujets sont significatives à  $p<.1$ ).

### III. Comparaison entre les résultats des expériences 1 et 2



Nous avons souhaité comparer ces résultats à ceux obtenus dans l'expérience 1. Il faut pour cela bien avoir en tête la différence fondamentale de nature de ces données :

- dans l'expérience 1 on a calculé, pour chaque échelle et chaque paire de sons, le pourcentage de sujets qui ont verbalisé spontanément que "le premier son de la paire était plus XXX que le deuxième" (ce qui aboutit par exemple à "le son TPT est *plus aigu* que celui TBN" pour 75% des locuteurs spontanés de cette comparaison, les 25% restant ayant verbalisé quelque chose du type : " le son TPT est *moins aigu* que le son TBN). Il est important de souligner qu'on n'a jamais tous les sujets qui s'expriment avec chaque mot sur toutes les paires de sons.
- dans l'expérience 2, on calcule la position moyenne (pour l'ensemble des sujets) sur chacune des échelles sémantiques de chacun des 12 sons. On peut ainsi en déduire leurs positions relatives 2 à 2 et comparer ces positions relatives, soit aux verbalisations spontanées - les deux sons sont-ils positionnés de la même façon ou de façon opposée vis-à-vis du même mot dans les deux contextes - soit aux positions relatives des sons sur les dimensions de l'espace perceptif.

### III.1. Comparaison des verbalisations spontanées et des positionnements sur les échelles

Nous avons comparé pour les deux expériences la valeur pour chacune des 66 paires de sons, d'une part du pourcentage de sujets ayant considéré que le premier son de la paire était plus "XXX" que le deuxième (expérience 1) et d'autre part la distance orientée entre les deux sons de la paire par rapport à l'échelle sémantique considérée (expérience 2). Nous avons rassemblé dans le tableau 5.12 les coefficients de corrélation entre ces deux valeurs pour chaque échelle et chaque paire de sons.

En ce qui concerne les comparaisons des valeurs correspondant à chaque paire de sons sur chaque échelle sémantique, on constate que le nombre de degrés de liberté varie en fonction de l'échelle. Ceci reflète le fait que pour l'expérience 1, nous n'avions pas de verbalisations libres systématiques qu'"un son était plus XXX que l'autre" pour chaque mot considéré et chacune des 66 paires de sons. Le nombre de degrés de liberté correspond donc au nombre de paires (sur les 66) pour lesquelles au moins un sujet s'était exprimé spontanément.

Tableau 5.12. Corrélations entre les données des expériences 1 et 2

mot / échelle	r	p	dl	mot / échelle	r	p	dl
aigu	0.73 *		55	métallique	0.75 *		31
attaque	<b>0.87</b> *		45	nasillard	0.69 *		25
bas	0.55 †		12	net	0.60 *		48
brillant	0.72 *		12	pincé	<b>0.33</b>		16
clair	0.75 *		36	résonnant	0.70 *		28
diffus	0.69 *		11	riche	<b>0.31</b>		22
doux	0.60 *		22	rond	0.79 *		37
fort	<b>0.38</b> †		31	sec	0.57 *		28
grave	0.77 *		48	soufflé	<b>0.87</b> *		10
haut	<b>0.88</b> *		8	sourd	0.78 *		39
large	<b>0.22</b>		20	vibré	<b>0.36</b>		27
long	0.71 *		36				
moyenne	0.64						
écart type	0.19						

avec \*:  $p < .01$  ; †:  $p < .05$  ; en gras, les valeurs s'écartant de la moyenne de plus d'un écart type.

On remarque que globalement le coefficient de corrélation moyen est de .64 et que pour la plupart des échelles sémantiques (19 sur 23) ces données sont significativement corrélées. On retrouve dans les

positions moyennes de chaque son correspondant à chacune de ces échelles sémantiques les positions relatives correspondant aux comparaisons spontanées qu'avaient effectuées les sujets lors des verbalisations libres de l'expérience 1. Ceci rejoint un résultat obtenu par Parizet et Nosulenko (1999) sur des sons de voiture.

C'est d'autant plus vrai pour les 3 échelles dont les coefficients de corrélation sont significativement supérieurs à ceux de la moyenne, c'est-à-dire les échelles liées à *haut*, *attaque rapide* et *soufflé*. Il est étonnant de retrouver parmi celles-ci l'échelle liée à *haut*, qui faisait partie de celles les moins verbalisées spontanément lors de l'expérience 1 (40 % des sujets avec un nombre d'occurrences moyen de 4.25 par sujet) et qui fait partie des échelles considérées comme les moins pertinentes par les sujets de l'expérience 2 (26% seulement la considèrent comme globalement pertinente). C'est un peu le même cas pour l'échelle liée à *soufflé* (globalement pertinente pour 48% des sujets). Ces échelles ont donc été utilisées de manière très similaire lors des verbalisations libres et lors de l'utilisation des échelles sémantiques.

Ce n'est pas le cas pour 4 échelles sémantiques, celles liées à *large*, *pincé*, *riche* et *vibré*. On peut remarquer que ces échelles font partie de celles dont le degré de liberté dans la comparaison est dans les plus faibles, c'est à dire aux comparaisons effectuées le moins souvent spontanément lors de l'expérience 1. Pourtant d'autres échelles ont de très bons coefficients de corrélation malgré un nombre de paires comparées spontanément aussi faible. C'est aussi le cas, dans une moindre mesure pour les échelles liées à *bas* et *fort*, dont les coefficients de corrélation sont significativement inférieurs à la moyenne.

Doit-on supposer que ces mots ont été utilisés dans des sens différents lors d'une utilisation spontanée pour comparer les sons de chaque paire dans l'expérience 1 et lors du positionnement de chacun des 12 sons sur l'échelle sémantique associée dans l'expérience 2 ?

### III.2. Profils sémantiques de chaque dimension perceptive

Nous allons comparer ici la façon dont, dans les 2 expériences, les données verbales sont corrélées à chacune des 4 dimensions de l'espace perceptif. Pour cela on définit un "profil sémantique" d'une dimension perceptive comme l'ensemble des coefficients de corrélation qu'elle entretient avec les 23 mots testés. Si on compare, pour chacune des 4 dimensions perceptives, les profils obtenus avec les coefficients de corrélation des 23 échelles sémantiques d'une part (expérience 2) et les profils obtenus avec les z de Kendall de ces mêmes 23 échelles d'autre part (expérience 1), on obtient le tableau 5.13 suivant.

Tableau 5.13. Corrélation des profils sémantiques des dimensions perceptives réalisés par les z de Kendall dans la corrélation de rangs pour l'expérience 1 et par les coefficients de corrélation de chacune des échelles sémantiques pour l'expérience 2. (dl=21)

---

Expérience 2

		D1.4	D2.4	D3.4	D4.4
Expérience 1	D1.4	0.82 *	0.61 *	0.56 *	0.36
	D2.4	0.52 †	0.90 *	0.30	0.68 *
	D3.4	-0.08	0.05	0.67 *	0.00
	D4.4	0.35	0.73 *	0.29	0.81 *

avec \*:  $p < .01$  ; †:  $p < .05$

On remarque que la distribution des corrélations de chacune des dimensions perceptives sur les 23 échelles sémantiques est corrélée de manière significative dans les deux expériences (toutes les valeurs sur la diagonale sont significatives (avec  $r(21) > .67$  et  $p < .01$ ). On obtient de plus des corrélations significatives entre les profils des dimensions 2 et 4 à travers les deux expériences (D2 de l'expérience 1 étant corrélée à D4 de l'expérience 2 et réciproquement). C'est aussi le cas des dimensions 2 et 1. Ceci peut s'expliquer par le fait que dans le cadre de chacune des expériences, les profils obtenus pour la dimension 2 sont à chaque fois corrélés à ceux des dimensions 1 et 4. Ceci est peut être relié au fait que dans le cadre de l'expérience 1, la plupart des corrélats verbaux étaient composites des dimensions 1 et 2 et donc conjointement corrélés à ces deux dimensions. On observe aussi une corrélation significative entre les profils des dimensions 1 de l'expérience 1 et celui de la dimension 3 de l'expérience 2.

### III.3. Profils perceptifs de chaque échelle sémantique

De la même façon, nous avons défini un "profil perceptif" comme l'ensemble des coefficients de corrélation avec chacune des 4 dimensions perceptives, d'une part de chaque "mot" utilisé spontanément et d'autre part de chacune des échelles sémantiques. En effet, lors de l'expérience 1, nous avons calculé des  $z$  de Kendall entre l'utilisation de chaque mot dans la comparaison de paires de sons et chaque dimension perceptive. On obtient ainsi un profil de corrélation de chacun des mots utilisés pour comparer les deux sons de chaque paire sur les 4 dimensions perceptives qu'on peut comparer au profil perceptif de chaque échelle sémantique correspondante de l'expérience 2. Nous avons comparé dans le tableau 5.14 l'utilisation des échelles lors des deux expériences dans la manière dont les données étaient corrélées à nos 4 dimensions perceptives.

On remarque que, bien que le nombre de degrés de liberté soit très faible ( $df=3$ ), le coefficient de corrélation moyen est de .80 et 10 des échelles ont des corrélations significatives entre les deux expériences avec  $r > .94$ . Ce qui est par contre étonnant c'est que l'échelle pour laquelle le profil de corrélation sur les 4 dimensions est le plus proche dans les deux expériences est l'échelle liée à *diffus*. C'était pourtant un des mots les moins employés spontanément (par 15% des sujets seulement) et une échelle globalement pertinente pour 57% des sujets seulement dans l'expérience 2. En fait, seules 7 échelles ont des coefficients de corrélations inférieur à .80.

Tableau 5.14. Corrélations entre les profils perceptifs des mots (Expérience 1) et des échelles sémantiques (Expérience 2) ( $df=3$ )

mot / échelle	r		p	
	mot / échelle	r	mot / échelle	p

aigu	0.96 †	métallique	0.99 †
attaque	0.98 †	nasillard	0.99 †
bas	0.94	net	0.67
brillant	0.82	pincé	<b>0.08</b>
clair	0.95 †	résonnant	0.96 †
diffus	0.99 *	riche	0.65
doux	0.90	rond	0.96 †
fort	0.99 †	sec	0.76
grave	0.92	soufflé	0.98 †
haut	0.87	sourd	0.89
large	<b>-0.71</b>	vibré	<b>0.36</b>
long	<b>-0.09</b>		
moyenne	0.80		
écart type	0.27		

avec \*:  $p < .01$  ; †:  $p < .05$  ; en gras, les valeurs s'écartant de la moyenne de plus d'un écart type.

Deux échelles ont des coefficients de corrélation négatifs : celles liées à *long*, mais cette corrélation est très faible, mais aussi celle liée à *large* avec  $r = -.71$ . Les utilisations spontanées de ce terme pour comparer les sons 2 à 2 sont donc corrélées aux dimensions perceptives de manière opposée à la façon dont est utilisée l'échelle sémantique "*peu large - très large*". En effet, les verbalisations spontanées de *large* dans des comparaisons entre les 2 sons de chaque paire sont corrélées de manière négative aux dimensions 1 et 2 (voir tableau 5.I.2), alors que l'échelle sémantique correspondante est corrélée de manière négative aux dimensions 2, 3 et 4 (voir tableau 5.II.6). On peut noter que c'est un terme qui a été peu utilisé spontanément et par seulement 20% des sujets.

#### III.4. Comparaisons des conclusions des deux expériences

Nous allons comparer les conclusions obtenues à partir des résultats analysés dans les parties I et II de ce chapitre, notamment à partir des 3 tableaux de résumés suivants (5.15 à 5.17). Il s'agit en effet tout d'abord de comparer les corrélats verbaux de chaque dimension perceptive obtenus d'une part à partir des verbalisations libres de comparaison de qualités entre les deux sons de chaque paire, et d'autre part à partir des échelles sémantiques. Puis nous comparerons les résultats obtenus par régression multiple incrémentale dans les deux cas, c'est à dire la notion de corrélat verbal composite de plusieurs dimensions perceptives.

##### III.4.1. Corrélat verbaux des dimensions perceptives

Tout d'abord, on obtient beaucoup plus de corrélats verbaux dans la première expérience que dans la deuxième. C'est-à-dire que si on analyse l'utilisation de termes de qualités produits spontanément pour comparer les sons 2 à 2, et qu'on compare cette utilisation à un positionnement des sons sur des dimensions perceptives, on obtient beaucoup plus de termes dont l'utilisation est corrélée aux dimensions perceptives, que si on analyse le positionnement de ces mêmes sons sur des échelles sémantiques directement issues des productions spontanées.

Tableau 5.15. Ensemble des corrélats verbaux des dimensions perceptives obtenus lors de l'expérience 1.

D1.4	D2.4	D3.4	D4.4
<b>net</b>	<b>pas grave</b>	pas long	<b>pas doux</b>
<b>attaque rapide</b>	<b>pas rond</b>	pas riche	<b>pas rond</b>
<b>clair</b>	<b>aigu</b>		pas agréable
<b>métallique</b>	<b>pas chaud</b>		pas long
<b>pas sourd</b>	<b>pas agréable</b>		pas homogène
<b>percussif</b>	<b>pas sourd</b>		nasillard
<b>pas homogène</b>	<b>nasillard</b>		pas grave
<b>pas long</b>	métallique		
<b>pas large</b>			
<b>résonnant</b>			
<b>pas grave</b>			
sec			
pas soufflé			
aigu			
naturel			

En gras, les attributs avec  $p < .01$ , les autres avec  $p < .05$ , correction selon Bonférioni pour 23 analyses.

Tableau 5.16. Ensemble des corrélats verbaux des dimensions perceptives obtenus lors de l'expérience 2.

D1.4	D2.4	D3.4	D4.4
<b>pas soufflé</b>	<b>pas grave</b>		<i>(pas doux)</i>
<b>pincé</b>	<b>pas rond</b>		<i>(nasillard)</i>
<b>attaque rapide</b>	<i>haut</i>		
résonnant	<i>pas bas</i>		
<i>clair</i>	<i>aigu</i>		
<i>métallique</i>	<i>(pas large)</i>	∅	
<i>vibré</i>	<i>(pas doux)</i>		
<i>(brillant)</i>	<i>(nasillard)</i>		
<i>(net)</i>	<i>(métallique)</i>		
<i>(sec)</i>			
<i>(pas sourd)</i>			

Avec, en gras :  $p < .01$ , sinon  $p < .05$ , corrigé pour 23 analyses ; en italique : sans correction de Bonférioni :  $p < .01$ , entre parenthèses  $p < .05$ .

On peut proposer deux interprétations de ce résultat.

1- Les jugements de dissemblance perceptive ont été recueillis lors de l'expérience 1, c'est à dire conjointement aux verbalisations libres. Il se peut donc que les sujets ayant passé la deuxième expérience n'aient pas jugé de la dissemblance perceptive des sons de la même façon que ceux de l'expérience 1, ce qui aurait abouti à un espace perceptif légèrement différent, dont les dimensions perceptives auraient été mieux corrélées au positionnement sur les échelles sémantiques. Or nous avons comparé les données verbales recueillies dans les 2 expériences aux données perceptives de la première expérience. Il est possible de refuser cet argument : on observe en effet sur ce même ensemble de sons une grande similarité des espaces perceptifs obtenus à partir de groupes de sujets différents (McAdams et al., 1995 et résultats chapitre 4, §I.5.).

Il se peut aussi que d'avoir à faire les deux tâches en même temps, les sujets qui verbalisaient spontanément étaient en même temps concentrés sur les différences et similarités perceptives entre les

sons. Les sujets qui passaient l'expérience 2 n'étaient orientés ni vers une tâche de comparaison des sons, ni vers une tâche d'ordre perceptif puisqu'ils devaient positionner chacun des sons sur chacune des échelles sémantiques. Il se peut donc que les sujets aient plus en tête l'idée d'une représentation perceptive des sons lors de la première expérience par rapport à la deuxième expérience. On peut néanmoins nuancer cette remarque en disant que de positionner 12 sons sur une même échelle, cela implique de les positionner aussi les uns par rapport aux autres.

2- Cela peut être dû à la différence fondamentale de nature des données recueillies dans les deux expériences. Quand un sujet verbalise spontanément une comparaison de qualités entre deux sons, il le fait uniquement pour les sons et les qualités qui lui paraissent saillants perceptivement. Au contraire, quand un sujet doit positionner tous les sons sur une échelle sémantique, il est obligé de faire plus de "compromis" avec ses perceptions...

Si le fait de trouver moins de corrélats verbaux des dimensions perceptives dans l'expérience 2 que dans l'expérience 1 est une observation générale, cela devient d'autant plus frappant en ce qui concerne les dimensions 3 et 4. En effet, alors qu'on trouve dans les verbalisations spontanées *long* et *riche* utilisées de manière corrélée à la **dimension 3**, on n'obtient aucune échelle sémantique corrélée à cette dimension. De même, la **dimension 4** recueille 7 corrélats verbaux parmi les adjectifs verbalisés spontanément pour comparer les sons, alors que seulement deux d'entre eux (*doux* et *nasillard*) se retrouvent faiblement corrélés à cette dimension si on les propose sous forme d'échelle sémantique. Pour ces deux dimensions perceptives, l'argument d'une possible différence d'espace perceptif en fonction du groupe de sujets et de la tâche à effectuer peut tout de même être considéré, du fait que ces deux dimensions sont moins stables entre les diverses expériences (voir chapitre 4 §I.6). La réalité perceptive de la dimension 3 pourrait être éventuellement questionnée.

Par contre, pour ce qui est de la **dimension 1**, on obtient à chaque fois le maximum de termes corrélés à cette dimension : 15 dans l'expérience 1, 4 (voir 11 au maximum, sans prendre en compte les corrections de seuil de significativité pour analyses multiples) dans l'expérience 2, contre au maximum 9 pour les autres dimensions. Ceci peut souligner le caractère saillant de la perception du temps d'attaque des sons de notre ensemble : un grand nombre de termes permettant de qualifier ces différences serait alors produit spontanément pour décrire cette perception. Le fait que ces différences perceptives se retrouvent dans la première dimension de l'espace perceptif modélisé, étant dû au fait que la plupart des différences perçues entre les sons relèvent de cette dimension.

Pour ce qui est des corrélats de la dimension 1 obtenus dans les deux expériences, sur les 15 corrélats obtenus dans les deux expériences et testés dans les deux conditions, 8 se retrouvent corrélats de cette dimension dans les deux conditions. Pourtant, ce n'est pas toujours le cas avec la même importance. Par exemple, alors que **net** était le premier corrélat de la dimension 1 dans l'expérience 1, il n'apparaît qu'en 9ème position (et de manière peu significative) lors de l'expérience 2. Au contraire, *peu soufflé* n'est corrélat de la dimension 1 qu'en 13ème position lors de l'expérience 1 alors qu'il arrive en première position lors de l'expérience 2.

On peut aussi noter l'absence dans l'expérience 2 de *long*, *large*, *grave* et *aigu* comme corrélats de la première dimension, dont les 4 principaux corrélats ne présentent qu'un caractère temporel. Inversement, on peut noter l'absence dans l'expérience 1 de *pincé*, (et dans une moindre mesure de *vibré* et *brillant*) comme corrélat de la dimension 1, ceci étant sans doute dû au fait que ce terme n'avait pas été très souvent verbalisé spontanément.

Pour ce qui est de la **dimension 2**, on obtient à peu près le même nombre de corrélats à partir des deux expériences, si on néglige les corrections pour analyses multiples dans la deuxième expérience. Cette fois-ci, on retrouve les mêmes deux premiers corrélats de cette dimension : *pas grave* et *pas rond*. Les autres corrélats communs aux deux expériences étant *aigu*, *nasillard* et *métallique*. On obtient cette fois-ci plus de différences entre les deux groupes (seulement 5 communs sur un ensemble de 12 au total). En effet, *pas chaud* et *pas sourd* sont corrélats de la dimension 2 uniquement dans le contexte des verbalisations spontanées, *haut*, *pas bas*, *pas large* et *pas doux*, uniquement dans le contexte des échelles sémantiques.

### III.4.2. Corrélats composites de plusieurs dimensions perceptives

On peut comparer aussi les différences entre les deux expériences en terme de corrélats composites de plusieurs dimensions perceptives. En effet, alors que c'est un phénomène dominant dans l'expérience 1, il n'apparaît que beaucoup plus rarement dans l'expérience 2. Les principaux résultats des analyses de régressions multiples incrémentales effectuées pour les deux expériences sont rassemblés dans le tableau 5.17.

Tableau 5.17. Comparaison des corrélats composites de plusieurs dimensions dans les expériences 1 et 2.

	Expérience 1	Expérience 2
D1D2D3	<b>nasillard (-+-)</b>	
D1D2D4	<b>rond (---) grave (---)</b> <b>aigu (+++)</b>	<b>métallique (+++) rond (---)</b> <b>grave (---) nasillard (-++)</b>
D2D3D4		large (---)
D1D3D4	<b>long (---)</b>	
D1D2	<b>sourd (--) agréable (+-)</b> <b>métallique (++) large (--)</b>	<b>pincé (++) bas (--) aigu (++) brillant (++)</b> haut (++) ( <i>sourd (--)</i> )
D1D3	<b>attaque (++)</b> ( <i>brillant (+-)</i> )	( <i>riche (+-)</i> )
D1D4	<b>percussif (++) homogène (--)</b> ( <i>naturel +-)</i> )	<b>attaque rapide (++)</b>
D2D4	<b>chaud (--)</b>	doux (--)
D1	<b>soufflé(-) net clair résonnant (sec)</b>	<b>soufflé (-) résonnant (clair) (vibré) (net)</b>
D2	( <i>haut</i> ) ( <i>bas (-)</i> ) ( <i>pincé</i> )	
D3	( <i>riche (-)</i> ) ( <i>vibré</i> )	
D4	<b>doux(-)</b>	

En gras :  $p < .01$ , normal :  $p < .05$ , corrigé selon Bonféroni pour 23 analyses (en italique : sans correction), "+" et "-" : signes des coefficients de corrélation significatifs avec les différentes dimensions.

Globalement, on retrouve à peu près le même nombre de corrélats de 3 dimensions différentes - 5 dans chacune des 2 expériences - et de corrélats de 2 dimensions différentes - 10 dans l'expérience 1

contre 9 dans l'expérience 2. Alors que certains corrélats ne changent pas entre les deux expériences (c'est par exemple le cas de *rond, grave, sourd, soufflé, net, clair et résonnant*), la majorité d'entre eux subissent des modifications dans leurs corrélations avec les dimensions entre les deux expériences. Par exemple, *nasillard*, qui était corrélé aux dimensions 1, 2 et 3 dans la première expérience devient corrélé aux dimensions 1 2 et 4 dans la seconde. *Attaquerapide* et *brillant* passent d'une double corrélation aux dimensions 1 et 3 dans la première expérience respectivement aux dimensions 1 et 4 et aux dimensions 1 et 2 dans la seconde. Quant à *vibré*, corrélé à la dimension 3 lors de la première expérience, il se retrouve corrélé à la première dans l'expérience 2.

Certains corrélats subissent des "réductions" dans leur nombre de corrélations différentes entre les 2 expériences. C'est par exemple le cas de *aigu*, qui était corrélé aux dimensions 1, 2 et 4 dans la première expérience, et ne l'est plus qu'aux dimensions 1 et 2 dans la seconde. C'est le contraire pour *métallique* et *large*, respectivement corrélés aux dimensions 1, 2 et 4 et 2, 3 et 4 dans la seconde expérience et seulement aux dimensions 1 et 2 dans la première. Il en est de même pour *doux, riche* et *pincé*, respectivement corrélés aux dimensions 2 et 4, 1 et 3, 1 et 2 dans la seconde expérience et seulement respectivement aux dimensions 4, 3 et 2 dans la première expérience. Enfin, on retrouve le fait que par exemple *long, chaud, haut, bas et sec* n'apparaissent comme corrélats que pour une des deux expériences.

#### **IV. Analyse de profils sémantiques des sons obtenus à partir des échelles**

Nous avons souhaité construire, à partir des résultats obtenus dans l'expérience 2 avec les 23 échelles sémantiques, des profils sémantiques de chacun de nos 12 sons, afin de les comparer aux dimensions perceptives du timbre. La construction de profils sémantiques est en effet une utilisation courante des échelles sémantiques, notamment dans le domaine de l'olfaction (Dravnieks & Bock, 1978 ; Dravnieks, 1982 ; Callegari, 1994).

Pour cela, nous avons calculé différents types de distances (euclidiennes ou suivant le modèle de similarité de Tversky (1977)) entre les timbres, afin d'obtenir à chaque fois un espace que nous avons qualifié de "sémantique", vu son obtention à partir d'échelles sémantiques, grâce à une modélisation par analyse multidimensionnelle de proximité. Ces espaces sémantiques ont été comparés à l'espace perceptif dégagé chapitre 4 §I.6.

Chaque timbre T possède donc un profil sémantique constitué des valeurs  $x_{iT}$  comprises entre "peu XXX" et "très XXX" pour chaque sujet et chacune des 23 échelles (i) ainsi que l'avis de chaque sujet sur son adéquation pour décrire le timbre considéré. A partir de ceci, nous avons calculé 5 distances différentes. Pour chacune d'elles, nous effectuons une analyse MDS avec le programme Exscal développé par Winsberg & Carroll (1989a). Ces résultats ont été présentés au 4ème Congrès Français d'Acoustique à Marseille et publiés dans les actes correspondants (Faure & McAdams, 1997).



#### IV.1. Calcul de distances entre les sons à partir des profils sémantiques

Nous avons choisi de calculer deux distances de type euclidien et trois distances dérivées de la notion de similarité proposée par Tversky (1977). Ainsi  $D_1$  est une distance euclidienne simple (eq. 1).  $D_2$  est aussi une distance euclidienne mais calculée avec  $x_{iT} = 0$  pour tous les mots  $i$  considérés comme non pertinents pour décrire le timbre  $T$  pour un sujet donné.

$$d_1(T_1, T_2) = \sqrt{\sum_{i(\text{mots})} (x_{iT_1} - x_{iT_2})^2} \quad (1)$$

$$d_2(T_1, T_2) = -\alpha \cdot f(T_1 - T_2) + \beta \cdot [f(T_1 - T_2) + f(T_2 - T_1)] \quad (2)$$

Pour les distances calculées à partir du modèle de Tversky (eq. 2),  $d_1$  est calculée en considérant que les mots communs aux deux sons ont autant d'importance que les mots différent ( $\alpha = \beta = 1$ ). La fonction  $f$  est le cardinal de l'ensemble considéré et un mot est considéré comme définissant le timbre  $T$  si la valeur de  $x_{iT}$  considérée dépasse un seuil arbitrairement fixé ici à 0.75. Pour la distance  $d_2$ , un mot est considéré comme définissant le timbre  $T$  s'il est considéré comme pertinent pour décrire ce timbre. Enfin  $d_3$  ne diffère de  $d_2$  que par les valeurs de  $\alpha$  et  $\beta$  : on pose que  $x_{iT} = 0$  pour tous les mots non pertinents pour le timbre  $T$ .

$$d_3(T_1, T_2) = -\sum_{i \in T_1 - T_2} (1 - |x_{iT_1} - x_{iT_2}|) + \sum_{i \in T_1 - T_2 \text{ ou } T_1 - T_2} x_{iT_1 \text{ ou } 2} \quad (3)$$

#### IV.2. Analyses multidimensionnelles

La procédure utilisée pour déterminer le modèle le plus approprié aux données, c'est à dire le nombre de dimensions pour lesquelles les distances représenteront le mieux la structure des relations de dissemblances, consiste en la comparaison des critères d'information AIC et BIC (Winsberg & Carroll, 1989a et chapitre 4 §I.6.). On cherche donc le modèle donnant la valeur minimale de ces indices.

Les modèles obtenus à partir des deux distances de type euclidien ne font apparaître aucune structure d'espace (BIC du modèle spatial > au modèle nul ou nombre de dimensions trop important (7) vu le nombre de timbres étudiés (12)). Les modèles les plus appropriés pour les autres types de distances sont 2 dimensions sans spécificités pour la distance  $d_1$  et une dimension avec spécificités pour les distances  $d_2$  et  $d_3$ .

#### IV.3. Corrélations avec les dimensions perceptives, les paramètres acoustiques et les positions moyennes sur chacune des échelles sémantiques.

Nous avons tout d'abord calculé les corrélations entre les positions des timbres sur les dimensions des trois modèles sémantiques obtenus ci-dessus, et les positions de ces timbres sur les dimensions de l'espace perceptif dégagé dans l'étude précédente (Faure, 1996). Les seules corrélations significatives obtenues sont : D4 perceptive avec la dimension 2 dérivée de la distance d1 ( $r(10)=-.61$  ;  $p<.05$ ) et la dimension dérivée de la distance d2 avec D3 et D4 perceptives (respectivement  $r(10)=-.55$  et  $r(10)=-.57$  ;  $p<.1$ ). Les deux espaces perceptifs et sémantiques ne sont donc pas isomorphes.

Les corrélations calculées entre les positions des timbres sur les dimensions "sémantiques" et les différents paramètres acoustiques proposés par Krimphoff et al. (1994) ne sont significatives qu'entre les dimensions 1, dérivée des distances d1 et d2, et une mesure de flux spectral (respectivement :  $r(10)=-.58$  ;  $p<.05$  ; et  $r(10)=-.76$  ;  $p<.01$ ) Les corrélations effectuées avec les positions moyennes sur chacune des 23 échelles sémantiques aboutissent elles aussi à peu de corrélations significatives. Pour l'espace dérivé de la distance d1, la première dimension sémantique est fortement corrélée ( $p<.01$  ;  $d1=10$ ) à *net, pas sourd* et *pas diffus* et assez bien corrélée ( $p<.05$  ;  $d1=10$ ) à *clair, aigu, brillant, pas bas* et *haut*. La deuxième dimension sémantique étant faiblement corrélée ( $p<.1$  ;  $d1=10$ ) à *long* et *pas bas*. La dimension dérivée de la distance d2 ne présente qu'une corrélation significative avec *long* ( $r(10)=-.75$  ;  $p<.01$ ), aucune corrélation significative n'apparaissant avec la dimension dérivée de la distance d3.

#### IV.4. Discussion

Les 5 modèles multidimensionnels de la représentation sémantique du timbre, proposés ici en calculant des distances à partir de profils sémantiques présentent donc une structure très différente des espaces perceptifs obtenus dans la littérature. Ils sont de plus difficilement reliables à des paramètres acoustiques permettant de décrire les sons. Nos représentations perceptives et sémantiques du timbre semblent donc assez éloignées les unes des autres. De façon plus étonnante, nous n'avons trouvé de liens significatifs entre les dimensions perceptives calculées et les échelles sémantiques proposées que pour le modèle obtenu à partir de la distance d1 faisant pourtant appel à un seuil arbitraire pour déterminer les mots appartenant au profil sémantique de chacun des timbres. Il conviendrait dans le futur de faire varier ce seuil afin d'observer les modifications engendrées sur l'espace sémantique calculé.

De plus toutes les analyses effectuées dans ce paragraphe ont considéré chaque échelle avec le même "poids". Or, nous avons vu qu'un grand nombre d'entre elles étaient partiellement corrélées entre elles. On peut donc envisager d'améliorer la recherche de profils sémantiques des sons en pondérant chaque échelle en fonction de sa corrélation avec un plus ou moins grand nombre d'autres échelles. Cette technique est notamment utilisée dans des recherches pédagogiques sur la notation (les notes de mathématique et de physique étant plus souvent corrélées que celles de mathématique et d'anglais, comment donner une note générale rendant compte de toute la richesse d'un candidat, sans donner une importance excessive à un ensemble de disciplines qui soulignent la même qualité).

## V. Conclusion

Nous avons donc recherché des corrélats verbaux des dimensions perceptives, en nous inspirant de la recherche de corrélats acoustiques de ces dimensions, effectuée par Grey puis Krimphoff et al. (1994). Il apparaît d'emblée que cette analogie entre corrélats acoustiques et corrélats verbaux n'est pas du tout évidente. On s'attendait à trouver des adjectifs qui nous permettraient de comprendre le sens associé par les participants à chacune des dimensions perceptives obtenues par analyse multidimensionnelle des jugements de dissemblances perceptives, on obtient la plupart du temps des corrélats composites de dimensions perceptives, et la plupart des adjectifs de qualité analysés correspondent à des variations sur les dimensions 1 et 2. On peut ainsi se poser la question de la réalité perceptive des dimensions 3 et 4. Il est en effet peu réaliste d'obtenir un espace en 4 dimensions à partir de 12 sons. De plus, si on retrouve toujours les dimensions correspondant au temps d'attaque et au centre de gravité spectral dans l'ensemble des espaces obtenus en analysant des dissemblances perceptives entre timbres, les dimensions supplémentaires sont différentes d'une étude à l'autre. On peut donc se demander s'il ne s'agit pas d'un artefact de la technique d'analyse MDS.

Nous avons recherché ces corrélats verbaux à partir de deux types de recueil de verbalisations très différents : des verbalisations libres et des positionnements sur des échelles sémantiques. Les résultats nous permettent de résumer les avantages et inconvénients de ces deux méthodes.

L'avantage d'utiliser des verbalisations libres est de ne considérer que les comparaisons spontanées (et donc pertinentes) entre deux sons par rapport à une qualité donnée. Par contre, il est difficile d'obtenir des données suffisamment systématiques pour qu'une analyse quantitative soit significative. Inversement, le positionnement de sons sur un ensemble d'échelles sémantiques force les auditeurs à les comparer vis-à-vis de qualités qui ne les concernent peut-être pas directement. Ils sont donc obligés d'utiliser comme sens de ses attributs verbaux un sens assez général pour pouvoir s'appliquer à tous les sons de l'ensemble. Il est donc possible que le sens d'un même adjectif ne soit pas le même dans les deux contextes (verbalisations libres vs échelles sémantiques), même s'ils servent à décrire le même ensemble de sons. Par contre, les données recueillies avec un questionnaire sous forme d'échelles sémantiques présentent une forme tout à fait analysable avec des outils statistiques (données nombreuses et systématiques).

Il pourrait être intéressant de comparer ces deux points de vue dans d'autres études afin de mieux comprendre comment chacune de ces techniques influence les résultats obtenus. Ainsi, on comprendra mieux la nature de ces résultats et on pourra considérer ces méthodes non plus comme opposées mais complémentaires.

## Chapitre 6 : Description des sons

Depuis que les sons sont enregistrés, analysés, mixés, travaillés, créés par synthèse, indexés dans des bases de données (notamment sur le Web) ; depuis que des compositeurs, des architectes, des ingénieurs du son, des psychoacousticiens parlent de "design sonore" ; depuis que la matière sonore est devenue un objet reproductible et mesurable, nous sommes obligés de parler des sons, et en particulier de leur timbre. Cependant, le vocabulaire correspondant semble seulement en cours de construction et d'élaboration.

Ce chapitre est centré autour de la question : comment parle-t-on des sons ?

Cette question s'inscrit dans la question plus générale de savoir comment on parle de nos perceptions et donc dans une des problématiques actuelles du champ des sciences cognitives qui consiste à étudier les liens entre nos percepts et nos concepts (Dretske, 1969 ; Meini, 1996 ; Dubois, 1998). Des tentatives de description de notre vocabulaire en lien avec nos perceptions ont été en effet abordées dans différentes modalités, notamment en vision, dans le domaine de la couleur (Berlin & Kay, 1969 ; Dubois et al., 1997) ou dans la description de textures visuelles (Bhushan et al., 1997), en olfaction (Dravnieks & Bock, 1978 ; David et al., 1997) et à propos du goût (Urdapilleta, 1998).

Dans le domaine de l'audition, des chercheurs ont décrit la façon dont des participants parlent de sonneries de téléphone (Wright, 1971), de sons purs (Taylor et al., 1974), ou de bruits (Miller et al., 1976), et dans le domaine du timbre, des tentatives de descriptions sémantiques ont été développées en parallèle des descriptions acoustiques, souvent au sein du paradigme de la sémantique différentielle (Bismarck, 1974a ; Kendall & Carterette, 1992a,b).

Samoylenko et al. (1996) ont déjà donné un premier aperçu de la manière dont on parlait des mêmes sons que nous avons fait écouter à nos participants. Ils ont construit une base de données afin d'analyser des verbalisations libres de participants russes non-musiciens dans une expérience dont le protocole était analogue à celui de notre première expérience. Leur étude était centrée sur l'analyse des liens entre d'une part les différences de jugements numériques de dissimilarité inter et intra participants et d'autre part leurs correspondances avec des descriptions de similarité et de différence. Leur analyse étudiait aussi le caractère spécifique de l'identification des différents timbres sous forme d'instruments de musique.

La spécificité de la présente étude est due d'une part au fait que les verbalisations analysées ont été produites par des participants français de compétence musicale variée et d'autre part au fait que nous nous centrons surtout sur une description globale de la façon dont les participants parlent du timbre. Nous répondrons donc à la question : comment parle-t-on des sons ? essentiellement à travers l'analyse du contenu sémantique des verbalisations libres recueillies lors de notre première expérience.

Pourtant, comme nous avons aussi demandé aux participants de notre troisième expérience de décrire chacun des 12 sons présentés selon 4 consignes différentes, nous analyserons succinctement ces résultats et les comparerons à ceux recueillis dans notre première expérience.

Enfin, nous avons choisi ici de décrire les données verbales à l'aide d'un logiciel de traitement hiérarchique des propriétés : STONE, élaboré par Sébastien Poitrenaud (1998) dans l'équipe Processus Inférentiels du laboratoire Cognition Finalisée de l'Université Paris VIII-St Denis. Nous allons donc tout d'abord présenter ce logiciel et son intérêt dans l'étude des champs sémantiques de verbalisations libres, puis expliquer comment nous avons construit l'arborescence correspondant à nos données avant d'analyser la forme et le contenu de cette arborescence.

## **I. Le logiciel STONE : un arbre hiérarchique de propriété**

### I.1. Formalisme : des objets, des propriétés

La plupart des systèmes de gestion de bases de données utilisent un codage des propriétés des entités représentées à l'aide de la notion de couple attribut-valeur. Cette approche permet des implantations informatiques très performantes et très sûres. Cependant, beaucoup de données ne se laissent pas facilement représenter à l'aide d'un tel format. C'est par exemple le cas quand les valeurs ne sont pas exclusives mais constituent des listes. C'est aussi une structure très contraignante qui ne permet pas de rendre compte par exemple des relations entre attributs indépendamment des données, telles les implications entre les attributs.

De plus, avec les bases de données usuelles, on a d'une part le cadre de description des données (nombre d'attributs et structure des attributs) et d'autre part les données elles-mêmes qui s'inscrivent dans ce cadre. Les données ne permettent pas de modifier le cadre de description : on ne peut pas modifier d'une manière souple le cadre de description sauf à le changer, voire refaire une autre base avec une autre structure, ce qui sort du domaine de compétence de l'utilisateur qui saisit les données et les exploite. Or dans un domaine mal connu, on est amené très souvent à restructurer la description en fonction des cas rencontrés : introduction de nouveaux descripteurs, compactage de valeurs, dissociation de descripteurs, modification du découpage des objets, etc. Il faut disposer d'un moyen très rapide de faire cette modification, et surtout d'obtenir la mise à jour des données déjà décrites avec l'ancienne description.

Le modèle de description Stone a été implémenté par S. Poitrenaud (Poitrenaud, 1995, 1998). Il est basé sur la notion de propriété, définie comme une description arborescente et d'implication de propriétés. Les catégories se réalisent en s'organisant sous forme de réseaux sémantiques qui sont des réseaux de propriétés utilisant la méthode des treillis de Gallois (Barbut & Monjardet, 1970).

Formellement, un réseau se construit de la façon dont se construisent les réseaux sémantiques d'inclusion de propriétés pour décrire l'organisation sémantique des objets (Collins & Quilian, 1969). Dans une description Stone, il n'y a ni table, ni attribut, ni valeur, mais des descripteurs organisés en une arborescence. Dans sa forme actuelle, le programme Stone comporte une interface graphique qui intègre, dans un contexte unique, l'interrogation et l'édition des données.

Ce formalisme a été utilisé pour étudier la catégorisation d'objets, mais aussi pour étudier l'interprétation d'énoncés d'histoires drôles (Tijus & Moulin, 1997) comme modèle de la Compréhension par Assignment Dynamique de Signification (CASD). L'hypothèse défendue est que, dans ces textes, le contexte initial induit une mauvaise catégorisation du sens de certains mots clés polysémiques au début de l'énoncé, la chute de l'histoire amenant un basculement dans un autre sens possible qui crée un effet de surprise.

Nos données se situent à la frontière entre deux types d'utilisation de ce formalisme, c'est-à-dire entre, d'une part, une conception uniquement sous forme d'arborescence liée à la catégorisation des objets, et d'autre part, une analyse visant à comprendre le sens porté par un énoncé langagier.

## I.2. Construction de l'arborescence

Nous avons utilisé ce logiciel pour faire apparaître les grandes lignes du champ sémantique recueilli lors de la description de nos sons. L'avantage d'utiliser le formalisme de Stone pour analyser le sens de verbalisations libres tient dans deux de ses propriétés :

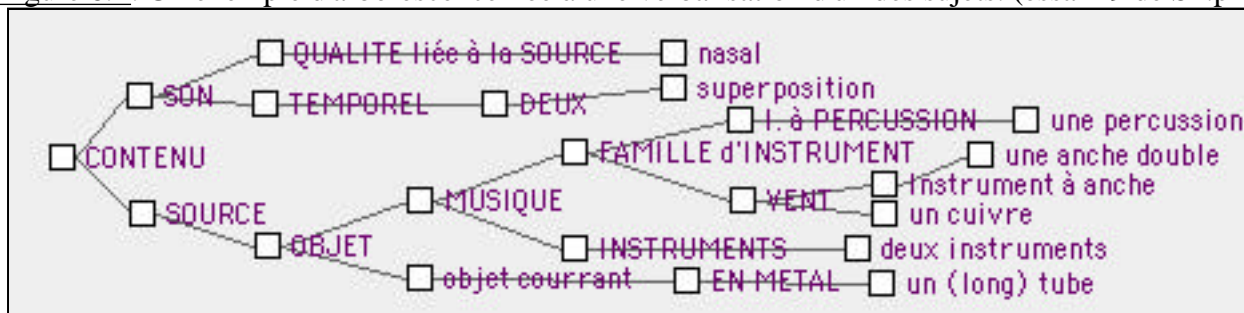
- Il est d'une part extrêmement souple d'utilisation : les catégories et sous-catégories peuvent être modifiées en permanence au cours du codage, permettant ainsi de rendre compte de la particularité de certains énoncés dans l'arborescence générale sans nécessiter le re-codage de l'ensemble des données par rapport à ce nouveau critère découvert dans les données. Ceci est un avantage majeur de ce type de formalisme par rapport à des bases de données plus classiques dont les différents champs ou niveaux d'analyses doivent être conçus avant l'expérience afin que toutes les verbalisations soient codées selon ces différents niveaux préétablis.
- D'autre part, il permet de réunir des énoncés dont les occurrences sont différentes à un certain niveau, mais dont le sens est identique à un niveau supérieur. Ainsi par exemple les verbalisations : "*se fait attendre*", "*n'arrive pas à sortir*", "*du mal à démarrer*", "*partlement*", bien que différentes, seront à un niveau donné considérées comme décrivant toutes quatre des caractéristiques temporelles des sons relatives à leur départ lent et seront donc rassemblées dans l'arborescence sous "*parle du son/temporel/début/début lent*".

Enfin, le codage comme l'interrogation des données peut se faire dans l'arborescence présentée soit sous forme de listes incluses les unes dans les autres, soit d'une manière graphique qui figure

l'arborescence grâce à quelques codes graphiques très simples (voir la figure en annexe A.6.1 pour un aperçu de l'interface).

Nous allons en effet utiliser ces deux qualités comme fondement de la construction de l'arborescence qui sera décrite par la suite. Il est en effet important de souligner que la structure arborée que nous allons analyser s'est construite au fur et à mesure du codage des verbalisations, celui-ci permettant à tout moment de regrouper certaines locutions, ou de créer une nouvelle catégorie spécifique (qui pourra par la suite être rassemblée à d'autres catégories qui apparaîtront lors du codage d'autres verbalisations). Ainsi par exemple, l'énoncé produit par le participant S2.pro lors de son essai 29 : "*Deux instruments. Deux sons superposés pour le premier: une percussion et un cuivre, comme peut-être un son de percussion, le premier prolongé par un long tube, et une anche double, un son assez nasal pour le deuxième.*" (essai 29 du participant S2.pro ), sera finalement instancié dans l'arborescence de la manière suivante (figure 6.1) :

Figure 6.1. Un exemple d'arborescence liée à une verbalisation d'un des sujets. (essai 29 de S2.pro)



Mais cela résulte du fait qu'au départ, cette production contenait : *Deux instruments, une percussion, un cuivre* : trois locutions qui parlaient d'instruments de musique, *un long tube, une anche double* : deux locutions qui parlaient de parties de l'objet supposé avoir produit le son, *Deux sons superposés, un son assez nasal* : deux locutions décrivant cette fois-ci le son et non sa source supposée, la première se focalisant plutôt sur des qualités temporelles, la seconde plutôt sur une qualité liée à la source (nasal = qui vient du nez). Et qu'au fil des codages des autres productions de l'ensemble des participants, nous avons été amenés à créer les catégories intermédiaires, par exemple, l'idée que "un tube" soit codé sous : objet/objet courant/en métal/ est venue du fait que d'autres productions faisaient apparaître la notion d'objets différents des objets courants (par exemple des objets indéfinis : "un truc") ou des objets courants qu'on pouvait regrouper autrement que par la particularité "en métal" (par exemple des tissus).

Il est nécessaire d'explicitier ici les symboles utilisés dans l'interface graphique de Stone. Dans les figures de ce chapitre, nous présenterons en général une partie des propriétés d'un sous-ensemble d'objets de notre base de données, un objet correspondant à une production d'un participant lors de la description d'une des 66 paires qu'il a écoutées.

Les propriétés sont en général en minuscules quand elles correspondent directement à des locutions verbalisées dont on a observé des occurrences, en MAJUSCULES quand elles ont été ajoutées lors du codage pour regrouper d'autres locutions (par exemple, dans la figure 1, MUSIQUE a été rajouté pour regrouper les propriétés sous cette catégorie, bien qu'aucune de celles-ci ne contienne le mot "musique" dans leurs locutions). Le mot en minuscule a sa première lettre en majuscule quand on trouve regroupée dessous cette propriété (qui correspond donc à une locution réelle d'un participant) d'autres locutions contenant ce mot mais plus spécifiques (par exemple, une sous-catégorie d'Instrument INDÉFINI est "deux instruments").

Les propriétés sont reliées entre elles par des segments qui correspondent à leur positionnement dans l'arborescence. Elles sont précédées d'un symbole et d'un chiffre. Le chiffre indique le nombre d'objets sélectionnés qui possèdent la propriété donnée, le symbole indique si cette propriété est partagée par tous les objets sélectionnés (un rond s'il y a plusieurs objets, un carré s'il n'y en a qu'un, comme dans la figure 6.1), un triangle indique que toutes les sous-catégories de la propriété donnée ne sont pas montrées dans le graphique (certaines ont été cachées pour des raisons de lisibilité). (Un rond dans un triangle indique que tous les objets sélectionnés contiennent la propriété mais que toutes les sous-catégories de cette propriété ne sont pas représentées.) On peut donc visualiser à tout moment l'intersection ou l'union de plusieurs objets choisis, à partir d'une propriété quelconque sur le ou les niveaux de codage que l'on souhaite.

Un inconvénient de notre codage est de ne pas différencier à l'intérieur des verbalisations les différentes parties décrivant respectivement le premier son, le deuxième et l'ensemble des deux. Ceci était difficile principalement pour des raisons techniques : en effet, Stone gère chacune des 66 verbalisations produites par chacun des 20 participants comme  $66 \times 20 = 1320$  objets ayant un certain nombre de propriétés spécifiques ou partagées (une trentaine dans l'exemple précédent, qui aboutissent pour l'ensemble des objets à plus de 1000 propriétés). Le fait de considérer pour chacune des verbalisations produites 3 objets différents, nous ferait passer à près de 4000 objets. Actuellement, les possibilités du logiciel et des ordinateurs sur lequel il tourne ne permettent pas de gérer autant d'objets et autant de propriétés conjointement.

Il est évident que la structure de l'arborescence dépend de ce qu'on voudra tester par la suite. Il n'y a donc pas une arborescence unique associée à un ensemble de verbalisations recueillies. Ainsi, comme nous avons voulu tester l'adéquation des verbalisations par rapport aux sons décrits, nous avons donné à l'arborescence construite à partir du contenu des verbalisations des participants une forme comparable à une classification possible des sons testés. Ainsi par exemple, nous avons réparti l'ensemble des noms d'instruments de musique produits par les participants en fonction de la famille d'instruments dans laquelle ceux-ci sont conventionnellement classés. Ceci permet à un niveau de l'arborescence de rassembler par exemple les verbalisations de *trompette*, *trombone*, *tuba*, *cor*, *cornet à piston* et *trompe* parmi des verbalisations où le participant "*parle de la source/objets/musique/instrument/instrument à vent/cuivre/*" et de les différencier à ce



niveau de verbalisations de *clarinette, flûte, hautbois...* rassemblées dans la sous-catégorie "*~/instrument à vent/bois/*", toutes ces verbalisations étant tout de même rassemblées au niveau des "*~/instrument à vent/*". Les sons testés contenant des imitations de sons d'instruments à vent, de la famille des cuivres (TPT et TBN) et de la famille des bois (CRA et CLA), ce niveaux de l'arborescence nous permettra de tester directement l'adéquation entre les verbalisations sur la perception d'un coté, et la source sonore visée par le luthier électronique de l'autre.

## II. Expérience 1 : Description de sons présentés par paires

Bien que la construction de l'arborescence ai été effectuée en partant des occurrences et en les regroupant peu à peu dans les catégories relativement spécifiques au début, et de plus en plus générales, nous allons analyser l'arborescence résultant de ce long codage en partant de son sommet (ou peut-être de sa racine, puis des catégories les plus génériques).

### II.1. Description générale

Ainsi, nous avons fini par répartir l'ensemble des verbalisations des participants sur chacune des paires de sons en 4 grandes catégories (voir tableau 6.1). Le contenu des verbalisations concernant chacune des paires de sons relève en général de plusieurs catégories, de natures différentes, ce qui fait que le codage ne répartit pas les productions entre les 4 catégories principales de manière exclusive. Une même paire peut être codée à la fois dans plusieurs catégories différentes.

Tableau 6.1. Répartition du contenu des verbalisations à propos des 1320 paires de sons dans les 4 premières catégories.

Verbalisations	NB de paires	% de paires
parlent du son	1300	98
parlent des ressemblance / dissemblance	842	64
parlent de la source	787	60
parlent du sujet	298	23

Le locuteur parle du ou des sons entendus (ce qui correspond *a priori* à la consigne). Ces verbalisations sont présentes dans 98% des productions (c'est à dire des essais correspondant chacun à une paire de sons). Dans les 2% de cas restants, 12 paires sont décrites uniquement en terme de source sonore, ce qui concerne 6 participants différents, principalement le musicien professionnel S2.pro (pour plus de la moitié des cas). Les 8 derniers cas concernent soit des cas où les participants n'ont parlé que de la ressemblance ou de la dissemblance des deux sons de la paire (*Ils sont différents.. c'est tout*), ce qu'on peut voir comme un signe de fatigue (5 cas), soit comme des erreurs d'enregistrement qui ne nous ont pas permises de recueillir leur verbalisations (3 cas).

Dans 64% des paires décrites, le locuteur parle des ressemblances ou dissemblances entre les sons (*ils se ressemblent, ils sont très différents...*), c'est-à-dire de leur activité de comparaison des sons, qui

correspond à l'autre partie de la consigne : donner un jugement de dissemblance entre les deux sons de chacune des paires écoutées.

Dans 60% des paires décrites, le locuteur décrit une source sonore qui aurait pu produire le son (*j'entends la corde et la bouteille*). Ceci n'était pas induit *a priori* par la consigne. C'est donc un résultat qui nous semble important. Quand on demande à des gens de décrire *les similarités et les différences entre les deux sons* qu'on leur fait entendre, dans presque deux tiers des cas, ils décrivent aussi un objet (musical, comme un instrument de musique, ou non), ou une action qui a produit ou qui aurait pu produire le son. Ceci pourrait être lié à la manière dont nous mémorisons les sons et à notre utilisation de ceux-ci. Si le timbre des sons que nous entendons est bien, comme le définit Schaeffer (1966, p55): "ce à quoi on reconnaît que divers sons proviennent du même instrument", et si les sons sont directement liés à des objets sonores ou des actions sur des objets qui produisent des sons, il est en effet logique que ces verbalisations apparaissent avec autant de force, bien que la consigne induise ici plutôt des descriptions de qualités sonores.

Enfin, dans 23% des paires décrites, le locuteur parlait de lui-même en émettant des jugements de valeurs (*j'aime bien, je déteste*), des commentaires, des explications (*quand je dis aigu, je veux dire surtout...*), ou en exprimant des difficultés (*c'est dur, j'arrive pas...*) ou encore en verbalisant des activités extérieures à la tâche (*j'en ai marre, j'ai faim...*). Ce type de verbalisations touche très exactement aux critiques effectuées sur l'utilisation de verbalisations libres comme données en psychologie, nous décrirons donc leur contenu.

### **II.1.1. Répartition en fonction des participants**

Nous avons souhaité analyser la répartition entre participants musiciens et non-musiciens de ces différents types de verbalisations. Les différences de comportement entre les participants des deux groupes : musiciens et non-musiciens pour chacune des 4 catégories de verbalisations, ne sont significatives que dans le cas où le participant parle de lui-même ( $\chi^2(1)=20, p=10^{-14}$ ). La compétence musicale n'influence donc pas le fait que le locuteur parle "des dissemblances" entre les deux sons de la paire ou "du son" ou "de la source" de celui-ci (dus respectivement pour 53, 50 et 51% aux participants musiciens). Par contre, les non-musiciens parlent plus souvent d'eux-mêmes que les non-musiciens : 72% de ce type de verbalisations leur sont dues. Ils expriment notamment plus de difficultés.

En fait, la répartition des verbalisations de chaque participant entre les 4 grandes catégories de verbalisations définies ici dépend plus de la personnalité de chacun des participants que de sa compétence musicale. En effet, seules les verbalisations à propos des sons ne dépendent pas du participant. Pour ce qui est des verbalisations des dissemblances entre les sons, comme pour celles où le participant parle de lui ou d'une source sonore, leur fréquence dépend du participant ( $\chi^2(19)>110, p<10^{-15}$ ). Ceci est particulièrement vrai pour les verbalisations sur soi-même.

Une analyse de cluster (voir la figure de l'annexe A.6.2) effectuée sur le nombre de paires pour chaque participant contenant des verbalisations de chacun des 4 types décrits dans ce premier niveau d'analyse, montre trois grands groupes de participants. Le premier groupe contient les 6 participants (dont un musicien) qui ont parlé d'eux-mêmes plus souvent que les autres (en moyenne dans 57% de leurs verbalisations alors que sur l'ensemble des participants  $M=23\pm 26\%$ ). Le deuxième groupe contient les 8 participants (dont 5 musiciens) qui ont parlé de la source plus souvent que les autres (en moyenne dans 80% de leurs verbalisations), mais aussi des participants dont la quantité de production de chaque type correspond aux productions moyennes. On peut noter que les trois musiciens professionnels qui ont participé à notre expérience en font partie. Enfin, le dernier groupe contient les 6 participants (dont 4 musiciens) qui ont parlé moins souvent que les autres de la source sonore (en moyenne dans 27% de leurs verbalisations alors que sur l'ensemble des participants  $M=59\pm 30\%$ ).

### II.1.2. Répartition en fonction des sons

Nous avons souhaité vérifier si le son présent dans les paires décrites, ou le type de paire de sons considérée (contenant deux hybrides, une ou deux imitations de sons soufflés, une ou deux imitations de sons pincés-frappés ou une imitation de son frotté), pouvait influencer les verbalisations produites par les participants. Aucune des différences considérées n'est significative, ni en fonction des 12 sons ( $\chi^2(11) < 8, p > .70$ ), ni en fonction des 11 types de paires de sons possibles ( $\chi^2(10) < 7, p > .73$ ).

On peut tout de même noter quelques tendances intéressantes. En effet, les participants verbalisent plus à propos des dissemblances entre les deux sons d'une paire s'il s'agit d'une paire comportant deux sons imitant des sons soufflés (dans 72% de ces paires alors qu'en moyenne  $M=64\pm 3\%$ ). Ces sons sont parmi les sons les plus proches, donc les plus semblables. De même, les participants verbalisent un peu plus de commentaires personnels, de difficultés... quand les deux sons de la paire sont tous les deux des hybrides (dans 30% de ces paires alors qu'en moyenne  $M=23\pm 2\%$ ) et moins si ce sont tous les deux des imitations de sons pincés ou frappés (dans 20% de ces paires). Les sons imitant des sons pincés ou frappés étant les imitations les plus vraisemblables donc sans doute les plus facilement identifiables à des sons d'instruments de musique connus. On trouve donc une tendance à verbaliser plus de commentaires personnels quand les sons sont moins facilement identifiables (ou reconnaissables ou connus).

Enfin, les participants décrivent moins souvent une source sonore quand les deux sons de la paire sont tous les deux des hybrides ou tous les deux des imitations de sons soufflés (respectivement dans 47 et 48% de ces paires alors qu'en moyenne  $M=58\pm 6\%$ ). Par contre ils décrivent plus souvent la source quand un des deux sons de la paire est le son CFT (dans 65% de ces paires). De la même façon, les participants ont tendance à parler plus souvent de la source sonore pour les paires dont un des deux timbres est soit le son VBR soit le son CVC (respectivement dans 145 et 147 des 220 paires verbalisées contenant chacun de ces sons alors que  $M=132\pm 10$  paires). Il est intéressant de remarquer que ces deux sons sont parmi les plus "vraisemblables", c'est à dire que l'imitation du son de l'instrument considéré est la plus plausible. Les cas où les verbalisations de ce type sont les moins

fréquentes sont les paires contenant les 4 timbres CRA, SNO, VBN et OBC, qui sont respectivement l'imitation de cor anglais (dont le résultat donne un effet particulièrement synthétique) et les trois hybrides testés dans cette expérience.

### II.1.3. Résumé

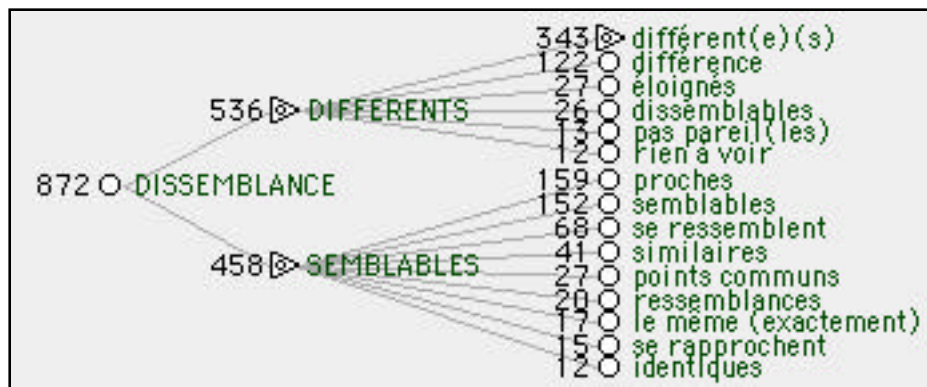
On observe 4 grandes catégories de verbalisations selon que le locuteur parle du son, des différences entre les sons, d'une source sonore possible ou de lui-même. L'importance de chacune de ces catégories dépend plus de la personnalité des participants que de leur compétence musicale, bien qu'on observe une plus grande proportion de verbalisations sur soi-même chez les locuteurs non-musiciens. La répartition des verbalisations entre ces 4 grandes catégories ne dépend pas des sons présents dans les paires écoutées ni du type de ces paires. Malgré tout, on note une tendance à parler d'autant plus de la source qui a pu produire le son entendu que le son paraît réellement être émis par un instrument de musique connu. Il semble que quand une paire contient deux sons soufflés ou deux hybrides (qui ont des caractères de surfaces proches), les participants les décrivent moins sous forme de source sonore mais ont plus tendance à comparer directement les sons.

Chacune de ces 4 grandes catégories de verbalisation va être étudiée dans le détail dans les paragraphes suivants. Nous commencerons par celles qui demandent le moins d'analyses, soit celles qui concernent les verbalisations de ressemblances et de dissemblances, plus celles où le locuteur parle de lui-même. Nous aborderons enfin les verbalisations où le locuteur parle du son, puis d'une source sonore possible.

## II.2. Parler de ressemblances et dissemblances

Nous avons déjà souligné l'importance des verbalisations sur les ressemblances et dissemblances entre les deux sons de chacune des paires (présentes dans 64% des 1320 paires). En effet, confrontés à la double tâche de donner, pour chacune des 66 paires de sons proposées, d'une part un jugement numérique de dissemblance (de 1 à 9) entre les deux sons et d'autre part de "décrire verbalement les similarités et les différences entre les deux sons entendus avec le plus de détails possibles", les participants ont très souvent commencé chacune de leurs verbalisations par des énoncés du type : "*ils sont très différents*", ou "*là ils sont très proches*". La figure II.1 donne un aperçu des sous-catégories de cette partie de l'arborescence, seules les verbalisations ayant totalisé plus de 10 occurrences ont été représentées sur le graphique.

Figure 6.2. Verbalisations codées sous *dissemblance* (représente 842 paires) Avec, en majuscule des regroupements de verbalisations, en minuscule des occurrences de verbalisation.



Il faut noter que les verbalisations : *ils sont très différents* ou *ils sont peu différents* ont été codées toutes les deux sous la catégorie *différent(e)(s)*. Les verbalisations *très peu différents* et *très peu semblables* ont par contre été séparées chacune dans les deux catégories *différent / semblable*. De plus, sur les 842 paires concernées par des verbalisations de type ressemblances et dissemblances, 120 d'entre elles (soit 14%) présentent à la fois des verbalisations de type *semblables* ET *différents*. En effet, les sons peuvent être considérés comme proches pour certaines qualités et différents pour d'autres. Les conclusions que cette partie de l'analyse peut proposer sont donc assez limitées.

Pourtant, il est intéressant de comparer ces verbalisations avec les jugements numériques effectués par les participants pour juger la dissemblance perceptive des deux sons de chacune des paires. En effet, les différences de répartition des verbalisations appartenant aux catégories *différents* et *semblables* (de même que les verbalisations relevant des deux catégories *différents* ET *semblables*) en fonction du jugement numérique de dissemblance (1 à 9) donné par les participants sont significatives (respectivement  $\chi^2(8) = 53, 136$  et  $18, p=10^{-9}, 10^{-25}$  et  $0.02$ ). Lorsque les participants jugent les sons d'une paire comme très ressemblants (jugement numérique de 1) ils produisent des verbalisations catégorisées comme *semblables* dans 72% des cas, alors que ces verbalisations sont présentes dans seulement 4% des productions effectuées lors des paires jugées comme très différentes (jugements numériques de 9). Au contraire, des verbalisations catégorisées comme différent sont présentes dans respectivement 62% des paires jugées numériquement très différentes (9) et 19% des paires jugées très semblables (1). Par contre, on trouve le plus souvent des verbalisations catégorisées à la fois comme *différents* ET *semblables* dans les paires jugées entre les deux extrémités de l'échelle numérique (19% des paires jugées 5 alors qu'en moyenne, c'est le cas de  $9.2 \pm 4.6\%$  des paires jugées de 1 à 9).

De plus, si on effectue une corrélation de rang de Kendall (pour l'ensemble des 842 paires concernées) entre d'une part le jugement de dissemblance émis par les participants et la présence de verbalisations de type *semblables* ou *différents* (en codant les verbalisations 1 pour *semblables*, 3 pour *différents* et 2 pour *semblables* ET *différents*), on obtient une corrélation très significative ( $z(840)=22, p<.0001$ , corrigé pour les ex-aequo). Il y a donc redondance entre les jugements de dissemblance et les verbalisations des participants sur les différences et ressemblances entre les sons.

Si on ne peut donc pas considérer que la tâche de verbalisation gêne la tâche de jugement de dissemblance (Faure et al., en préparation), par contre le contenu des verbalisations est influencé par les jugements perceptifs effectués par les participants. En effet, les participants ont tendance à verbaliser la tâche de jugement de dissemblance qui accompagne celle de verbalisation.

### II.3. Parler d'eux-mêmes

Quand les locuteurs en profitent pour parler d'eux-mêmes (dans 23% des 1320 paires décrites), ils abordent des sujets très variés au moyen de locutions pour la plupart produites une seule fois par un seul participant. Nous avons tenté de les rassembler dans 6 sous-catégories que nous allons décrire brièvement. Puis nous analyserons comment ces verbalisations se répartissent en fonction des différents groupes de participants. En effet, parmi les 298 paires où un locuteur a parlé de lui-même :

1- Il exprime des *difficultés* (dans 40% des cas), soit d'ordre général (*j'ai du mal* ou *je sais pas*), soit des difficultés d'expressions (*c'est difficile à dire*), soit des exclamations (*ou là !*), des difficultés de cohérence (*c'est pas cohérent*) ou des erreurs commises (*j'ai fait une fausse manip*).

2- Il verbalise des activités extérieures à la tâche proposée (dans 34% des cas). Ce peut être ce que j'ai appelé des *délires* (*après avoir mangé je suis plus gentil* ou *je vais vous parler un peu de moi*), des passages de fatigue (*je baille un coup*), des moments de rire ou d'amusement (*je le fais bégayer quand j'appuie plusieurs fois sur le bouton*), ce peut être aussi plus sérieusement des remarques par rapport à l'enregistrement ou à la passation de l'expérience (*oui, je regarde, ça enregistre, c'est bon*), des exhortations à se concentrer (*il faut se remettre dans le bain*) particulièrement après la pause, des remarques sur leur jugement numérique de dissemblance (*j'ai décalé mon jugement*), ou parfois des choses qui n'ont vraiment rien à voir (*il faut que j'arrête de manger des cookies*).

3- Il peut aussi exprimer une impression de *répétition* (dans 20% de ces cas), soit qu'il a l'impression générale que c'est toujours pareil (*toujours la même chose*), qu'il a déjà entendu les sons ou la paire considérée (*le même son que tout à l'heure*), qu'il a l'impression de dire la même chose (*je me répète*) ou qu'il va donner le même jugement numérique (*même jugement*).

4- Il exprime parfois des jugements de valeur sur les sons entendus (dans 19% des cas), soit pour dire qu'il aime (*"j'ai un petit faible pour le violoncelle"*) ou au contraire qu'il n'aime pas certains sons (*j'aime pas les sons aigus*), soit pour émettre un avis indifférent (*ça me choque pas*).

5- Le locuteur peut aussi exprimer des questions qu'il se pose (dans 12% des cas où il *parle de lui*), soit sur la manière d'exprimer ce qu'ils perçoivent (*on dit vibrato, qu'est-ce qu'on dit ?*), soit sur l'identité de ce qu'ils vient d'entendre (*c'est quoi ce cirque ?*), soit encore sur le jugement de dissemblance qu'il va donner (*combien je vais mettre ?*).

6- Enfin, il peut ressentir le besoin d'exprimer des explications (dans 11% des cas) soit sur le sens des mots qu'il emploie (*pincé dans le sens nasillard*), sur les procédures qu'il utilise (*j'essaye de prendre le parti de faire la différence entre instruments à cordes et instruments à vent*), soit en exprimant son désarroi par rapport à ses attentes (*c'est bizarre, enfin je me comprends*).

Ces verbalisations sont donc de nature et d'intérêt très différents : autant les expressions d'activités annexes à la tâche demandée ne nous informent que sur le niveau d'attention du locuteur, autant des explications sur le sens des mots utilisés peuvent être précieuses.

#### **Répartition en fonction des participants**

Nous avons déjà remarqué que la fréquence de ces verbalisations ne dépendait pas des sons écoutés, mais par contre de la compétence et de la personnalité des locuteurs. En effet, ces verbalisations sont surtout le fait de non-musiciens (pour 72% d'entre elles) et tout particulièrement de 4 d'entre eux. Les différences entre locuteurs sont encore plus marquées si on considère les sous-catégories de ce type de verbalisations.

Si on regarde la répartition des verbalisations sur les 6 sous-catégories où le participant parle de lui-même, les différences entre musiciens et non-musiciens sont toujours significatives ( $\chi^2(5)=51, p=9.10^{-10}$ ). Ainsi, les jugements de valeur (*j'aime* ou *j'aime pas*) sont quasiment exclusivement dus aux non-musiciens (54 paires sur 56). Pour les verbalisations de type *répétition* ou *autres activités* ou encore *questionnement*, la participation des musiciens est toujours inférieure à 28% de ces verbalisations. Pourtant, dans la sous-catégorie de verbalisation *autres activités*, les musiciens produisent 50% des verbalisations sur *l'enregistrement* ou la *passation* ainsi que des remarques sur les jugements de dissemblances. Par contre, aucun d'entre eux ne se permet de verbaliser des *délires* ou des choses n'ayant rien à voir avec l'expérience qu'ils sont en train de passer. La présence de ce type de verbalisations peut donc être un indice de difficulté de la part du locuteur. Enfin, 43% des paires où un participant verbalise une *difficulté* sont dues aux musiciens (et ils sont auteurs de 52% des verbalisations de difficulté d'expression et de 67% des verbalisations d'erreurs). Les musiciens sont aussi auteurs de 50% des verbalisations d'*explications*, et c'est d'autant plus vrai s'il s'agit d'explications sur le sens de mots utilisés, puisqu'ils ont produit 69% de ce type de verbalisations.

Nous avons déjà vu que ce type de verbalisations était surtout le fait de 4 participants, tous non-musiciens, pour lesquels ces verbalisations interviennent dans respectivement 76, 73, 65 et 50 % de leurs productions ( $M=23\%$ ). Si on reprend les sous-catégories ayant recueilli plus de 10 occurrences, la répartition entre les participants devient encore plus extrême, certains types de verbalisations n'étant en fait dus qu'à un ou deux d'entre eux (par exemple, 73% des verbalisations de *délires* sont dues au même participant). Même si 85% des participants produisent au moins une fois une verbalisation sur eux-mêmes, c'est en général le cas pour très peu de paires (moins de 6 sur 66 pour 60% d'entre eux). Sur l'ensemble des 25 sous-catégories constituées, seules 7 d'entre elles (exprimant la fatigue, des difficultés d'ordre général ou d'expression, des erreurs, des exclamations, le fait d'avoir l'impression

d'avoir déjà entendu un son ou des questionnements sur *comment dire* ce qu'ils perçoivent ou ressentent) ont été produites par plus d'un tiers des participants.

Les moments où les locuteurs parlent d'eux-mêmes sont donc relativement rares. Ces verbalisations présentent une grande diversité et révèlent d'éventuelles distractions, difficultés ou préoccupations extérieures à ce qui leur est demandé. Elles semblent très spécifiques d'un ou deux participants qui expriment ainsi plus leur personnalité propre, alors que dans les autres types de verbalisations, la conformité aux stimuli ainsi que le caractère conventionnel du langage, nécessaire si on veut être compris, rendent les verbalisations plus homogènes entre les locuteurs. Loin de considérer ces verbalisations comme des preuves de l'impossibilité de recueillir des verbalisations libres, nous pensons que c'est un moyen pour les sujets de "dire tout haut ce qu'ils pensent tout bas", et donc pour l'expérimentateur, de mieux comprendre les difficultés liées à la tâche demandée.

#### II.4. Décrire les sons

Ces verbalisations sont *a priori* ce que la consigne induisait comme réponse. Nous avons déjà remarqué que les participants s'y sont conformés dans la grande majorité des cas. Nous avons souhaité analyser la répartition de ce type de verbalisations en différentes sous-catégories. Un aperçu de ce niveau de l'arborescence est donné par la figure 6.3.

Figure 6.3. Verbalisations codées sous *le son*, qui représente 1300 paires. (Sous-catégories rassemblant plus de 10 occurrences)



Dans 80% des cas où un participant parle des sons écoutés, il va se consacrer à décrire une caractéristique *temporelle* de celui-ci, principalement la façon dont il commence, se déroule et finit. Ceci est à relier à la nature de la première dimension perceptive dans notre espace perceptif (Faure et al., 1996), corrélée au logarithme du temps de montée de l'enveloppe temporelle, qui a donc un caractère temporel. Il s'ensuit que les 12 sons testés dans notre expérience diffèrent par leurs caractères temporels de manière très saillante, ce qui est systématiquement noté dans les verbalisations. En fait, la plupart des descriptions d'objets et d'actions (analysés dans le paragraphe suivant) reflètent aussi des caractéristiques temporelles : les instruments frottés et soufflés ont des attaques lentes alors que les instruments pincés et les frappés ont des attaques rapides. Nous touchons ici à une des caractéristiques principales des sons testés que nous analyserons en détail.



Ensuite, 73% des verbalisations sur le son contiennent des locutions décrivant des *sensations* relevant des 5 sens. Nous avons regroupé sous cette catégorie toutes les verbalisations ayant trait au sens de l'audition bien sûr, mais aussi à la vision, au toucher et au goût. En effet, nombre des verbalisations à propos des sons se font par métaphore avec un autre de nos sens. Un son peut ainsi être jugé *brillant, clair, mou, chaud* ou *acide*.

Les locuteurs décrivent aussi la *forme* qu'ils perçoivent du son (dans 44% des cas où ils décrivent le son). Ces verbalisations concernent à la fois des qualités qu'on a coutume de considérer comme visuelles et qui, dans notre cas, se rapprochent beaucoup des descriptions temporelles. En effet, les participants décrivent les sons un peu comme s'ils avaient à décrire l'enveloppe de la forme d'onde de ce son obtenu grâce à un oscilloscope, c'est-à-dire une représentation graphique de son déroulement temporel. Ainsi, le début, le milieu puis la fin du son sont décrits comme si cela constituait sa forme. Le son aurait ainsi une forme temporelle au contraire de l'image qui aurait une forme spatiale. Il est intéressant de montrer que ces formes décrites par les participants, et ce sans distinction de compétence musicale, concernent des sons dont la durée moyenne est de l'ordre de 836 ms. Nous sommes donc tout à fait capables de décrire des phénomènes qui se déroulent pendant des périodes de temps très brèves.

Puis, les participants décrivent les émotions évoquées par les sons ou émettent des jugements de valeur (dans 25% des cas). Il s'agit ici principalement de donner leur impression sur le caractère agréable ou désagréable des sons entendus.

Enfin, dans la partie générale, nous avons regroupé un certain nombre de qualités que nous n'avons pas su classer ailleurs, par exemple les qualités liées aux notions de *riche-pauvre, propre-sale*, ou de *pureté*. Mais nous y avons aussi fait figurer les notions générales sur le type de son ou au contraire les marques de spécificité d'un son qui est reconnu à chaque nouvelle écoute. Ces verbalisations représentent 21% des cas où les participants parlent des sons. Nous avons souhaité séparer les descriptions de qualités des sons, liées à l'identité de la source sonore (21% des cas). Ce sont en effet des verbalisations intermédiaires entre le fait de décrire le son ou sa source : *un son cristallin* est une façon de décrire un son au moyen d'un adjectif directement dérivé d'un nom de matériau. Par contre, les verbalisations *c'est vraiment du cristal, ou ressemble à du cristal*, seront codées dans la partie *source/objet/matériau*. Nous avons aussi mis en exergue les qualités liées à l'*intensité* ou à l'*énergie* perçue et véhiculée par les sons entendus. Ces verbalisations représentent respectivement 11 et 6% des cas. Enfin, nous avons codé les verbalisations où les participants donnaient un *nom* au son (*une note, un grésillement*) ainsi que les cas où, désespérant de ne pouvoir les décrire, ils tentaient d'imiter les sons. Ceci représente respectivement 12 et 2% des verbalisations et renvoie à l'idée de lexicalisation du domaine sonore.

Comme les verbalisations sur *le son* sont produites à propos de 98% des paires de sons, il est normal que la répartition des participants sur ce niveau ne dépende ni du participant, ni du groupe de

participants considéré. Nous nous intéresserons donc à la répartition des participants au niveau des sous-catégories.

#### II.4.1. Déroulement temporel

Nous avons déjà souligné l'ampleur du phénomène : tous les participants produisent ce type de verbalisations et à propos d'un grand nombre de paires de sons (79% de l'ensemble des paires). Nous allons analyser la répartition de ces verbalisations selon les sous-catégories dégagées.

Les verbalisations d'ordre temporel sont essentiellement composées de la description des différentes parties des sons. Nous avons regroupé les descriptions concernant le *début*, le *milieu* et la *fin* des sons (l'ensemble de ces trois catégories concerne 70% des 1320 paires). Les verbalisations concernant le *milieu* du son (*évolution / pas d'évolution*) sont les plus nombreuses (53% des paires contenant une description du déroulement temporel). Ensuite viennent les verbalisations concernant la *fin* des sons (48% des paires) et celles concernant le *début* des sons (47% des paires). Comme les caractéristiques verbalisées ici correspondent au paramètre perceptif qui semble le plus saillant dans l'ensemble de ces sons (la première dimension perceptive de ces sons est liée au logarithme du temps de montée de l'enveloppe temporelle, soit justement le temps d'attaque (McAdams et al., 1995)), nous en ferons la description systématique dans les paragraphes suivants.

Différentes *qualités temporelles* sont aussi décrites (32% des paires), notamment des qualités liées à la *durée* (*court/long*) ou à la *vitesse* (*vite/lent*) mais aussi des qualités moins marquées temporellement, comme les notions de *sec* ou *incisif*, mais qui nous semblent décrire une forme de rapidité d'attaque. Enfin, on peut noter des verbalisations ayant trait au caractère unique ou non de chacun des deux sons de la paire (*un-deux sons*), considérées comme d'ordre temporel du fait que ce caractère *double* est souvent lié au début de certains sons. En effet, parmi les trois sons hybrides proposés, certains gardaient une trace perceptive de la manière dont ils avaient été construits, ce qui pouvait amener les participants à mentionner une *juxtaposition*, *deux sons*, ou *un son doublé*... et donc par opposition *un seul son*, ou *un son unique*. Ces deux types de verbalisations ne concernent que 8% des paires mais seront intéressantes à confronter aux sons écoutés.

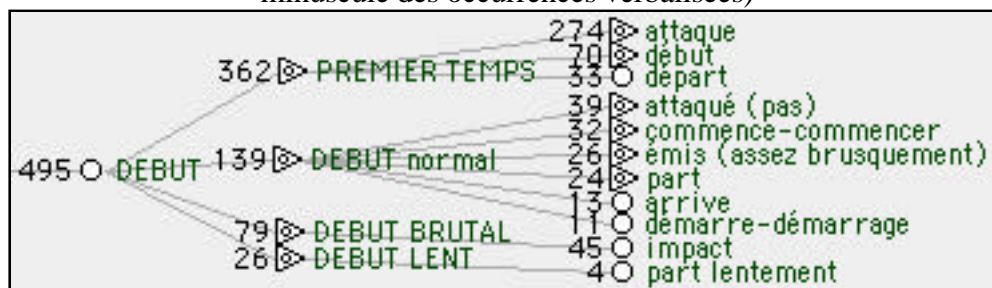
Ces verbalisations concernant la totalité des participants, on ne s'attend pas à observer de différences de comportement entre musiciens et non-musiciens et en effet, globalement, sur l'ensemble des verbalisations temporelles, la contribution des musiciens et des non-musiciens est identique (50%). Pourtant, les différences de répartition entre musiciens et non-musiciens sont significatives entre les 5 sous-catégories de ce niveau d'analyse ( $\chi^2(4)=33, p=8.10^{-7}$ ). Les musiciens sont locuteurs de 77% des remarques sur le caractère *double* de certains sons, mais seulement de 40% des descriptions du *milieu* des sons. On peut rappeler qu'objectivement parlant, les sons écoutés avaient été égalisés en durée globale perçue.

La répartition de ces verbalisations en fonction des sons présents dans les paires décrites nous paraît plus intéressante à étudier. En effet, ces paires de sons reflètent justement des différences de caractère temporel et il serait intéressant de voir si les verbalisations en rendent compte. Pourtant, à ce niveau de description, tous les sons sont représentés de la même façon dans les descriptions appartenants à chacune des 5 grandes catégories décrites ci-dessus. Ce résultat est sans doute dû au caractère assez systématique de la description des caractéristiques temporelles des sons.

#### Le début des sons

Les verbalisations à propos du *début* des sons sont présentes pour 495 paires de sons, soit 37% des paires et 47% des verbalisations d'ordre temporel. Un aperçu de la répartition de ces verbalisations dans ses sous-catégories est donné dans la figure 6.4.

**Figure 6.4.** Verbalisations codées sous : *parlent du son / temporel / début* (495 paires, sous-catégories rassemblant le plus d'occurrences. Les mots en majuscule représentent des catégories, ceux en minuscule des occurrences verbalisées)



Ainsi, parmi les verbalisations où un locuteur décrit le début des sons entendus, il parle très souvent de ce premier temps (73% des cas), désignant ce moment principalement comme l'*attaque* (dans 274 paires, soit 55% des cas), mais aussi le *début* ou le *départ* des sons. Ensuite les participants caractérisent souvent ce premier temps du son à l'aide d'un des verbes rassemblés dans la sous-catégorie *début normal* (55% des cas). Les verbes les plus fréquemment utilisés dans ce contexte sont : *attaquer*, *commencer*, *émettre*, *partir*, *arriver* ou *démarrer*. Il est étonnant de remarquer que dans ce contexte les verbes *partir* et *arriver* sont utilisés comme synonymes. (par exemple : *il part d'un coup* et *il arrive d'un coup*). Tout se passe comme si l'auditeur se place soit du point de vue du son soit du point de vue du récepteur du son. La description de ce premier temps des sons peut révéler déjà la notion de début rapide ou brutal ou au contraire lent. Nous avons donc réparti les locutions des participants dans ces deux catégories de verbalisations qui représentent à elles deux 21% des verbalisations sur le début des sons. La catégorie *début brutal* rassemble 20 locutions différentes, parmi lesquelles les notions d'*impact* et d'*impulsion* rassemblent le plus grand nombre d'occurrences. La catégorie *début lent*, réunie 19 locutions différentes, ayant chacune moins de 4 occurrences atteintes par *part lentement* et *viennent comme ça du vide*.

La répartition de ces verbalisations, en fonction des timbres contenus dans chacune des paires, présente quelques différences significatives (tableau 6.2), chaque son étant écouté par chaque sujet

dans 11 paires, soit 220 fois pour l'ensemble des sujets. Les deux différences qui sont significatives concernent la répartition des verbalisations de *début lent* et paradoxalement de *début normal*.

Tableau 6.2. Répartition du nombre de paires décrites par : *le son / temporel / début* en fonction des timbres contenus (au maximum 220 = 11 paires \* 20 sujets contiennent chacun des sons).

Timbre	<i>début</i>	<i>début brutal</i>	<i>début lent</i>	<i>début normal</i>	<i>premier temps</i>			
TPT	76	12	2	29	50			
TBN	82	10	6	25	61			
CRA	66	-	13	21	43			
CLA	74	8	-	28	46			
CFT	100	+	16	<b>13</b>	+	<b>39</b>	+	69
VBR	84	18	+	2	19	60		
PNE	78	12	6	<b>14</b>	-	67		
PNO	88	15	1	24	65			
CVC	82	15	4	19	62			
OBC	80	8	-	2	22	63		
VBN	91	20	+	6	21	68		
SNO	87	11	6	<b>17</b>	-	68		
moyenne	82.33	13.17	4.33	23.17	60.17			
écart type	8.76	3.76	3.39	6.60	8.95			
$\chi^2$	10.26	11.82	<b>29.23</b>	<b>20.70</b>	14.65			
p	0.51	0.38	0.002	0.036	0.20			

avec + (respectivement -) pour les valeurs s'écartant de la moyenne de plus d'un écart type. Les différences significatives sont en gras (valeurs de  $\chi^2$  par rapport à la moyenne attendue).

En effet, les verbalisations de *début lent* sont beaucoup plus fréquentes quand un des deux sons de la paire est le son CFT, ce qui est bien normal considérant que c'est le seul son ayant une attaque franchement progressive. Par contre, les verbalisations de *début normal* ne présentaient a priori pas de caractère discriminant par rapport aux sons écoutés et ont, malgré tout, été réparties de façon significativement différente entre les sons. Elles apparaissent plus souvent quand un des sons de la paire est le son CFT et moins souvent quand c'est un des deux sons SNO et PNE. Ce résultat est difficile à interpréter, le son SNO étant un des plus proches du son CFT comme hybride de corde frottée et de piano.

Les verbalisations sur le *premier temps* sont très fréquentes et ne présentent *a priori* pas de caractère lié à un mode de production des sons, les différences de répartition en fonction des sons présents ne sont d'ailleurs pas significatives. Pourtant, on observe une légère tendance à une fréquence moins élevée quand un des deux sons de la paire est un des deux sons CRA ou CLA qui font partie des sons ayant un temps d'attaque le plus lent et donc une attaque moins marquée.

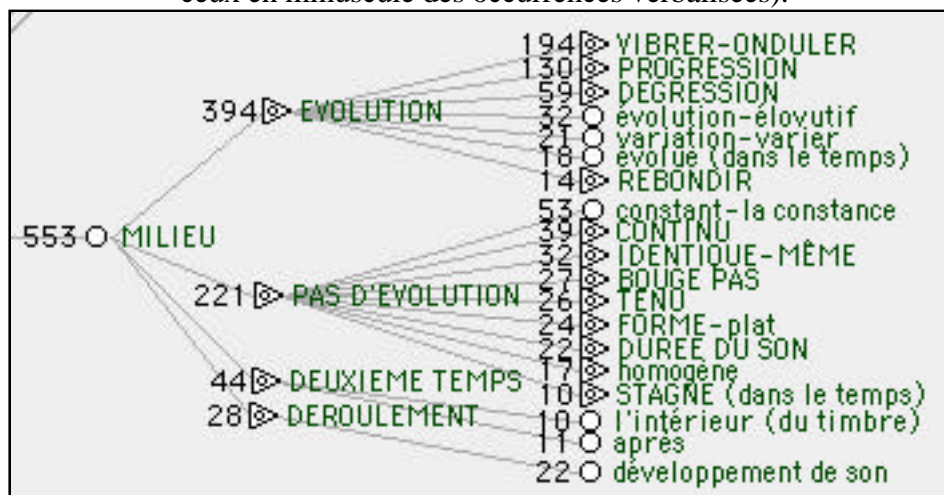
On remarque aussi une tendance à parler plus souvent du *début* des sons quand un des deux timbres de la paire est le son CFT, et moins souvent quand un des deux sons de la paire est le son CRA. On retrouve cette tendance quand les deux sons de la paire sont des hybrides. La tendance inverse se manifeste quand les deux sons sont des imitations de sons soufflés, dont le son CRA fait partie (respectivement pour 43 et 28% de ces types de paires alors que  $M=38\pm5\%$ ).

De même, on observe une tendance à parler plus souvent de *débutbrutal* quand un des deux sons de la paire est le son VBN ou le son VBR, tous deux percussifs, et moins souvent quand un des deux sons de la paire est le son CLA (imitation de clarinette donc de son soufflé) mais aussi du son OBC (hybride perçu comme percussif) ce qui s'explique moins bien. On note aussi une tendance à parler moins souvent de *débutbrutal* quand les deux sons de la paire sont des imitations de sons soufflés (dans 2% de ce type de paire alors que  $M=6\pm 2\%$ ). Ceci s'explique très facilement, les sons soufflés ayant des temps d'attaque beaucoup plus lents que les autres sons.

#### Le milieu des sons

C'est le type de verbalisation le plus fréquent dans l'ensemble des verbalisations d'ordre temporel puisqu'on le rencontre à propos de 553 paires de sons soit 42% d'entre elles. Nous allons en faire une description dans la figure 6.5.

**Figure 6.5.** Verbalisations codées sous : *parlent du son / temporel / milieu* (553 paires, sous-catégories rassemblant le plus d'occurrences. Les mots en majuscule représentent des catégories, et ceux en minuscule des occurrences verbalisées).



On peut remarquer que les verbalisations sur le milieu des sons sont dominés par la notion d'*évolution* au cours de l'événement (elle représente 30% des paires de sons). Nous avons subdivisé cette catégorie en rassemblant d'une part, toutes les verbalisations qui avaient trait à la notion de vibration ou d'ondulation (194 paires), principalement représentée par des occurrences de *vibrer* ou de *vibration*, et d'autre part des évolutions du type *progression* ou *dégression*. La sous-catégorie ayant trait à la notion de *dégression* est surtout représentée par des verbalisations faisant appel à la notion de *descendre*, et représentée par exemple par des occurrences de *redescend* ou *va vers le bas* (respectivement 10 et 12 occurrences). La sous-catégorie ayant trait à la notion de *progression*, plus fréquente (dans 130 paires de sons), se répartit entre plusieurs tendances : progresser en hauteur avec la notion de *monter* (principalement représentée par 16 occurrences de *va vers le haut*), progresser en volume avec la notion de *plus de place* (principalement représentée par les 11 occurrences de *s'étale*),

en taille avec *plus grand* (principalement représenté par 8 occurrences de *se développe*), auxquelles on ajoute la notion de *mouvement* (représentée par des occurrences de *poussé* ou *repart*).

On trouve symétriquement à la notion d'*évolution*, celle de *non-évolution* qui se répartit en verbalisations de *constant, continu, identique* (avec 20 occurrences de (*reste*) *le même (pendant toute la durée du son)*), *bouge pas, tenu, plat, homogène, stagne*, ainsi que des verbalisations sur la durée de cette non-évolution avec par exemple 21 occurrences de *du début à la fin*.

Enfin, on trouve des occurrences de verbalisations dont le but est de localiser ce moment du son, rassemblées dans les catégories *déroulement* et *deuxième temps*, mais qui ont une importance moindre dans les verbalisations sur le milieu du son qu'en avaient les localisations sur le *premier temps* des sons dans la partie *début* (12% de ces verbalisations contre 73% pour les verbalisations sur le début des sons). On y trouve notamment des occurrences de *développement du son, après ou l'intérieur du son*.

Les verbalisations des deux catégories *évolution / non-évolution* vont être confrontées à leur répartition en fonction des sons des paires entendues. En effet, on s'attend à ce que les sons imitant des sons soufflés, particulièrement synthétiques notamment à cause de leur non-évolution dans le temps, soient plus fréquemment décrits à travers la catégorie *non-évolution*.

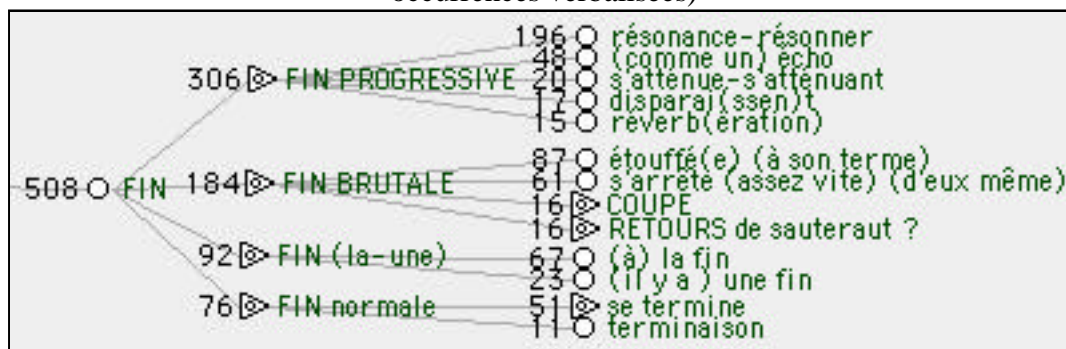
Si les différences de répartition ayant trait aux verbalisations de type *évolution* ainsi que deux de ses principales sous-catégories *vibrer-onduler* et *dégression* ne sont pas significatives ici ( $\chi^2(11) < 15$ ,  $p > .18$ ), celles de type *progression* ou *pas d'évolution* sont significatives (respectivement  $\chi^2(11) = 20$  et 53 et  $p = .04$  et  $2.10^{-7}$ ). En effet, on rencontre plus souvent des verbalisations de *progression* dans les paires dont un des deux sons est le son CFT et on rencontre plus souvent des verbalisations soulignant un caractère non évolutif quand un des sons de la paire est soit le son CLA, soit le son TPT, donc deux des imitations de sons soufflés.

On avait remarqué que l'ensemble des verbalisations sur le milieu des sons est dû plus fréquemment aux non-musiciens (pour 60%). Les différences de répartition des verbalisations appartenant à ses sous-catégories en fonction des groupes musiciens et non-musiciens sont très significatives ( $\chi^2(7) = 78$ ,  $p = 3.10^{-14}$ ). Ceci est particulièrement marqué en ce qui concerne les verbalisations sur le *deuxième temps* du son (dues à 86% aux musiciens) et celles sur les notions de *progression, dégression* et *déroulement* (dues respectivement à 74, 81 et 82% aux non-musiciens). Ceci s'explique peut-être en considérant que les verbalisations de type *progression* et *dégression* sont dominées respectivement par des occurrences de *va vers le haut* et *va vers le bas* qui font référence à la notion de hauteur, *a priori* non pertinente dans notre expérience (vu que les sons étaient égalisés en hauteur), et que les musiciens ont donc verbalisé avec parcimonie.

La fin des sons

La troisième sous-catégorie des verbalisations d'ordre temporel rassemble celles qui ont trait à la *fin* du son. Elles représentent elles aussi un grand nombre de paires (38% de la totalité d'entre elles), ce qui justifie la description plus détaillée que nous allons en faire (figure 6.6).

Figure 6.6. Verbalisations codées sous *parlent du son / fin* (508 paires, sous-catégories rassemblant le plus d'occurrences. Les mots en majuscule représentent des catégories, et ceux en minuscule des occurrences verbalisées)



Les verbalisations se répartissent entre une *fin progressive*, dominée par les occurrences de *résonance-résonner* (196 paires) mais aussi par les notions d'*écho*, d'*atténuation*, de *disparition* et de *réverbération*. Celles-ci s'opposent à la notion de *fin brutale*, dominée par les occurrences de *étouffé*, *s'arrête*, *coupé* et d'un ensemble de verbalisations que l'on a rassemblé en faisant l'hypothèse qu'elles se réfèrent à la perception du *retour du sautereau* qu'on entend dans le son imitant le clavecin (CVC) et dans l'hybride de clavecin et de hautbois (OBC).

On trouve aussi des verbalisations consistant à nommer cette partie du son avec de nombreuses occurrences de *à la fin* ou *(il y a) une fin* par opposition à *il n'y en a pas*. Enfin, un certain nombre de verbalisations désignent la fin du son sans caractériser le fait qu'elle se fasse brutalement ou progressivement. Nous les avons donc rassemblées sous la catégorie *fin normale* principalement due aux occurrences de *se termine* et de *terminaison*.

Ici aussi, on observe des différences de répartition significatives des différentes catégories de verbalisations en fonction des sons contenus dans les paires (tableau 6.3). Ainsi la différence de répartition la plus importante concerne les verbalisations regroupées sous la notion de *retour du sautereau* qui concernent bien principalement des paires dont un des sons est le son CVC pour lequel on perçoit bien un *accoup final*. De même, les verbalisations à propos de *réverbération* ou *réverbérant* concernent principalement les paires qui contiennent le son VBR. On peut noter qu'un son naturel de vibraphone est produit par une percussion sur une lame de métal qui vibre longtemps. Ce type de verbalisation n'apparaît par contre jamais dans les paires qui contiennent le son PNE qui fait pourtant partie des sons de type *pincé-frappé*.

On retrouve le son VBR comme un des plus fréquents dans les paires où la notion de *fin progressive* est la plus souvent verbalisée, mais aussi le son PNO. Ces deux sons se retrouvent aussi parmi les

plus fréquents dans les paires, à propos desquelles les participants parlent plus souvent de *résonance* ou *résonner*. Par contre, c'est le son CRA qui est le plus souvent présent dans les paires à propos desquelles les participants verbalisent le plus d'occurrences d'*étouffé*.

Tableau 6.3. Répartition du nombre de paires décrites par le *son / temporel / fin* en fonction des timbres contenus (au maximum 220 = 11 paires \* 20 sujets contiennent chacun des sons).

Timbre	<i>retour du sautereau</i>	<i>étouffé</i>	<i>fin progressive</i>	<i>résonnance-résonner</i>	<i>réverbération-réverbérant</i>
TPT	2	13	40	24	1
TBN	1	18	42	33	2
CRA	2	<b>30</b> +	49	31	5
CLA	2	13	44	<b>21</b> -	2
CFT	2	<b>8</b> -	51	35	1
VBR	1	14	<b>76</b> +	<b>51</b> +	<b>12</b> +
PNO	1	11	<b>70</b> +	<b>48</b> +	4
PNE	3	11	50	27	<b>0</b> -
CVC	<b>12</b> +	16	44	27	3
OBC	2	16	46	30	4
VBN	2	10	62	40	2
SNO	1	14	<b>36</b> -	25	2
moyenne	2.58	14.50	50.83	32.67	3.17
écart type	3.03	5.63	12.29	9.41	3.13
$\chi^2$	<b>16.38</b>	<b>6.25</b>	<b>11.80</b>	<b>9.87</b>	<b>12.72</b>
p	5.10 <sup>-5</sup>	0.01	6.10 <sup>-4</sup>	2.10 <sup>-3</sup>	4.10 <sup>-4</sup>

avec + (respectivement -) pour les valeurs s'écartant de la moyenne de plus d'un écart type. En gras, différences significatives (valeurs de  $\chi^2$  par rapport à la moyenne attendue).

On peut aussi noter une tendance à verbaliser plus souvent à propos de la *fin* des sons quand un des sons de la paire est une imitation de son pincé-frappé et moins quand les deux sons sont des imitations de soufflé (respectivement dans 49 et 26% de ces paires alors que M=38±6%). La même remarque est vraie à propos des verbalisations sur la notion de *fin progressive*, plus fréquentes quand la paire contient une imitation de son pincé-frappé (32% alors que M=23±6%) et moins souvent si les paires contiennent deux imitations de sons soufflés (9%).

Ces analyses dévoilent une perception très fine et très pertinente des différents sons écoutés. La question de savoir si ces verbalisations dépendent de la compétence musicale des participants se pose ici avec plus de force. L'ensemble des verbalisations sur la fin des sons est dû pour 55% aux musiciens, et les différences de répartition entre les deux groupes de participants en terme de nombre de paires verbalisées dans chacune des sous-catégories sont significatives ( $\chi^2(12)=86$ ,  $p=2.10^{-13}$ ). Ces différences sont particulièrement marquées pour les verbalisations sur le *retour du sautereau*, sur la *résonance* mais aussi sur le verbe *disparaître*, dues aux musiciens pour respectivement 88, 75 et 100% d'entre elles, et pour les verbalisations sur les verbes *s'arrêter* et *s'atténuer*, dues aux non-musiciens pour respectivement 72 et 58% d'entre elles.

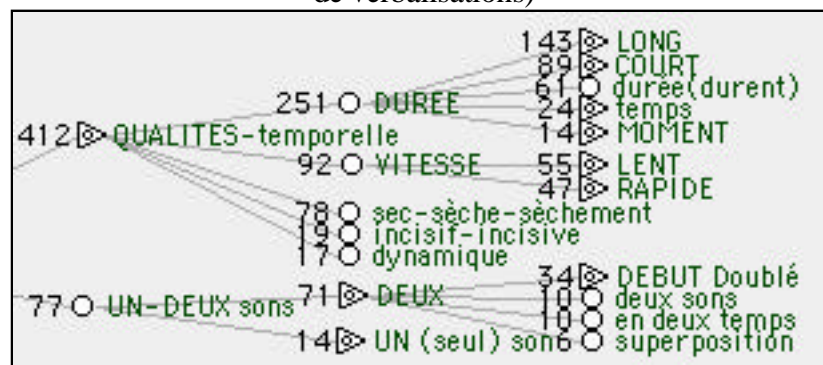


Si on retrouve la notion de *retour du sautereau* et de *résonance* à la fois dans les verbalisations principalement dues aux musiciens et réparties sur quelques sons de façon spécifique, ce n'est pas une généralité. En effet, certaines verbalisations, elles aussi principalement dues aux musiciens, ne sont pas produites de façon spécifique en présence d'un des 12 sons (c'est le cas des verbalisations de *disparaissent*). D'autre part, certaines verbalisations dont les différences de répartition en fonction des sons étaient significatives sont produites également par les musiciens et les non-musiciens (c'est le cas des verbalisations de type *fin progressive*, *réverbération* et *étouffé*). Par contre, aucun des deux types de verbalisations principalement produits par les non-musiciens (*s'arrête* et *s'atténue*), n'est réparti de manière spécifique sur quelques-uns des 12 sons testés.

#### Autres catégories temporelles

Afin de terminer la description de cette catégorie de verbalisations si importante par sa fréquence, nous allons décrire brièvement les autres niveaux codés sous *parle du son / temporel* dans la figure 6.7.

Figure 6.7. Partie des verbalisations codées sous *parlent du son / temporel* (458 paire, sous-catégories rassemblant plus de 10 occurrences, avec en majuscule des catégories, en minuscule des occurrences de verbalisations)



La catégorie notée *qualités-temporelle* est principalement due à deux qualités, celles ayant trait à la *durée* et à la *vitesse*. Mais elle contient aussi de nombreuses occurrences de *sec-sèche* et *incisif-incisive* qui nous semblent caractériser une attaque rapide. Les qualités ayant trait à la *durée* se répartissent principalement dans les sous-catégories de *long* et *court*, elles-mêmes principalement dues respectivement aux occurrences de *long-longue* et *se prolonge* d'une part et de *court(e)* et *bref-brève* d'autre part. On trouve aussi dans cette catégorie des occurrences de *durée* ou du verbe *durer*, ainsi que des verbalisations du type *prend du temps* et des verbalisations sur le *moment* comme *instantané* ou au contraire *tout le long*.

La catégorie *vitesse* se répartit en deux sous-catégories : *lent* et *rapide* qui sont respectivement dues principalement aux occurrences de *lent(e)-lentement* et *progressivement* d'une part, et de *rapide*, *rapidité*, *rapidement* d'autre part. Enfin, la catégorie rassemblant les verbalisations d'un côté double de certains sons est principalement représentée par les occurrences de *doublé-double*, *deux sons* et *en deux temps*. Les verbalisations du type *deux sons* sont considérées comme appartenant à cette catégorie quand elles figurent dans des phrases explicites du type : *le premier son, on dirait deux*

sons, le deuxième, un seul son afin de ne pas confondre ces verbalisations avec le fait que dans chacune des paires que les participants écoutaient on entendait effectivement deux sons différents.

Ces différents types de verbalisations de qualités temporelles sont souvent co-occurentes à d'autres verbalisations de type temporel, comme par exemple les notions de *début*, *milieu* et *fin*. En effet, si l'ensemble des trois catégories de verbalisations appelées *durée*, *vitesse* et *qualités-temporelles* représente 395 paires de sons, seules 106 (soit 27% d'entre elles) n'apparaissent pas dans une des autres catégories temporelles.

La différence de répartition des verbalisations sur la *vitesse* en fonction des différents timbres contenus dans les paires décrites est significative ( $\chi^2(11)=22$ ,  $p=3.10^{-6}$ ). C'est vrai aussi pour les verbalisations de *vitesse/lent* et du caractère double de certains sons, mais pas pour les verbalisations de *vitesse/rapide*. Ceci s'explique par le fait que le son présent le plus souvent dans les paires à propos desquelles les participants verbalisent des qualités de vitesse est le son CFT, que les participants perçoivent avec un départ lent. Les descriptions de *vitesse/rapide* sont souvent dues à la présence de ce même son CFT et semblent donc apparaître par contraste ou comparaison avec ce son. Par contre on retrouve très souvent deux des trois sons hybrides (les sons de VBN et d'OBC) dans les verbalisations sur le caractère *double* des sons ( $\chi^2(11)=52$ ,  $p=2.10^{-7}$ ). Le troisième son hybride, SNO, étant construit comme un mélange de corde frottée et de piano, et ayant le profil temporel de la corde frottée, n'a donc pas ce caractère double contenu par les deux autres hybrides.

#### Lien avec des mesures acoustiques

Sachant que nous possédons un grand nombre de mesures acoustiques effectuées sur ces mêmes sons par Krimphoff et al (1994), il était tentant de vérifier si on pouvait avoir des corrélations significatives entre certaines de ces mesures et certaines verbalisations d'ordre temporel.

Nous avons testé la corrélation entre la répartition de 31 types de verbalisations d'ordre temporel sur chacun des 12 sons contenus dans les paires décrites, en rapport avec la mesure acoustique de leur logarithme du temps de montée de l'enveloppe temporelle, qui est le corrélat acoustique de la première dimension perceptive des sons testés. Seul 3 des 31 types de verbalisations testés avaient une corrélation significative ( $p(9)<.05$ ) avec cette mesure acoustique. Il s'agit des verbalisations sur *la fin* des sons ( $r=-.70$ ,  $p=.01$ ), et des occurrences de *continu* ( $r=-.77$ ,  $p=.004$ ) et *réverbération* ( $r=-.61$ ,  $p=.04$ ). De plus, la répartition des occurrences de *étouffé* sur chacun des 12 sons contenus dans chacune des paires était corrélée à une mesure de flux spectral de manière significative ( $r=-.61$ ,  $p=.05$ ).

Ces liens entre verbalisations et mesures acoustiques sont donc limités.

#### **II.4.2. La forme des sons**

Décrire la forme d'un son ou utiliser des verbalisations d'ordre temporel revient souvent au même. Ce n'est pas un hasard si les acousticiens ont choisi de représenter le déroulement temporel des sons par la forme graphique de l'enveloppe temporelle. Est-ce parce que spontanément un son est considéré comme une forme se déroulant dans le temps que les acousticiens l'ont mesurée et dessinée, ou est-ce l'existence de cette technique qui influence les verbalisations des participants ? Notre étude ne pourra pas trancher, étant donné que la plupart des non-musiciens interrogés étaient des étudiants en physique.

Quand les participants décrivent la forme des sons qu'ils entendent (figure 6.8), ils caractérisent le *trait*, catégorie principalement due aux occurrences de *précis(e)-pas précis* et *défini(e)* et qui concerne 266 paires de sons, soit 20% de l'ensemble des paires et 46% des paires où il est question de la forme des sons. Ensuite, les participants caractérisent le *dessin* dans 215 paires (soit 16% de l'ensemble et 37% des paires où on parle de la forme des sons). Il s'agit ici soit de considération sur la *rondeur* des sons (le terme *rond(e)* est employé dans 106 paires soit dans 8% des paires décrites) soit des verbalisations empruntant au vocabulaire de la *géométrie* (avec en particulier les verbalisations de caractère *horizontal* et *triangulaire*), soit de verbalisations ayant trait à la notion de *pointu*, soit encore d'occurrences de *plat(e)* ou de *un halos*.

Les participants verbalisent aussi à propos de la *taille* des sons qui sont principalement décrits comme *large*, *ample* ou *petit*, la notion de *volume* apparaissant aussi assez souvent. On trouve ensuite la notion de *localisation* qui recouvre en grande partie la notion de *hauteur*, qu'on retrouvera dans la catégorie relevant de l'*Audition*, et qui est principalement due aux verbalisations de *haut*, *bas*, ainsi que des localisations comme *dans l'espace*.

Figure 6.8. Partie des verbalisations codées sous : *parlent du son / forme* (576 paires, sous-catégories rassemblant plus de 10 occurrences, avec en majuscule des catégories, en minuscule des occurrences de verbalisations)

266 ▶ TRAIT	54 ▶ Précis(e)-pas précise	
	24 ▶ Défini(e)-définition	
215 ▶ DESSIN	114 ▶ ROND	106 ○ rond(e)-rondeur
	54 ▶ GEOMETRIE	17 ▶ horizontal-horizontaux
	27 ○ plat(e)	8 ○ triangulaire-triangle
	24 ▶ POINTU	20 ○ pointu(e)
	13 ○ halos (un)	
70 ▶ TAILLE	24 ○ large	
	17 ○ ample-ampleur	
	13 ○ petit	
	11 ○ volume	
69 ▶ LOCALISATION	29 ○ HAUT	
	19 ○ BAS	
	13 ○ dans l'espace	
53 ▶ COMPOSITION	18 ○ structure	
	16 ○ construit-construction	
48 ▶ DENSITE	12 ○ ramassé	
	11 ○ serré-resséré	
	10 ○ serré	
39 ▶ OUVERT-FERME	20 ▶ fermé(e)	
21 ▶ La forme (de son)		
10 ▶ DIRECTION	8 ▶ direction	

Les participants verbalisent enfin la notion de *composition* des sons, principalement grâce à des occurrences de *structure* ou *construit-construction*, et celle de *densité* avec les occurrences de *ramassé*, *serré* et *resserré*. On trouve enfin les notions de *ouvert-fermé* et de *direction* des sons ainsi que des occurrences de *la forme(des sons)*.

Globalement, l'ensemble de ce niveau est produit par les participants des deux groupes (57% par les musiciens). Pourtant, si on regarde la répartition des verbalisations des sous-catégories qui recueillent plus de 10 occurrences de verbalisations, les différences entre musiciens et non-musiciens sont très significatives ( $\chi^2(17)=102$ ,  $p=3.10^{-14}$ ). Ces différences sont pourtant principalement dues à des catégories qui rassemblent un faible nombre d'occurrences (inférieur à 54 paires). En effet, les musiciens verbalisent plus à propos de la *précision* du trait, et des sons considérés comme *ouverts ou fermés* (respectivement 98 et 85% de ces verbalisations leur sont dues), les non-musiciens par contre sont plus nombreux à verbaliser la notion de *composition* des sons et le caractère *bas* de ceux-ci (respectivement 64 et 89% des verbalisations de ce type leurs sont dues).

Certaines différences de répartitions des sous-catégories en fonction des sons contenus dans les paires sont significatives, en particulier pour 3 catégories de verbalisations qui sont liées : celles concernant le *dessin*, *dessin/rond* et *dessin/plat*. En effet, quand les participants parlent de son *rond*, c'est plus souvent pour décrire une paire contenant un des sons TBN, VBR ou CVC, et moins souvent pour décrire un des sons CRA, CFT ou VBN ( $\chi^2(11)=28$ ,  $p=.004$ ). Ces répartitions sont difficiles à interpréter. En effet, la rondeur des sons CVC et VBR (et la non rondeur du son CRA) pourrait se comprendre par le côté rebondi de l'enveloppe temporelle de ces sons, alors que la rondeur des sons TBN et VBR (et la non rondeur des sons CRA et CFT) pourrait se comprendre en terme de composantes spectrales qui les rendent "doux" (avec des harmoniques basses). Par contre la non rondeur du son VBN doit recouvrir encore autre chose, ce son a en effet à la fois des harmoniques basses et une enveloppe temporelle de type rebondie, mais il est aussi considéré souvent comme "raté" et "désagréable", ce qui n'est sans doute pas compatible avec la notion de rondeur.

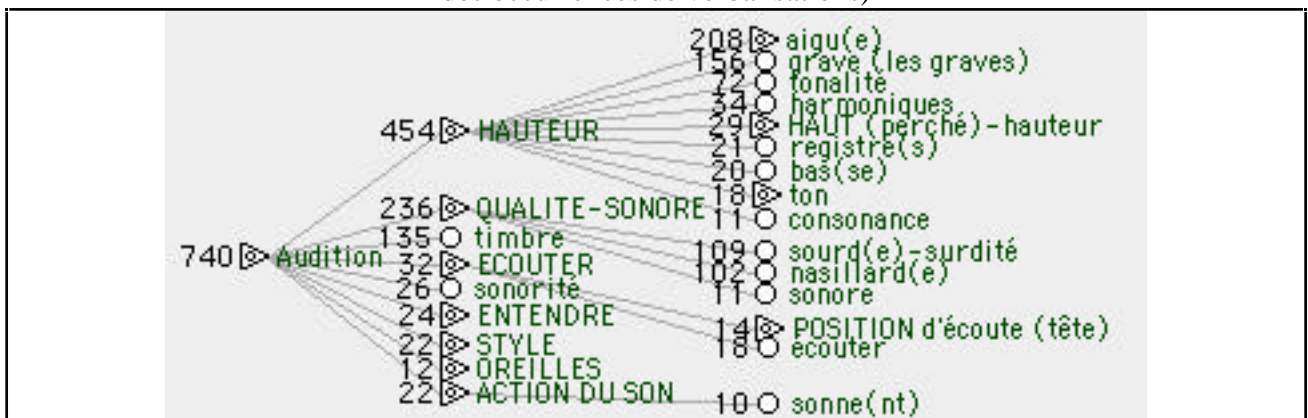
Pour ce qui est des différences concernant les occurrences de *plat*, elles sont plus souvent produites quand un des sons de la paire est le son CLA qui est un des sons les plus continus et stables temporellement parlant ( $\chi^2(11)=24$ ,  $p=.01$ ), mais aussi quand un des sons de la paire est le son VBR, ce qui paraît par contre assez inexplicable compte tenu de sa résonance. Ces verbalisations sont moins fréquentes quand un des sons de la paire est l'hybride SNO, ce qui peut être dû à son enveloppe temporelle imitant une corde frottée donc plutôt progressive. Ces différences de répartition se retrouvent au niveau supérieur des descriptions générales de type *dessin* ( $\chi^2(11)=23$ ,  $p=.02$ ) où on retrouve la répartition obtenue avec les verbalisations de *rondeur*. Par contre, les verbalisations de *direction* des sons sont beaucoup plus fréquentes quand un des sons de la paire est le son CFT ( $\chi^2(11)=22$ ,  $p=.02$ ), ce qui peut se comprendre par la direction de l'archet frottant les cordes qui produit un son dont l'évolution est très progressive.

### II.4.3. Vocabulaire relevant des 5 sens

Nous avons rassemblé dans cette partie l'ensemble des qualités attribuées aux sons et qui semblent relever d'un des 4 domaines sensoriels : audition, toucher, vision et goût. Seul l'odorat ne semble pas mis à contribution et nous verrons que les 4 sens cités ne le sont pas avec la même fréquence (figure 6.9). En effet, alors que l'ensemble des verbalisations rassemblées sous la catégorie *sensation-perception* concernent 944 paires de sons (soit 75% des paires), celles faisant appel au domaine de *l'audition* représentent 740 paires de sons (soit 56% des paires et 78% de celles de cette catégorie). Ceci est à nuancer, sachant que parmi les verbalisations sur la source sonore certaines concernent des instruments de musique et peuvent donc elles aussi être considérées comme relevant directement du domaine du sonore.

Les participants décrivent alors souvent la notion de *hauteur* (pour 454 paires soit dans 34% des paires et 61% de celles de la catégorie *audition*) qui est *a priori* non pertinente dans le cadre de notre expérience puisque les sons proposés sont tous joués sur la même note. Néanmoins, la fréquence de ce genre de considérations nous permet d'affirmer que les notions de hauteur tonale et de hauteur spectrale sont souvent confondues par les participants. La notion de *hauteur* relève soit spécifiquement du domaine musical avec *aigu* et *grave*, soit elle est empruntée à la notion de localisation avec *haut* et *bas*. Participent aussi à cette notion de hauteur des occurrences de termes plus abstraits ou plus techniques comme *harmoniques*, *registre*, *ton* et *consonance*, qui ne permettent pas d'envisager une confusion avec la notion de hauteur tonale.

Figure 6.9. Partie des verbalisations codées sous *parlent du son / sensation-perception* (944 paires) (sous-catégories rassemblant plus de 10 occurrences, avec en majuscule des catégories, en minuscule des occurrences de verbalisations)



		109	▷	TEXTURE	76	○	doux-douce-douceur
		52	▷	TEMPERATURE	10	○	piqué(e)
264	▷	41	▷	PLEIN-CREUX	29	○	chaud(e)
		40	▷	DUR-MOU	16	○	froid(e)
		25	▷	ENVELOPPE	16	○	creux-creuse
		19	▷	POIDS	16	○	plein(e)
		124	▷	LUMINEUX	20	○	mou-molle
244	▷	72	▷	COULEUR	10	○	enrobé(e)s
		51	▷	SOMBRE	10	○	lourd(e)
		12	○	acide	88	○	clair(e)
30	▷	7	○	fade	20	○	brillant(e)-brillance
					70	○	couleur
					17	○	bleu
					16	○	mat(e)
13	▷						

On trouve ensuite des *qualités sonores*, principalement celles de *sourd*, *nasillard* et *sonore*. Il peut paraître étonnant d'utiliser *sourd*, qui a pris les sens d'"idiot" et "qui n'entend pas", pour décrire un son que l'on entend. De plus, le mot *nasillard* vient sans doute de l'effet produit quand on parle du nez. Ces deux termes ne relèvent donc peut-être pas directement du domaine de l'audition...

On peut ensuite observer de nombreuses occurrences de *timbre* et *sonorité* ainsi que d'*écouter* et *entendre*. Schaeffer avait précisé leurs différences fondamentales dans son traité des objets musicaux, (Schaeffer, 1966, p104). On peut remarquer qu'une partie des occurrences d'*écouter* concerne des remarques sur la position d'écoute, qui sont dues à un seul des participants qui les associait aux notions de aigu et grave : les sons aigus étant pour ce sujet entendus dans le *haut de la tête* et les sons graves *dans le bas du corps*. Enfin, quelques unes de ces verbalisations concernent des descriptions de *style musical* (*campagnard, moderne, musique classique...*), des allusions aux *oreilles* (*sonne à l'oreille, saute dans l'oreille, fait mal aux oreilles...*), ou des descriptions d'*actions du son* (comme *sonner, imiter, couiner...*)

Ensuite, on trouve des verbalisations relevant du domaine du *toucher* produites dans 264 paires (soit 20% des paires et 28% de celles codées *sensation-perception*). Nous avons rassemblé ici des notions aussi différentes que celles de *texture* (principalement due aux occurrences de *doux* et *piqué*, mais aussi due à *grain-granuleux* (8 paires), et puis *feutré, cotonneux, velouté* (4 paires)), celle de *température* (principalement due à *chaud* et *froid*), ce qu'on aurait pu appeler la "ductilité" (principalement dû à *mou* mais regroupant aussi les notions de *souple, malléable, dur, flasque* ou *ferme*), la notion d'*enveloppe* (principalement due à *enrobé* mais aussi à *enveloppé* ou *calfeutré*) et la notion de *poids* (principalement due à *lourd* mais aussi à *léger, chargé, costaud, fluet...*). Enfin, la seule sous-catégorie non citée dans la figure (parce que regroupant 9 occurrences), se rapporte à la notion de *liquide* rassemblant les occurrences de *coulant, fluide, huileux* et *liquide*.

Ensuite seulement, apparaissent des notions appartenant au domaine de la *vision* avec 244 paires de sons concernées (soit 18% des paires et 26% de celles codées *sensation-perception*). Il faut noter ici que nous n'avons pas considéré que les verbalisations décrivant la *forme* des sons relevaient exclusivement du domaine visuel, bien que le domaine visuel soit souvent considéré comme prépondérant lorsqu'on parle de forme, de trait, de dessin, de taille, de localisations... Ces notions nous paraissent en effet dépasser le seul cadre visuel, nous avons donc rassemblé dans cette partie uniquement les verbalisations qui relèvent sans aucun doute du domaine de la vision, c'est à dire

principalement les notions de lumière avec les verbalisations rassemblées sous la catégorie *lumineux* (principalement dues aux occurrences de *clair* et *brillant*) et celles rassemblées sous la catégorie *sombre* (principalement dues aux occurrences de *mat* mais aussi de *terne*, *tamisé* et *sombre...*). Enfin apparaît aussi la notion de *couleur*, la couleur *bleu* étant citée la plus souvent mais au total 9 couleurs différentes sont citées par les participants, non sans rappeler le phénomène de synesthésie et en particulier d'audition colorée (Karwoski & Odbert, 1938 ; Odbert et al., 1942 ; Pecjak, 1970).

Enfin, quelques verbalisations font référence au domaine du *goût* (30 paires en tout), et regroupent les occurrences de *acide*, *aigre*, *fade*, *âpre*, *sucré*, *suave*, *douceux...*

#### Répartition en fonction des participants

Si on considère la répartition des verbalisations relevant des 4 différents sens cités en fonction de la compétence musicale des participants, il n'y a pas de différence significative ( $\chi^2=5.6$ ,  $p=.33$ ,  $df=3$ ), même si on peut observer une légère tendance à ce que les musiciens soient plus nombreux à faire appel au *goût* (53%) et les non-musiciens plus nombreux à faire appel à la *vision* (56%). Pourtant, si on prend en compte l'ensemble des niveaux qu'on a rassemblé dans la figure précédente (toutes les catégories et les occurrences concernant plus de 10 paires différentes), la différence entre musiciens et non-musiciens devient très significative ( $\chi^2(54)=353$ ,  $p=6.10^{-46}$ ). On peut donc noter que chacune des sous-catégories rassemblant les verbalisations qui relèvent des différents sens sont très spécifiques d'un groupe de participants, voire même souvent de quelques participants seulement.

#### *Audition*

Les verbalisations relevant du domaine de l'audition concernent tous les participants et semblent également réparties entre les musiciens et les non-musiciens (51% dues aux musiciens). Ainsi, même si aucune des catégories rassemblant plus de 30 occurrences ne présente de différence de répartition entre musiciens et non-musiciens, on observe des différences significatives pour *entendre* (produit à 79% par les musiciens) et pour les *styles musicaux* (produits à 91% par les non-musiciens). De plus, parmi les verbalisations de *hauteur*, les occurrences de *haut* sont produites à 72% par des musiciens, alors que celles de *bas* le sont à 85% par des non-musiciens. Ceci souligne la non symétrie de ces deux termes pourtant souvent considérés comme antonymes. De plus, les non-musiciens sont les principaux locuteurs des occurrences de *ton* et de *tonalité* (respectivement pour 83 et 76% d'entre elles). Enfin les verbalisations de *positions d'écoutes* sont dues à une seule personne, musicienne amateur. Il est intéressant de remarquer que les occurrences de *harmonique* et *registres*, pourtant des mots techniques du vocabulaire musical, sont principalement employées par des participants non-musiciens. (Il faut dire que le participant qui utilise le plus le mot *harmonique* est un thésard en physique des liquides pour qui ce terme est donc courant.).

#### *Toucher*

De même, alors qu'au niveau de l'ensemble des verbalisations relevant du sens du *toucher* on trouve 50% des verbalisations dues aux musiciens, ils sont beaucoup plus nombreux à verbaliser la notion de

son *plein* ou *creux* (pour respectivement 73 et 100% des paires concernées). Au contraire, les verbalisations de sons *froids* ou *enrobés* sont principalement dues aux non-musiciens (qui ont produit respectivement 94 et 100% de ces verbalisations). En fait la plupart de ces verbalisations est due à un très petit nombre de participants. En effet, si 18 participants (soit 90% d'entre eux) produisent au moins une fois des verbalisations relevant du domaine du toucher, seule la catégorie de *texture* concerne plus de 50 % des participants. Il est en particulier intéressant de remarquer que même la qualité *doux*, qui est souvent considérée comme adéquate pour parler des sons, est produite principalement par un participant (pour 29% des occurrences). Il nous paraît aussi important de remarquer que sur les 16 sous-catégories de verbalisations relevant du domaine du toucher et étudiées ici, 6 d'entre elles sont principalement dues au participant S11.am3.

#### *Vision*

Au contraire des verbalisations se référant au domaine de l'*audition* et du *toucher*, l'ensemble des verbalisations se référant strictement au domaine de la *vision* semble produit un peu plus par les non-musiciens qui sont locuteurs de 56% de ce type de verbalisations. La répartition musiciens/non-musiciens semble redevenir équitable en ce qui concerne les verbalisations sur le côté *lumineux* ou *sombre* des sons. Par contre, parmi les verbalisations de couleur des sons, aucun musicien ne produit de verbalisation concernant la qualité *bleue* d'un son, ce qui est pourtant observé dans 17 paires de sons. Au contraire, ils sont responsables à 94% des verbalisations sur le caractère *mat* des sons. De même que pour le *toucher*, même si tous les participants produisent au moins une fois des verbalisations se référant au domaine de la *vision*, seules les sous-catégories correspondant aux verbalisations du caractère *lumineux* ou *sombre* et à la qualité *claire* des sons sont dues à plus de 50% des participants.

#### *Goût*

Les verbalisations de qualités relevant du domaine du goût concernent un faible nombre de paires (30 au total) dont 53% sont produites par des participants musiciens. Ce type de verbalisation concerne pourtant 11 participants soit un peu plus de la moitié d'entre eux. Les deux qualités qui regroupent le plus d'occurrences de verbalisations se référant au *goût* sont dues pour plus de 50% à un seul participant, musicien (S4.am1) pour *acide* et non-musicien (S13.am3) pour *fade*. On peut considérer ici que ces différences ne sont pas dues à la compétence musicale des participants mais plutôt à des différences individuelles. On peut tout de même noter que le participant qui a le plus employé le mot *acide* a aussi donné des explications sur l'emploi qu'il en faisait : *ce que j'appelle acide, c'est un peu pincé*. De plus sur les 6 adjectifs dont l'usage a été défini par leur locuteur (verbalisations codées dans *parle du participant / explication / sens d'un mot*), 2 d'entre eux sont du domaine du goût : *acide*, précédemment cité, et *aigre* défini par *aigre, ce serait l'équivalent pour la nourriture*.

### Répartition en fonction des sons

#### *Audition*



Les différences de répartition de certaines des verbalisations relevant du domaine de l'audition sont réparties de façon significativement différente entre les sons et, semble-t-il, en liaison avec la mesure de certains paramètres acoustiques d'ordre spectraux.

En effet, on remarque une tendance à verbaliser plus à propos de *hauteur* quand un des sons de la paire est soit le son CVC, soit le son d'OBC (avec respectivement 88 et 100 paires concernées alors que  $M=75\pm 10$ ). Or ces deux sons sont significativement plus souvent présents dans les paires à propos desquelles les participants verbalisent la notion de *aigu* ( $\chi^2(11)=36, p=10^{-4}$ ) et sont ceux pour lesquels la mesure du centre de gravité spectral est la plus élevée (McAdams et al., 1995). Par contre les occurrences de *grave* ne sont pas réparties de façon significativement différente entre les sons, bien qu'on observe une tendance à parler plus souvent de *grave* quand un des sons de la paire est le son CLA (qui possède le centre de gravité spectral le plus bas du groupe) mais aussi quand c'est le son CVC qui pourtant est un des plus souvent considéré comme *aigu* (respectivement dans 35 et 33 paires alors que  $M=26\pm 6$ ). On peut donc supposer que la présence du son CVC induit à la fois l'utilisation de *aigu* pour le caractériser et l'utilisation de *grave* pour parler des sons l'accompagnant dans la paire. Ceci peut être une explication du caractère non significatif de la différence de répartition de ce type de verbalisations entre les sons.

On peut remarquer, par contre, que ni les occurrences de *haut*, ni celles de *bas* ne présentent de différences significatives de répartition sur les différents sons. Par contre, les occurrences de *sourd* et de *nasillard* sont toutes deux réparties de manière significativement différente sur les sons présentés (respectivement  $\chi^2(11)=32$  et  $34, p=7.10^{-4}$  et  $4.10^{-4}$ ). On retrouve d'ailleurs un effet semblable à celui observé pour *aigu* et *grave* : le son CVC est le plus souvent présent dans les paires à propos desquelles les participants parlent de sons *sourds* mais aussi de sons *nasillard*. L'autre son qui induit plus de verbalisations de *sourd* est le son CLA, ce qui semble être pour son propre compte. En ce qui concerne *nasillard*, cette qualité est plus souvent verbalisée quand un des sons de la paire est le son CVC mais aussi quand c'est le son OBC ou de TPT. Elle est moins souvent verbalisée quand un des sons de la paire est le son TBN, de CFT ou de VBN. On peut noter que si les trois sons CVC, OBC et TPT ont tous trois des centres de gravité spectrale élevés, seuls les sons TBN et VBN ont des centres de gravité spectrale bas, ce qui n'est pas le cas du son CFT.

#### *Toucher*

Ici, seules deux différences de répartition se sont avérées significatives, celles concernant les verbalisations de *texture* et en particulier les occurrences de *doux* (respectivement  $\chi^2(11)=22$  et  $30, p=.02$  et  $.001$ ). Une partie de ces résultats est sans doute due au faible nombre d'occurrences des catégories considérées. Ainsi, on parle plus de *texture* si un des sons de la paire est le son VBR ou le son OBC et moins si c'est le son VBN. On retrouve cette répartition au niveau en dessous quand il s'agit des différences de verbalisations de la qualité *doux*, dont on parle moins quand le son TBN est présent. On observe de plus une légère tendance à parler plus de sons *rugueux* (ou *abrasif, aspérités, du grain, piquant, racloux*) quand un des sons de la paire est le son CFT ou le son TBN (pour 7 paires alors que  $M=4\pm 2$ ). Le son VBR semble par contre faire l'unanimité en temps que son *doux*.

### *Vision*

En ce qui concerne les verbalisations relevant du domaine de la *vision*, les différences de répartition sur l'ensemble des sons des catégories *vision*, *lumineux*, *sombre*, *couleur* et des occurrences de *clair*, *brillant* et *mat* se sont avérées non significatives. Pourtant, les verbalisations de la qualité *bleue* sont produites plus souvent quand un des sons de la paire est le son VBR ou le son OBC et moins souvent quand ce sont les sons TBN et PNE ( $\chi^2(11)=22$ ,  $p=.02$ ). On peut remarquer que ce type de répartitions rappelle beaucoup celles faites à propos de *doux*. On retrouve le même genre de résultat si on ajoute à *bleu* les autres couleurs froides citées, c'est à dire *vert* et *violet*, ce type de verbalisations devient même encore plus spécifique du timbre VBR ( $\chi^2(11)=28$ ,  $p=.003$ ). Par contre, la répartition des verbalisations des couleurs chaudes (regroupant les occurrences de *rouge*, *orange* et *jaune*) ne présente pas de différences significatives, même si on peut noter la tendance à parler plus souvent de *rouge* en présence du son TPT (pour 5 paires alors que  $M=2.0\pm 1.6$ ). Cette "audition colorée" semble permettre de différencier les sons, qui ne sont pas considérés comme *bleu*, *jaune* ou *rouge* au hasard mais de manière reproductible par les locuteurs. Il faut tout de même rappeler que les verbalisations de noms de couleurs ne concernent que 4 des participants et principalement le participant S16.néo.

### *Goût*

Aucune des différences de répartition des verbalisations relevant du domaine du *goût* en fonction des sons présents ne sont significatives ( $\chi^2(11)<9$ ,  $p>.6$ ). Pourtant on observe la tendance de produire plus souvent la qualité *acide* quand un des sons de la paire est le son SNO (4 fois alors que  $M=2\pm 1$ ). Ce son étant moins souvent décrit que les autres et très rarement de façon spécifique. Cette tendance, bien que de modeste importance, mérite d'être notée.

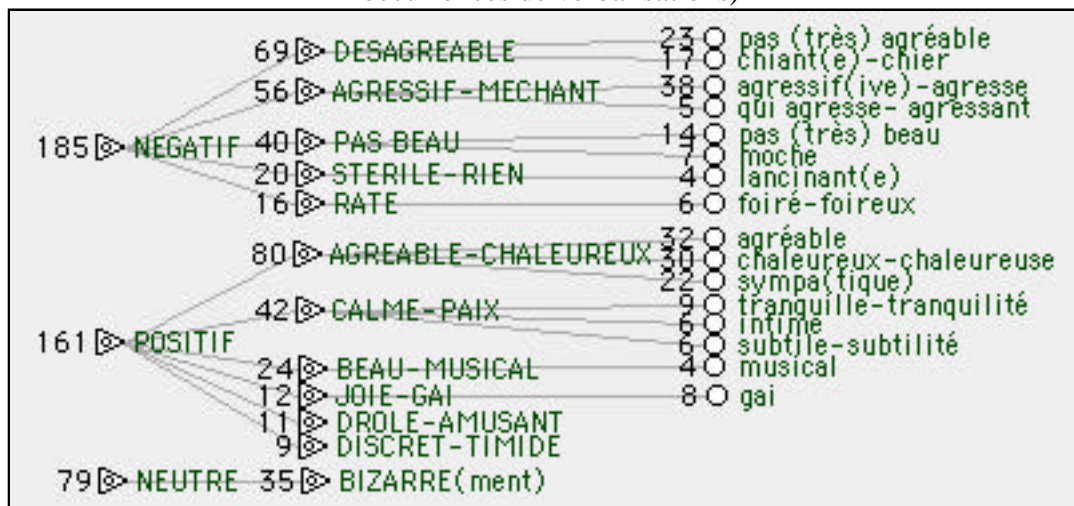
#### **II.4.4. Émotions**

Une autre façon de parler des sons est de décrire les émotions que nous y associons. Ces verbalisations sont présentes dans 331 paires (soit 25% des paires). Il s'agit des effets produits par les sons, *négatifs* (comme *désagréable*, *agressif*) ou *positifs* (comme *agréable*, *chaleureux*, *sympathique*) mais aussi des jugements de valeur sur eux, ici aussi soit *négatifs* (*pas beau*, *stérile*, *raté* ou *pas drôle*) soit *positifs* (*calme*, *beau*, *musical*, *gai*, *drôle*, *amusant*...). On trouve aussi quelques verbalisations ayant un caractère *neutre* comme par exemple *bizarre* (figure 6.10).

Ce qui frappe surtout dans cette catégorie de verbalisations c'est son extrême diversité. Nous avons en effet eu du mal à regrouper l'ensemble des verbalisations produites par les participants dans des catégories, tellement elles étaient diverses et rassemblaient très peu d'occurrences à chaque fois. Ainsi, par exemple, nous avons regroupé sous "*émotion-valeur/négatif*" 83 locutions différentes, et sous "*émotion-valeur/positif*" 61 locutions différentes. La totalité de la catégorie "*émotion-valeur*" rassemblant 183 locutions différentes (pour 331 paires concernées au total) de langage plus ou moins châtié, allant d'expressions comme *subtile*, *reposant*, *quiétude*, *paisible* à *chiant*, *déprimant*, *horripilant*, *insupportable*, *pénible*... On peut déjà remarquer une tendance à verbaliser plus souvent

des émotions désagréables que positives (respectivement 185 et 161 paires, soit 56% et 49% de celles de cette catégorie). Ceci peut être dû au caractère synthétique des sons qui ont par exemple amené les participants à les décrire comme des *sonnettes* ou des *alarmes*, mais aussi au caractère extrêmement répétitif de l'expérience.

Figure 6.10. Partie des verbalisations codées sous *parlent du son / émotion-valeur* (331 paires, sous-catégories rassemblant le plus d'occurrences, avec en majuscule des catégories, en minuscule des occurrences de verbalisations)



Bien que les participants aient tous au moins une fois produit des verbalisations de type *émotion-valeur*, ces verbalisations sont principalement dues aux non-musiciens (pour 73% d'entre elles) et de façon générale chacune des sous-catégories est due à très peu de participants différents (un seul participant peut facilement être responsable de plus de 40% des occurrences d'une sous-catégorie). Pourtant, les musiciens (et principalement un participant) sont plus nombreux à produire des verbalisations de type "neutre". C'est particulièrement le cas pour les occurrences de *bizarre*. Par contre, on ne trouve aucun musicien parmi les locuteurs de *stérile-rien*, *pas drôle*, *joie-gai* ni des occurrences de *chiant* qui sont dues quasi-exclusivement à un seul participant.

On peut noter une tendance ( $\chi^2(11)=19, p=.06$ ) à produire plus souvent ce type de verbalisations quand un des sons de la paire est un des 3 sons CRA, CFT et VBN. On verra par la suite que le son CRA fait partie de ceux qui ont été perçus comme les plus *désagréables*, le son VBN comme *bizarre* et *raté* et le son CFT comme un des plus *chaleureux* (avec le son VBR).

L'ensemble des verbalisations considérées comme *négatif* se répartit de façon significativement différente sur l'ensemble des sons testés ( $\chi^2(11)=26, p=.006$ ). En effet, les participants produisent plus de verbalisations de ce type quand un des sons de la paire est un des sons TPT, CRA ou VBN et moins souvent quand c'est le son PNO. On retrouve des différences significatives de répartition entre les sons dans la plupart des sous-catégories de *négatif*, et c'est souvent à propos des mêmes sons. Par exemple, les sons TPT et VBN sont plus souvent présents quand on parle de *désagréable* ( $\chi^2(11)=30, p=.001$ ), les sons SNO et PNE étant moins souvent présents. Les verbalisations de *agressif-méchant* ne sont pas réparties de façon significativement différente. Par contre, celles de *pas beau* font ressortir

surtout les sons CRA, CLA et VBN ( $\chi^2=26, p=.007, dl=11$ ). Pour ce qui est de *stérile*, c'est au tour des sons CRA et TPT, les sons PNO, VBR et PNE n'étant jamais cités à propos de cette catégorie de verbalisations ( $\chi^2(11)=25, p=.01$ ). Enfin les participants parlent plus souvent de son *raté* quand un des sons de la paire est le son VBN ( $\chi^2(11)=22, p=.02$ ).

On peut remarquer que dans l'ensemble des verbalisations de type *émotion-valeur/négatif*, les timbres à propos desquels les participants produisent le plus ce type de verbalisations sont toujours soit des sons imitant une production de type soufflé (sons TPT, CRA ou CLA mais pas le TBN : imitation de trombone) soit le son hybride VBN. Les sons les moins souvent présents dans les paires où les participants verbalisent des *émotion-valeur/négatif* sont tous des imitations de sons produits par frappement (PNO surtout mais aussi PNE et VBR). Ceci rejoint notre remarque sur le caractère synthétique des sons imitant des sons soufflés.

Les verbalisations de type *positif* ne sont pas réparties de façon significativement différente entre les sons ( $\chi^2(11)=9, p=0.6$ ), même si on peut observer une tendance à les rencontrer plus souvent dans des paires où le son VBR ou TPT est présent et moins souvent si c'est le son SNO (respectivement dans 33, 34 et 19 des paires contenant ces sons alors que  $M=27\pm 5$ ). Enfin, même si les verbalisations de type *neutre* ne se répartissent pas de manière significativement différentes entre les différents sons ( $\chi^2(11)=15, p=0.17$ ), les verbalisations de types *bizarre* sont significativement plus souvent produites quand un des sons de la paire est le son VBN (hybride de *vibrone*) ( $\chi^2(11)=36, p=2.10^{-4}$ ).

Les verbalisations rassemblant les jugements de valeur sur les sons ainsi que les émotions qu'ils produisent semblent donc très spécifiques à la fois des participants qui les produisent et des sons à propos desquelles elles sont verbalisées.

#### II.4.5. Autres manières de décrire les sons

Pour clore l'analyse de la partie des verbalisations qui parle des sons, nous allons analyser brièvement les autres sous-catégories qui rassemblent moins d'occurrences et que nous avons codées *qualité liée à la source, général, énergie et intensité, nom de son et imitation*.

##### Qualités liées à la source

Ces verbalisations sont intermédiaires entre celles où les participants parlent du son et celles où ils parlent de la source. En effet, c'est bien du son que les participants parlent mais en lui prêtant des qualités directement issues de caractéristiques de la source sonore, décrivant par exemple un son *métallique, synthétique, artificiel, électronique, cristallin, naturel, cuivré, percussif* ou *appuyé*... Même si 95% des participants ont au moins une fois produit des verbalisations de ce type, les musiciens en ont globalement produit plus que les non-musiciens (63% des occurrences leur sont dues dont 17% pour le seul participant S3.pro). De plus, les sous-catégories ont en général été produites par moins de 6 locuteurs différents, à part *métallique*, produit par plus de la moitié des participants.

Les seules différences significatives de répartition de ces verbalisations sur l'ensemble des sons concernent *synthétique* et *cuivré*. En effet, les participants parlent plus souvent de sons *synthétiques* quand un des sons de la paire est le son OBC et moins souvent si c'est le son VBR ou CFT ( $\chi^2(11)=20, p=.04$ ), et les participants parlent plus souvent de son *cuivré* si un des sons de la paire est le son TPT ou TBN ( $\chi^2(11)=26, p=.007$ ). Par contre, les participants ont tendance à parler plus souvent de *métallique* quand le son CVC est présent (dans 29 paires alors que  $M=18\pm5$ ). De même, ils ont tendance à parler plus souvent de *fabriqué* si un des sons CRA, VBN, OBC et SNO est présent et moins souvent si c'est le son VBR ou CFT (en moyenne respectivement dans 19 et 6 paires alors que  $M=13\pm5$ ). Tous les sons hybrides sont donc ici considérés plus souvent comme *fabriqués*.

Les participants ont aussi tendance à parler plus souvent de sons *électroniques* si un des sons CRA ou CLA est présent (respectivement dans 7 et 6 paires alors que  $M=3\pm2$ ), de son *cristallin* si le son VBR est présent (dans 12 paires alors que  $M=4\pm3$ ), de son *naturel* si un des sons VBR, PNE et TPT est présent (dans 7, 6 et 6 paires alors que  $M=4\pm2$ ) et de son *appuyé* si le son SNO est présent (dans 6 paires alors que  $M=2\pm1$ ). Toutes ces tendances semblent être en accord avec le mode de production imité lors de la synthèse de ces sons.

#### Divers

Les verbalisations codées sous *le son/général* sont celles que nous n'avons pas pu mettre ailleurs. En effet, il s'agissait soit de qualité des sons mais qui ne relevaient ni du domaine temporel, ni de la forme, ni d'un des 4 sens cités par les participants, ni d'ordre émotionnel. On trouve par exemple la notion de *pur-libre* principalement due aux occurrences de *pur*, *limpide*, celle de *riche-pauvre*, principalement due aux occurrences de *riche*, mais aussi la notion de *propre-sale*. Ensuite, on a rassemblé ici toutes les occurrences de dénominations générales que nous avons codées sous *type-nature* comme *le type de son*, *l'impression générale*, et des verbalisations que nous avons codées comme *spécifiques* parce qu'elles montraient qu'un des sons avait été reconnu comme particulièrement remarquable, par exemple *j'ai bien reconnu celui qui me gêne*.

On observe que les verbalisations de type *pur-libre* sont plus fréquentes quand un des sons de la paire est le son VBR ( $\chi^2(11)=30, p=.002$ ). Les différences concernant les verbalisations de type *riche-pauvre*, pratiquement uniquement dues aux occurrences de *riche* ou *richesse* ne sont pas significatives mais montrent tout de même une tendance à parler plus souvent de *richesse* quand un des sons de la paire est le son CRA qui a un des centres de gravité spectraux les plus élevées (Krimphoff et al., 1994)). Enfin, les verbalisations codées sous *spécifiques* ne sont pas réparties de manière significativement différentes entre les sons, même si on observe une tendance à produire plus souvent ce type de verbalisations quand un des sons de la paire est le son VBN (dans 6 paires alors que  $M=3\pm1$ ). On peut rappeler que ce son est aussi le plus souvent présent quand les participants parlent de son *bizarre* ou de son *raté*.

#### Énergie, Intensité

On trouve aussi 149 paires à propos desquelles les participants parlent de *l'intensité* des sons, principalement à travers les occurrences de *fort*, *lointain*, *présent* et *intensité* et 75 paires où ils parlent d'*énergie*, principalement à travers les occurrences de *déterminé* et *vif-vive*. Les paires contenant deux imitations de sons soufflés sont plus souvent décrites avec des verbalisations de type *intensité* (26% de ce type de paires alors que  $M=11\pm 5\%$ ), en particulier la paire TPT-CRA pour laquelle 11 des 20 participants ont spontanément verbalisé l'impression que le son TPT est *plus fort* que le son CRA, ou que le son CRA paraît *plus lointain* que le son TPT. Les autres paires de sons qui ont inspiré des verbalisations sur *l'intensité* de ces sons ne l'ont fait que pour 1 à 5 participants au maximum. On peut remarquer que sur les 9 paires rassemblant les occurrences de plus de 4 participants différents, 5 d'entre elles (soit 55%) contiennent le son CRA imitant le cor anglais et souvent considéré comme *faible* ou *lointain*, et 4 contiennent le son TPT, plutôt considéré comme *fort*.

#### Noms de sons et imitation

Enfin, nous ne pouvions terminer cette partie sans mentionner les noms que les participants emploient pour désigner les sons. En effet, de même que pour le domaine olfactif, le domaine sonore est peu lexicalisé (David, 1997). On retrouve pourtant ici les noms de *note*, *bruit*, *son*, *cri*, *trémolo*, *gargouillis* et *cacophonie*, mais aussi des occurrences de *frappement*, *frottement*, *pincement*, *bâillement*... Ceci rejoint les remarques faites par Frédérique Guyot sur un corpus de bruits domestiques (Guyot et al., 1998). Elle observait en effet que les noms de sons produits par ses participants relevaient tous d'adverbes en "-ment" obtenus à partir de verbes d'action permettant de produire ces sons. Les participants ont aussi tendance, quand ils veulent parler des sons, à les imiter à défaut de pouvoir les décrire ou les nommer. Ceci avait déjà été noté par Wright (1971). Pourtant dans le cadre de notre expérience, ce type de verbalisations n'est présent que dans 2% de celles-ci.

#### II.5. Décrire une source sonore

Nous avons vu que dans 60 % des descriptions produites, les participants décrivaient une source sonore possible des sons qu'ils entendaient. Le passage de la description du son lui-même à celle de la source du son se fait souvent grâce au processus de la métonymie. Le son et sa source semblent à ce point confondus dans la représentation des participants que ceux-ci mélangent ces deux types de descriptions. Un exemple typique de l'alternance et du mélange de ces deux modes de descriptions est donné dans l'essai 29 du participant S2.pro, déjà cité §3.1. :

*"Deux instruments. Deux sons superposés pour le premier : une percussion et un cuivre, comme peut-être un son de percussion, le premier prolongé par un long tube, et une anche double, un son assez nasal pour le deuxième."*

On voit que sont mis ici sur le même plan :

*-un son assez nasal*, décrivant une qualité sonore du son entendu, *un son de percussion* qui parle toujours du son mais en décrivant sa provenance,

-une *anche double* où le procédé de métonymie permet de désigner le son directement par l'objet qui a servi à le produire, ainsi que

- des mélanges comme *un son de percussion prolongé par un long tube* (on imagine l'hybride constitué par un son prolongé par un tube !) qui décrit sans doute le son produit par un vibraphone qui est un instrument à percussion possédant des tubes sous les lames qui sont frappées ce qui permet de prolonger les sons émis.

On peut observer la présence de ces glissements de sens entre son et source assez souvent chez les locuteurs, par exemple avec *une clarinette très douce, un piano assez acide, la flûte me paraît plus plate* (S18.néo) ou encore *comme un petit marteau qui frapperait le son pour qu'il soit produit* (S3.pro).

Nous avons construit 4 sous-catégories de verbalisations décrivant une source sonore possible des sons entendus.

1- On observe en effet des descriptions d'*objets* dans 631 paires (soit 48% des paires), qui se présentent soit directement comme *objet sonore* faisant partie des instruments de musique (*une trompette, un archet...*) ou d'objets sonores de la vie courante (*une sonnette* ou *un Klaxon*), soit d'objets qui peuvent produire un son même si ce n'est pas leur fonction première (*une bouteille, une scie*) ou encore un matériau (*du cristal*). On observe aussi des descriptions d'objets indéfinis du type *chose* ou *truc* et des objets animés (*un canard, une voix*).

2- Les participants décrivent une *action* dans 505 paires (soit 38% d'entre elles). La plupart de ces actions permettant d'émettre le son donc provoquant le début du son (*on frappe, un coup*).

3- Dans 149 paires (soit 18% des paires), codées *matières+excitation*, on trouve des descriptions faisant intervenir de façon plus élaborée à la fois une action et un objet ou une matière sur laquelle cette action est effectuée (*une corde frottée, on frappe sur une tôle*). On retrouve ici la définition physique de ce qu'est un son : une vibration transmise par un matériau donc résultant d'une mise en vibration d'une matière... Comme si le son produit portait la trace de l'action et de la matière responsable de son origine.

4- On trouve enfin des *descriptions* assez variées se rapportant à la source sonore, par exemple un *lieu* (*dans une forêt*) ou des précisions sur les *qualités des objets* ou *des actions* (*une petite flûte indienne, on frappe bizarrement*) mais dans seulement 11% des paires verbalisées.

Ces sous-catégories seront décrites de manière plus précise dans les paragraphes suivants.

La plupart des participants sont concernés par ce type de verbalisations. En moyenne, elles sont présentes dans 59% des paires que chaque participant décrit. Quatre participants les produisent moins souvent (entre 12 et 23 % des paires qu'ils décrivent) et 3 participants plus souvent que les autres (entre 97 et 98% des paires qu'ils décrivent), l'ensemble des autres participants produisant des

descriptions d'une source sonore dans 48 à 80% des paires qu'ils décrivent. Au total, seuls 6 participants (soit 30% d'entre eux) parlent de la source sonore dans moins de la moitié des paires qu'ils décrivent. La répartition de ces verbalisations en fonction de la compétence musicale des participants donne un léger avantage aux non-musiciens (qui produisent 54% de ces verbalisations), ce qui est accentué pour les *descriptions* (de lieu, de pays d'origine, de qualité de la source ou du geste...), produites à 72% par des non-musiciens. Les autres sous-catégories de sources sonores sont relativement bien réparties entre musiciens et non-musiciens.

On peut donc considérer que le fait de décrire un son par une source sonore possible de celui-ci est un phénomène général, peu lié à la compétence musicale des participants et que l'on rencontre assez fréquemment.

### II.5.1. Objets produisant des sons

Nous avons vu que quand les participants décrivent une source sonore possible du son qu'ils entendent, c'est très souvent en décrivant un objet (dans 80% des cas), qu'il ait pour principale fonction de produire des sons (musicaux ou non), ou que ses qualités sonores soient secondaires. Les sous-catégories d'objets verbalisées seront décrites ci-dessous (figure 6.11). La plupart des objets décrits appartiennent au domaine musical (dans 88% des cas), que ce soit pour donner le nom d'un instrument de musique, d'une famille d'instruments ou d'un objet participant à la construction d'un instrument de musique.

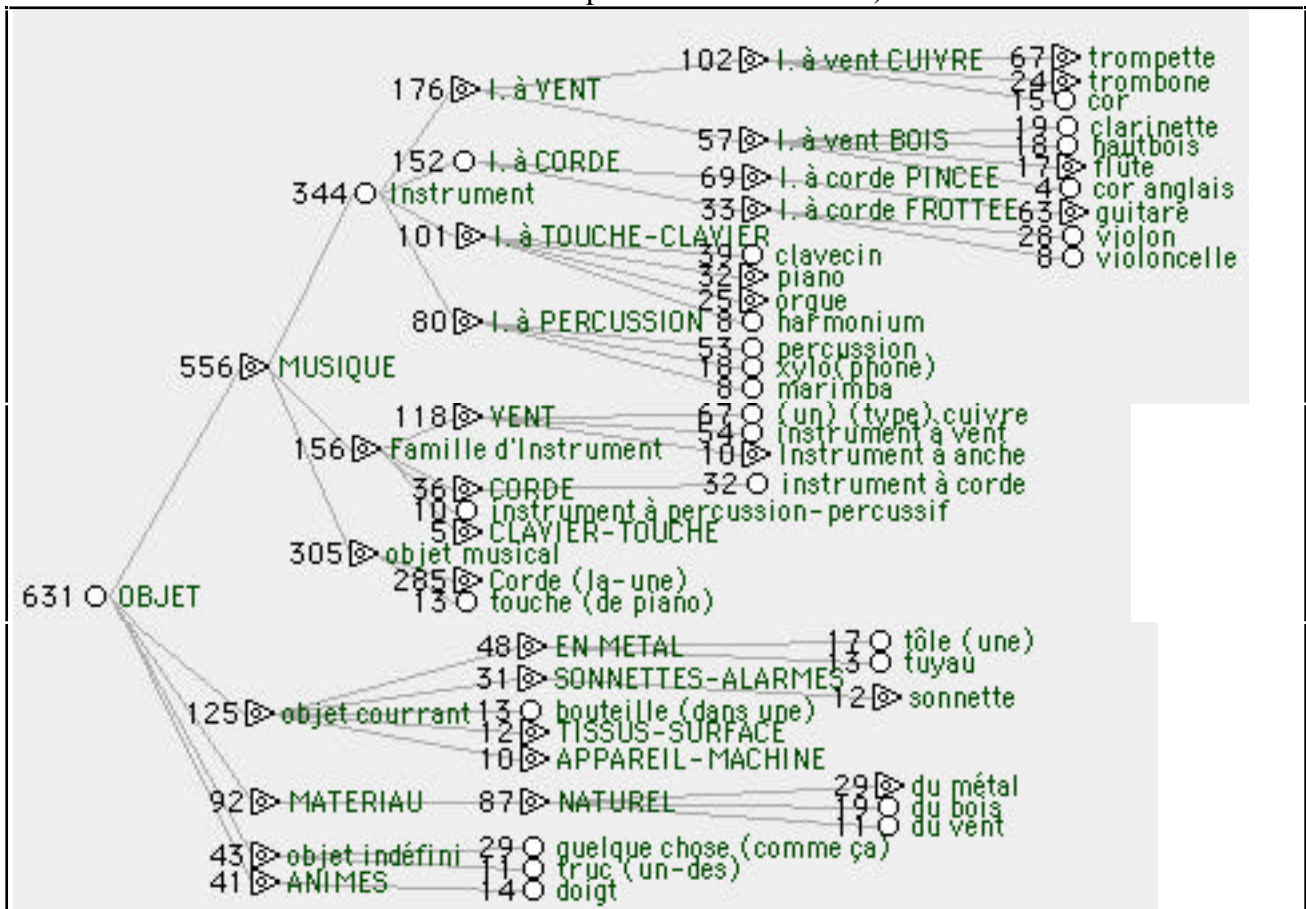
Nous avons réparti les 40 noms d'instruments cités par l'ensemble des participants dans des sous-catégories correspondant aux familles d'instruments dans lesquelles ils sont habituellement classés, ceci afin de pouvoir comparer ultérieurement la production de ces noms d'instruments avec les sons présents dans les paires décrites. Les *noms d'instruments* les plus fréquemment cités sont *trompette* et *guitare*, suivis de *percussion*, *clavecin*, *piano*, *violon*, *orgue*, *trombone*, *clarinette*, *hautbois*, *xylophone*, *flûte* et *cor*. Les participants citent plus souvent des noms d'instruments à vent, puis des noms d'instruments à cordes et enfin des noms d'instruments à clavier et à percussion, mais si nous avons rassemblé ces deux dernières catégories (un piano est un instrument à percussion, comme l'a magistralement rappelé la musique contemporaine...) on obtiendrait une répartition relativement équilibrée dans ces trois catégories.

Les participants donnent parfois plutôt le nom d'une *famille d'instruments*. Quand la locution *instrument* était absente, il nous a parfois été difficile de décider si un participant parlait de *percussion* pour décrire un instrument ou une famille d'instruments. De même, quand il parlait de *corde* et dans une moindre mesure quand il parlait de *cuivre*. Nous avons donc considéré au cas par cas le degré de généralité ou au contraire la désignation de l'objet en vibration. La plupart des occurrences de *corde* se sont donc vues classées dans la catégorie *objet musical* alors que nous avons rassemblé beaucoup d'occurrences de *cuivre* sous la catégorie *famille d'instruments/vent*. Pourtant, si on considère



l'ensemble des dénominations renvoyant respectivement aux instruments à vent, à cordes ou à percussion de la partie relevant du domaine musical, on obtient respectivement 386, 260 et 178 paires.

Figure 6.11. Verbalisations codées sous *parlent de la source/objet* (631 paires, sous-catégories rassemblant plus de 10 occurrences)



Les participants décrivent aussi des objets extérieurs au domaine musical dont la fonction première peut ou non être de produire des sons. Cinquante objets ont ainsi été cités, allant d'objets en métal comme des *tôles*, *tuyaux* ou *pièces de monnaies*, à toute une série de *sonnettes*, *sirènes*, *alarmes*, *klaxons* et autres *avertisseurs* en passant par des outils (*marteau*, *scie*, *couteau*...), des *machines* (*ordinateur*, *voiture*, *circuit électrique*...), des *tissus* ou des *membranes* (*feuilles*, *soie*, *drap*...) et puis aussi des *bouteilles*, *jouets*, *boîtes*... tout ceci souvent mis en scène pour expliquer dans quelles conditions on peut en tirer des sons.

Parfois les locuteurs citent un nom de *matériau*, principalement *du métal*, mais aussi *du bois* (dont on fait les pipes), *du vent* (dans les branches de *Sassafras*), *du verre*, *du cuivre* ou *du cristal*... Enfin ils parlent aussi d'*objets indéfini*, c'est-à-dire qu'ils désignent clairement un objet même si c'est de façon vague (*quelque chose* comme ça, ou *un truc*...), ou ils citent des *objets animés*. On trouve ainsi des parties du corps (essentiellement quand les participants décrivent un mécanisme fonctionnant avec un *doigt*, mais aussi la *bouche*, la *main*...). On rencontre enfin au détour de verbalisations quelques *araignées* ou *canards* mais, n'en déplaise à Prévert (1949), pas de *raton-laveur* !

#### Répartition en fonction des participants

L'ensemble des descriptions d'*objets* est produit à part égale par des musiciens et des non-musiciens. Pourtant cette répartition varie en fonction des sous-catégories. En effet, si les objets relevant du domaine de la *musique* sont produits eux aussi à 54% par des musiciens, ces derniers sont un peu plus nombreux à produire des *noms d'instruments* (65%) et significativement plus nombreux à produire des noms d'instruments de la famille des bois (*clarinette, hautbois, flûte...*). On peut même remarquer que le participant qui produit le plus de verbalisations de ce type (30% à lui tout seul) est un hautboïste professionnel et pratique donc un instrument de la famille des bois. Parmi les autres catégories de verbalisations principalement produites par des musiciens, on trouve aussi les noms d'instruments faisant partie de la classe des claviers (*clavecin, piano...*). Pourtant, chacune des sous-catégories de *noms d'instruments* de musique a toujours été produite par plus de la moitié des participants et, à part les deux exceptions précédemment citées, pas systématiquement par des musiciens.

Par contre, alors que les noms de *famille d'instruments* sont eux aussi produits à 54% par les musiciens, les noms d'objets du domaine *courant* sont cités à 64% par des non-musiciens. Parmi ceux-ci, les noms comme *tôle* et *tuyau*, produit par seulement respectivement 4 et 3 participants différents sont principalement dus aux musiciens, alors que les noms de *sonnettes-alarmes* ou de *bouteille* (produits eux aussi par moins de 6 participants) sont principalement dus aux non-musiciens. Enfin, les noms de *matériau* sont produits légèrement plus souvent par des musiciens (64%) et significativement plus souvent dans le cas *du métal* (79% du aux musiciens). Les noms d'*objets animés* ou d'*objets indéfinis* sont tous les deux significativement produits plus souvent par des non-musiciens.

Tableau 6.4. Répartition des verbalisations de *la source/objets/musique/instruments* en fonction des types de paires de sons (% de chaque type de paire).

	<i>Instrument</i>	<i>I. Vent</i>	<i>I. à vent Cuivre</i>	<i>I. à vent Bois</i>	<i>I à cordes</i>	<i>I. à cordes pincées</i>	<i>I. à cordes frottées</i>
<b>HYBRIDES</b>	22	11	6	3	8	4	2
2	15 -	7	5	2	<b>5 -</b>	3	2
1	23	11	6	3	9	4	2
<b>IMITATIONS</b>	27	14	8	4	12	5	3
frotté	28	10	3	4	15	3	<b>12 +</b>
pincé-frappé	26	10	6	3	15	8 +	2
2	26	12	7	4	14	7	2
1	25	<b>0 -</b>	<b>0 -</b>	0 -	<b>21 +</b>	9 +	2
soufflé	30 +	<b>21 +</b>	12	7 +	10	5	2
2	26	<b>23 +</b>	<b>18 +</b>	8 +	<b>1 -</b>	0 -	0
1	30 +	<b>20 +</b>	11	7 +	12	6	2
moyenne	25.27	12.58	7.44	4.08	11.06	4.94	2.60
écart type	4.18	6.68	4.71	2.43	5.41	2.51	3.26
khideux	0.12	<b>15.07</b>	<b>11.03</b>	2.07	<b>8.73</b>	1.39	<b>19.19</b>
test	0.73	1.10 <sup>-4</sup>	9.10 <sup>-4</sup>	0.15	0.003	0.24	1.10 <sup>-5</sup>

avec + (respectivement -) pour les valeurs s'écartant de la moyenne de plus d'un écart type. En gras, différences significatives.  
Sons hybrides : VBN, OBC, SNO ; Son frotté : CFT ; Sons pincés-frappés : PNO, PNE, CVC, VBR ; Sons Soufflés : TPT, TBN, CLA, CRA.

On peut donc observer une tendance à ce que les noms les plus techniques soient plutôt produits par des musiciens, les noms d'objets courants ou d'objets indéfinis étant plus souvent produits par des non-musiciens. Pourtant les différences ne sont significatives que pour des catégories de verbalisations produites par un nombre faible de participants différents.

#### Répartition en fonction des sons

La répartition des verbalisations de *noms d'instruments* et de *familles d'instruments* en fonction des types de paires de sons est très intéressante à observer. En effet, les verbalisations donnent ici une sorte de mesure directe de l'appartenance des sons aux types de paires que nous avons constitués.

Ainsi, on peut constater dans le tableau 6.4 que, alors que la répartition de l'ensemble des verbalisations de *nom d'instruments* se fait de façon homogène, quelles que soient les paires de sons, les noms *d'instruments à vent* sont plus souvent produits quand au moins un des sons de la paire imite un son "soufflé" et moins souvent si un des sons de la paire imite un son "pincé-frappé". On retrouve à peu près ces résultats dans le cas des noms d'instruments à vent de la famille des cuivres (*trompette, trombone...*), plus souvent produits quand les deux sons de la paire imitent des sons "soufflés". Par contre, les différences concernant les noms d'instruments à vent de la famille des bois (*clarinette, hautbois, flûte..*) ne sont pas significatives. On retrouve des différences significatives dans le cas de noms d'instruments à cordes, plus souvent produits quand un des sons de la paire imite un son "pincé-frappé" et moins souvent quand les deux sons de la paire sont soit des hybrides, soit tous les deux de type "soufflé". Enfin, alors que les différences concernant les noms d'instruments à cordes pincées (*guitare...*) ne sont pas significatives, celles concernant les noms d'instruments à cordes frottées (*violon...*) montrent une forte tendance à produire ces noms uniquement quand un des

sons de la paire imite une corde frottée (son CFT). On retrouve le même type de résultat à propos des noms de familles d'instruments et d'objets musicaux (tableau 6.5).

En effet, l'ensemble des noms de *familles d'instruments* se répartit déjà de façon significativement différente en fonction des paires de sons, les participants produisant plus souvent ces verbalisations quand les deux sons de la paire imitent tous les deux des sons "soufflés" et moins souvent quand ce sont tous les deux des hybrides. Avec la production de noms de famille d'instruments à vent, on retrouve des différences de répartitions significatives et comparables à celles dues aux verbalisations de *noms d'instruments* de musique de la famille des vents, à savoir que ces verbalisations sont plus souvent produites quand un au moins des deux sons de la paire imite une son "soufflé". Les différences concernant les verbalisations à propos de famille d'instruments à corde ne sont, par contre, pas significatives.

Les verbalisations rassemblées dans la catégorie *objet musical*, principalement représentée par des occurrences de *la-une corde*, apparaissent significativement plus souvent quand un des sons de la paire est le son CFT et moins souvent quand les deux sons de la paire imitent des sons "soufflés". On retrouve ces résultats encore plus significatifs avec les seules verbalisations de *la-une corde*. Toutes les autres catégories de verbalisations d'objets ne se répartissent pas de manière significativement différente en fonction des paires de sons.

Tableau 6.5. Répartition des verbalisations de *la source/objets/musique* en fonction des types de paires de sons (% de chaque type de paires).

	<i>Famille d'instruments</i>	<i>VENT</i>	<i>CORDES</i>	<i>Objet Musical</i>	<i>la-une corde</i>
<b>HYBRIDES</b>	10	8	3	23	21
2	<b>5 -</b>	5	2	15	13
1	11	8	3	24	21
<b>IMITATIONS</b>	12	9	3	23	22
frotté	14	5	8 +	<b>31 +</b>	<b>31 +</b>
pincé-frappé	10	6	2	28	26
2	11	7	3	29	27
1	<b>2 -</b>	<b>0 -</b>	1	22	20
<b>soufflé</b>	16	<b>14 +</b>	2	21	19
2	<b>18 +</b>	<b>17 +</b>	0 -	<b>2 -</b>	<b>1 -</b>
1	15	<b>14 +</b>	3	25	23
moyenne	11.17	8.41	2.58	21.90	20.41
écart type	4.68	4.82	1.94	7.98	7.95
khideux	<b>4.52</b>	<b>9.49</b>	2.09	<b>10.46</b>	<b>11.80</b>
test	0.03	0.002	0.15	0.001	0.0006

avec + (respectivement -) pour les valeurs s'écartant de la moyenne de plus d'un écart type. En gras, différences significatives.  
Sons hybrides : VBN, OBC, SNO ; Son frotté : CFT ; Sons pincés-frappés : PNO, PNE, CVC, VBR ; Sons Soufflés : TPT, TBN, CLA, CRA.

De plus, la quasi totalité des sous-catégories de verbalisations rassemblant plus de 10 occurrences sont réparties de façon significativement différente en fonction des sons présents dans les paires. En effet, seules les verbalisations *musique* en général, *tuyau*, *matériaux*, *objets indéfinis* et *animés* ne

sont pas réparties de manières significativement différentes en fonction des sons présents. On rencontre en effet plus de verbalisations de *noms d'instruments* en général quand un des sons de la paire est CLA, CVC ou TPT et moins souvent si c'est SNO, VBN et PNO ( $\chi^2=22$ ,  $p=.02$ ,  $df=11$ ). Si on regarde en détail les familles d'instruments auxquelles les *noms d'instruments* verbalisés appartiennent, les résultats sont toujours concordants avec ce que les sons concernés sont censés imiter. Ainsi, les participants parlent plus souvent de noms d'instruments à vent quand un des sons de la paire est le son CLA, TPT ou TBN, les deux derniers étant plus souvent présents quand il s'agit de noms d'instruments à vent de la famille des cuivres (*trompette, trombone...*) et le premier quand il s'agit de noms d'instruments à vent de la famille des bois (*clarinette, hautbois, flûte...*) ( $\chi^2(11)>68$ ,  $p<2.10^{-10}$ ). On peut noter ici l'absence du son CRA, censé imiter un son de cor anglais, qu'on retrouve plus souvent au rayon des *sonnettes*.

On observe le même genre de résultat en ce qui concerne les *noms d'instruments à cordes* en général, plus souvent produits quand un des sons de la paire est le son CVC et moins souvent si c'est le son VBN. Les noms d'instruments à cordes pincées (*guitare...*) étant plus souvent produits quand la paire contient un des sons CVC ou VBR et moins souvent si c'est le son VBN, les noms d'instruments à cordes frottées (*violon*) sont essentiellement produits quand un des sons de la paire est le son CFT ( $\chi^2(11)>27$ ,  $p<.004$ ). Les noms d'instruments à touche-clavier (*clavecin, piano, orgue...*) sont plus souvent produits quand un des sons CVC ou PNE est présent et moins souvent si c'est le son VBN ( $\chi^2=34$ ,  $p=4.10^{-4}$ ,  $df=11$ ). Enfin, les noms d'instruments à percussion (*percussion, xylophone...*) sont plus souvent produits quand un des sons de la paire est soit le VBR soit le VBN ( $\chi^2(11)=42$ ,  $p=2.10^{-5}$ ). On retrouve ainsi le caractère percussif de cet hybride de *vibraphone* et de *trombone*.

Les noms de *familles d'instruments* sont en général produits plus souvent quand un des sons de la paire est le TBN et moins souvent si c'est le CVC ou le PNO. Ceci nous paraît assez inexplicable. Par contre on comprend très bien que les noms de *famille d'instruments à vent* soient produits plus souvent quand un des sons de la paire est soit le TBN, soit le TPT (tous les deux des imitations de cuivres). L'importance de la présence ici du TBN explique sans doute le résultat précédent. Les noms de *familles d'instruments à cordes* sont plus souvent produits quand un des sons de la paire est le son CFT. Il est ici intéressant de remarquer qu'à chaque fois, les sons les plus souvent cités dans ces catégories de familles d'instruments (respectivement famille d'instruments à vent et à corde) font partie d'une sous-catégorie de celle énoncée (ils imitent respectivement des cuivres (TPT et TBN) et une corde frottée (CFT)). Est-on en présence d'un changement de niveau de base ?

Les verbalisations d'*objet musical* se rencontrent plus souvent quand un des sons de la paire est le CFT ou de CVC et moins souvent si c'est le CRA ou de SNO ( $\chi^2(11)=27$ ,  $p=4.10^{-3}$ ). Ces résultats semblent dus au fait que cette catégorie est dominée par les verbalisations de *la-une corde*. En effet, on retrouve ces résultats si on considère uniquement les occurrences correspondantes. On peut noter à ce propos que l'hybride SNO, bien qu'étant conçu comme un mélange de corde frottée et de piano, a perdu tout caractère de son d'instrument à corde.

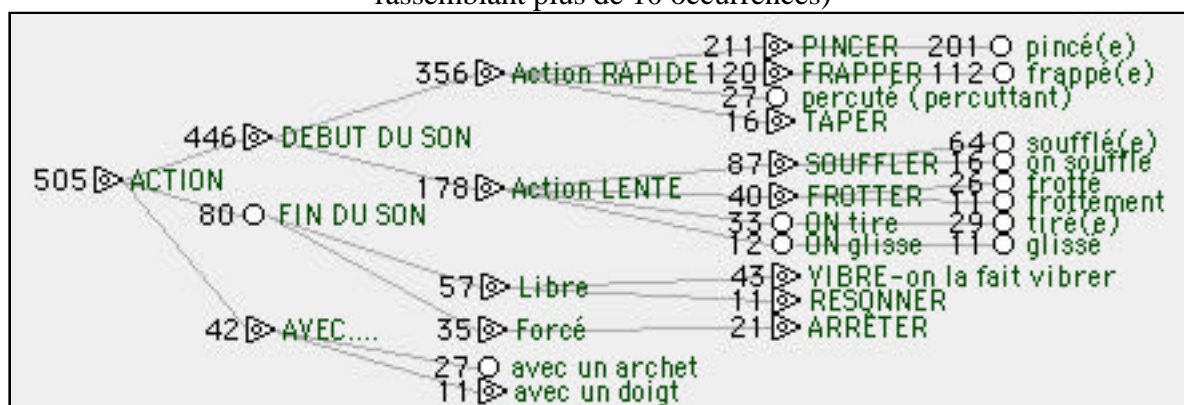
Enfin, les noms d'*objets courants* sont plus souvent produits si un des sons de la paire est le VBR, CRA ou VBN et moins souvent si c'est le SNO ( $\chi^2(11)=26, p=7.10^{-3}$ ). Ceci semble dû au fait que la plupart des sous-catégories de ces verbalisations ne sont produites qu'en présence d'un ou deux des sons cités : le VBN pour les verbalisations de *en métal* et *tôle*, les sons CRA et TPT, particulièrement synthétiques, pour les verbalisations de *sonnettes-alarmes* et le VBR pour les occurrences de *bouteille*.

On peut donc remarquer que la plupart des productions de noms d'objets, qu'ils soient musicaux ou non sont très spécifiques de la présence d'un son précis dans les paires décrites. On retrouve notamment le résultat décrit par Samoylenko et al. (1996) que les descriptions sous forme de nom d'instruments renvoient aux sons qui imitent ces instruments.

### II.5.2. Actions produisant des sons

Les verbalisations décrivant une source sonore, contiennent aussi beaucoup de descriptions d'*actions*, principalement d'actions permettant de produire un son, donc intervenant au début de celui-ci . On retrouve ainsi d'une certaine manière les types de paires de sons, puisque nous avons réparti les actions verbalisées par les participants en *actions rapides*, principalement dues aux occurrences des verbes *pincer*, *frapper*, *percuter* et *taper*, et des *actions lentes* (ou plutôt qui prennent plus de temps) et qui sont principalement dues aux occurrences des verbes *souffler*, *frotter*, *tirer* et *glisser*. On trouve aussi décrites quelques actions qui ont lieux à la fin des sons et dont les effets sont soit l'impression d'une *fin libre* avec les occurrences des verbes *résonner* et *vibrer*, soit d'une *fin forcée*, avec principalement les occurrences du verbe *arrêter*. On trouve aussi quelques descriptions de moyens d'action, par exemple *avec un archet* ou *avec un doigt*.

Figure 6.12. Verbalisations codées sous *parlent de la source/action* (505 paires, sous-catégories rassemblant plus de 10 occurrences)



Tous les participants produisent à un moment donné des verbalisations d'*action rapide* décrivant le début des sons, et 85% d'entre eux produisent aussi des verbalisations d'*action lente*. Cela souligne l'importance psychologique de décrire les actions qui permettent de produire les sons. En général, les verbalisations d'*actions* sont également produites par des participants musiciens ou non-musiciens,

même si les musiciens sont plus nombreux à produire les occurrences du verbe *glisser* (en fait essentiellement dues au seul participant S6.am1) et les non-musiciens sont plus nombreux à produire les occurrences des verbes *vibrer* et *arrêter* (67%). Pour ce qui est de la description de moyen d'action, les musiciens produisent plus souvent *avec un archet* et un des non-musiciens a produit la quasi totalité des occurrences de *avec un doigt*.

La répartition des sous-catégories de descriptions d'actions en fonction des types de paires de sons d'une part, et des sons présents dans ces paires d'autre part, est ici aussi très intéressante, du fait que nous avons coutume de classer les différents instruments de musique en fonction justement de leur mode de production. La répartition des verbalisations d'actions en fonction des paires de sons est significativement différente en général, ainsi que pour ce qui concerne la description des actions de *début du son*. En effet, ces deux types de verbalisations sont plus souvent produits quand un des sons de la paire est le CFT, seul exemplaire d'imitation de son frotté, et moins souvent si les deux sons de la paire imitent des sons "soufflés". Le résultat concernant le CFT est à relier au fait que ce son est significativement plus souvent présent dans les paires à propos desquelles les participants parlent d'*action lente* en général. CFT apparaît moins, soit si les deux sons de la paires sont tous les deux des hybrides, soit si un des deux sons est une imitation de son pincé ou frappé. Parmi les verbes dont les occurrences constituent cette catégorie, seul le verbe *frotter* est produit significativement plus souvent quand un des sons de la paire est le CFT, les répartitions des productions des autres verbes (*souffler*, *tirer*, *glisser*) sur les différents types de paires de sons n'étant pas significatives. Ceci nous permet de montrer que dans la constitution des différents niveaux de catégorisation des verbalisations libres produites par les participants, il y a un niveau qui est plus facilement comparable aux données physiques que les autres, et que souvent ce n'est pas le niveau directement produit par les participants, ce qui nécessite un rassemblement des verbalisations de sens proches.

**Tableau 6.6.** Répartition des verbalisations de *la source/action* en fonction des types de paires de sons (% de chaque type de paires).

	<i>Action</i>	<i>début du son</i>	<i>action rapide</i>	<i>action lente</i>	<i>frotter</i>	<i>avec un archet</i>
<b>HYBRIDES</b>	35	31	26	11	2	2
2	32	28	27	<b>5 -</b>	2	2
1	36	31	25	11	2	2
<b>IMITATIONS</b>	38	34	27	14	3	2
frotté	<b>49 +</b>	<b>43 +</b>	26	<b>32 +</b>	<b>15 +</b>	<b>11 +</b>
pincé-frappé	44	39	<b>34 +</b>	13	3	1
2	45	40	<b>35 +</b>	15	3	2
1	39	34	33	<b>3 -</b>	0	0
soufflé	35	31	23	13	2	1
2	<b>14 -</b>	<b>13 -</b>	<b>6 -</b>	8	0	0
1	39	35	26	14	2	1
moyenne	37.07	32.75	26.18	12.78	3.07	2.14
écart type	9.16	7.92	7.86	7.39	4.23	2.99
2	6.29	4.28	6.89	20.67	178.99	19.91
p	0.01	0.04	0.009	5.10 <sup>-6</sup>	8.10 <sup>-9</sup>	8.10 <sup>-6</sup>

avec + (respectivement -) pour les valeurs s'écartant de la moyenne de plus d'un écart type. En gras, différences significatives.  
Sons hybrides : VBN, OBC, SNO ; Son frotté : CFT ; Sons pincés-frappés : PNO, PNE, CVC, VBR ; Sons Soufflés : TPT, TBN, CLA, CRA.

Les verbalisations d'*actionsrapides* sont plus souvent produites quand au moins un des sons de la paire imite un son "pincé-frappé" et moins souvent si les deux sons imitent des sons "soufflés". Par contre, on ne retrouve pas ces différences significatives sur les occurrences des verbes *pincer*, *frapper*, *taper*, *percuter* qui sont sous-jacents à cette catégorie de verbalisations.

Par contre, la quasi totalité des sous-catégories sont réparties de manière significativement différente en fonction des sons présents dans les paires décrites. De plus, les sons les plus souvent présents dans les paires décrites sont toujours ceux qui sont censés imiter le mode de production décrit par l'action verbalisée. Ceci montre une fois de plus la justesse et la pertinence des verbalisations des participants, d'autant plus que ce type de verbalisations n'est que très rarement dû à un sous-groupe de participants ayant une compétence musicale établie. En effet, les occurrences du verbe *pincer* sont plus souvent produites quand un des sons de la paire est soit le CVC soit le OBC, qui possède l'enveloppe temporelle du clavecin, qui sont les deux seuls imitant des sons produits par un pincement ( $\chi^2(11)=61$ ,  $p=5.10^{-9}$ ). Au contraire, les occurrences du verbe *frapper* sont plus souvent produites quand un des sons de la paire est soit le son PNE, soit le son PNO ( $\chi^2(11)=33$ ,  $p=5.10^{-4}$ ). Seul le VBR imite lui aussi une production par frappement, mais peut-être sa particularité (ce n'est pas une corde frappée, il est donc plus souvent considéré comme provenant d'un instrument à percussion) explique le fait qu'il soit considéré différemment. De même, les occurrences du verbe *frotter* sont plus souvent produites si un des sons de la paire est le CFT qui est le seul son de notre ensemble à imiter une production par frottement ( $\chi^2(11)=126$ ,  $p=9.10^{-22}$ ). Le cas des occurrences du verbe *souffler* est un peu différent. En effet, il est produit plus souvent quand un des sons de la paire est le TBN ( $\chi^2(11)=36$ ,  $p=2.10^{-4}$ ), alors que 3 autres sons sont censés eux aussi imiter des sons produits en soufflant. Pourtant, le caractère très artificiel de la synthèse des sons d'instruments à vent sur le TX802 utilisé ici permet de comprendre ce résultat. On rappelle que les sons TPT et de CRA sont les plus souvent présents quand les participants parlent de *sonnettes*, d'*alarmes*, de *klaxons*... ce qui traduit bien leur caractère artificiel.

Ces résultats, concernant les occurrences des verbes *frotter*, *souffler*, *pincer* et *frapper*, se répercutent aux niveaux supérieurs de l'arborescence. Ceci mène à des différences de répartition des verbalisations d'*actionrapide* et *actionlente*, ainsi que pour celles de *début du son*, *fin du son* et même d'*action* en général ( $\chi^2 > 28$ ,  $p < .003$ ,  $dl=11$ ). Nous ne détaillerons que les différences concernant les verbalisations sur la *fin des sons*. En effet, celles-ci sont plus souvent produites quand un des sons de la paire est le son CVC ou VBN. Or le son CVC imite le son créé par la mécanique du retours du sautereau, qui bloque la corde et arrête donc la vibration. Il est donc intéressant de remarquer la présence de verbalisations d'*arrêter* ainsi que de *fin de son forcée*, plus souvent quand le CVC est dans la paire (respectivement  $\chi^2(11)=23$  et  $32$ ,  $p=.003$  et  $7.10^{-4}$ ). On peut rappeler que les verbalisations de *fin forcée* sont présentes dans 35 paires décrites par 11 participants différents, également répartis entre musiciens et non-musiciens. Les verbalisations sur la fin des sons concernant le VBN semblent plutôt dues aux verbalisations de fin *libre* et en particulier des occurrences de *on*



*laisse vibrer* (respectivement  $\chi^2(11)=23$  et  $26$ ,  $p=.02$  et  $.005$ ). Ce son, en temps qu'hybride de vibraphone et de trombone, possède une enveloppe temporelle ressemblant à celle du vibraphone, notamment dans sa résonance finale.

On ne peut que souligner à nouveau ici la finesse des descriptions faites par les participants, ainsi que leur adéquation aux sons entendus, qui ne dépend pas de la compétence des sujets. Ces résultats vont notamment à l'encontre de l'idée que perceptivement, on a tendance à confondre des sons émis d'une part de façon pincée ou frappée et d'autre part de façon soufflée ou frottée, et ce, même si la forme des enveloppes temporelles à l'intérieur de ces deux groupes de sons se ressemblent beaucoup.

### II.5.3. Mécanisme produisant des sons

Nous avons vu dans les deux paragraphes précédents que les participants qui devaient décrire des sons verbalisaient souvent des noms d'*objets* et des verbes d'*actions* permettant de produire des sons. En fait, une partie de ces verbalisations contiennent la description complète d'un mécanisme de production de ces sons, plus ou moins imaginaire mais toujours assez vraisemblable. Ces verbalisations, présentes dans 243 paires (soit 18% des paires) sont pourtant difficiles à décrire de façon générale. En effet, à part l'ensemble des descriptions de différents *types de cordes* (*corde pincée*, *corde frappée*, *corde frottée*, *corde tirée*), qui est d'ailleurs difficile à distinguer de verbalisations de noms de familles d'instruments. La caractéristique principale de ces verbalisations est justement le fait qu'elles soient à chaque fois très spécifiques et dans leur grande majorité citées une seule fois (et toujours moins de 5 fois).

Figure 6.13. Verbalisations codées sous *parlent de la source/matière + excitation* (243 paires, sous-catégories rassemblant le plus d'occurrences et quelques exemples)



Les regroupements effectués ici (figure 6.13) sont donc assez arbitraires et surtout basés sur l'action verbalisée. Nous ne testerons donc pas la répartition de ces verbalisations en fonction des sons présents dans les différentes paires décrites. Pourtant ces descriptions sont sans doute parmi les plus spécifiques de chacun des sons décrits et permettraient donc certainement de les identifier ou de les reconnaître facilement. Par exemple, l'ensemble des 14 occurrences de *un doigt coincé dans une bouteille* ou *le doigt qui sort de la bouteille*, sont toutes produites par le participant S19.néo pour décrire le son VBR. On peut noter que ce son a été aussi souvent décrit en référence à un matériau

comme le *verre* ou le *cristal* par d'autres participants. Cette impression est donc partagée par plusieurs participants et spécifiquement associée à ce son. Nous n'avons malheureusement pas de moyens statistiques de prouver leur pertinence. Nous ne pouvons que souligner le fait que ces descriptions paraissent évidentes et pertinentes à quelqu'un qui la lit et entend le son en même temps, ce qui est un point de vue subjectif et qualitatif. Certaines de ces locutions seront testées en reconnaissance de portraits verbaux dans notre troisième expérience, afin de donner un caractère quantitatif à cette remarque.

Ces verbalisations sont produites au moins une fois par la quasi-totalité des participants de notre expérience (95%). Pourtant, si l'ensemble de ce type de verbalisation est dû un peu plus souvent à des participants musiciens (58%), ils sont locuteurs de 71% des verbalisations des différents *types de cordes* alors que les non-musiciens sont locuteurs de 65% des autres verbalisations de cette partie. Les descriptions de mécanismes ayant pu produire les sons écoutés, codés *matière+excitation*, sont donc peut-être un moyen de décrire une source sonore plus accessible aux non-musiciens qui n'ont pas la possibilité de citer des noms d'instruments comme *cor anglais* ou *vibraphone*.

#### II.5.4. Description d'une source complexe

Enfin, nous avons rassemblé les autres types de descriptions de la source sonore dans la catégorie *description*, qui regroupe donc des verbalisations très variées et totalisant très souvent peu d'occurrences. Cependant, elles concernent quand même 149 paires de sons et on peut tout de même en dégager quelques sous-catégories. On trouve ainsi des indications sur l'*interprétation* d'un instrumentiste imaginé, sur le *style musical* auquel ces sons font penser, des noms de *lieux* où ses sons semblent produits ou de *pays* d'où ils semblent provenir, des indications de *mélange* ou d'associations d'instruments, ainsi que des noms de qualité s'appliquant cette fois-ci non pas au son mais à l'objet ou à l'action décrite (*une corde métallique, un coup rapide*).

La plupart des participants produisent au moins une fois ce type de verbalisations (85%), principalement des non-musiciens (pour 40% des occurrences de ce type de verbalisations), même si elles sont principalement dues à trois participants qui totalisent à eux trois 49% des occurrences de ce type. On retrouve parmi eux le participant S16.néo qui est celui qui a produit la plus grande quantité de verbalisations et qui produit ici 23% des verbalisations de ce type, mais aussi les participants S18.néo et S13.néo. Le faible nombre d'occurrences de ces verbalisations, rend difficile le test de leur répartition en fonction des sons présents dans les paires décrites. On peut tout de même remarquer que 5 des 6 citations de verbalisations de noms de pays évoquant *l'orient* ou *l'extrême orient* ont été produites en présence de l'hybride OBC, ce qui concernent tout de même 3 participants différents.

### **III. Expérience 3 : Description de sons en fonction de consignes différentes**

Nous avons demandé aux participants, au début de notre troisième expérience, d'écouter une fois l'ensemble des 12 sons de notre étude, puis de les décrire chacun leur tour en répondant à 4 consignes différentes selon le sous-groupe de portraits qu'ils allaient avoir à juger par la suite. Ainsi, ceux qui ont eu à associer les sons aux portraits de noms d'instruments, de famille d'instruments ou "d'instruments visés par le concepteur" ont dû répondre à la consigne C1 : "Décrire chaque son en donnant si possible un nom d'instrument ou de famille d'instruments " ; ceux qui ont eu à associer les sons aux portraits de source sonore ont dû répondre à la consigne C2 : "Décrire chaque son en donnant si possible un nom de source sonore (par exemple, pour une odeur, on pourrait dire "l'odeur du foin coupé après la pluie")" ; ceux qui ont eu à associer les sons aux portraits de qualités simples et triples ont dû répondre à la consigne C3 : "Décrire chaque son en donnant si possible un ou plusieurs noms de qualité (par exemple, pour une odeur, on pourrait dire "une odeur forte, acide, enivrante")" ; enfin, ceux qui ont eu à associer les sons aux portraits de qualités doubles et aux autres qualités simples ont dû répondre à la consigne C4 : "Décrire chacun des sons comme vous le voulez".

Ces demandes répondent à trois objectifs.

1- Tout d'abord, pouvoir comparer plus directement le vocabulaire testé sous forme de portrait à une production spontanée de ce vocabulaire. Les sujets de cette expérience étant beaucoup plus nombreux et différents des sujets de notre première expérience qui était notre seule source de verbalisations libres, on voulait savoir si un autre groupe de sujets produirait des descriptions comparables, c'est-à-dire si le vocabulaire testé sous forme de portrait était familier pour ces nouveaux participants.

2- Pouvoir recueillir des verbalisations libres décrivant chacun des sons et non plus les sons 2 à 2 dans une comparaison par paires. Ceci nous permettrait notamment d'enregistrer seulement 12 verbalisations par sujet, sans ambiguïté possible sur les sons décrits, au lieu de 66 correspondant aux 66 paires possibles de sons différents, où chaque verbalisation est à scinder en : ce qui concerne les deux sons, ce qui concerne le premier et ce qui concerne le second. Les possibilités d'analyse, notamment avec le logiciel Stone, seraient alors plus simples.

3- Vu que la consigne utilisée lors de l'expérience 1 : "décrire toutes les différences et ressemblances entre les deux sons de la paire" avait amené les participants à décrire une source possible de ces sons dans 60% de leurs descriptions, nous voulions tester d'éventuelles différences de vocabulaire recueillies en fonction des consignes demandées.

Malheureusement, compte tenu du temps que demande le codage de verbalisations libres, nous n'avons pas analysé de manière exhaustive l'ensemble du matériel recueilli ici. Seules les descriptions sous forme de nom d'instrument et de famille d'instruments (C1) ont été entièrement codées dans le logiciel Stone, ce qui nous permettra de comparer l'arborescence résultante de celle obtenue lors de

l'analyse des données de notre première expérience. Les descriptions correspondant aux 3 autres consignes n'ont été codées que vis à vis des champs sémantiques correspondant aux portraits testés dans l'ensemble de l'expérience 3, soit, les noms d'instruments, de familles d'instruments, de matériaux, de modes d'excitation, de combinaisons des deux et les adjectifs présentés soit seuls soit par 2 ou par 3. Nous ne pourrions donc comparer ces différentes consignes que vis à vis de ces locutions.

Nous pourrions aussi tester d'éventuelles différences de longueurs des descriptions en fonction de la consigne. Compte tenu qu'une observation de surface sur l'ensemble de ces données donne l'impression que les descriptions se ressemblent beaucoup, malgré les différences de consigne (les sujets étant en difficulté, ils finissent par dire ce qu'ils peuvent...), il aurait été extrêmement intéressant de pouvoir tester l'importance respective des descriptions où le locuteur "parle du son" vs "parle de la source" en fonction de la consigne...

Enfin, nous avons noté systématiquement les gestes effectués par les participants lors de leur description des sons. Nous pourrions faire une description de ces gestes et de leur éventuel lien privilégié avec une des 4 consignes demandées.

### III.1. Comparaison des 4 consignes

L'influence des 4 consignes données aux participants pour décrire les sons va être étudiée d'une part sur la longueur des descriptions obtenues, et d'autre part sur leur contenu en comparaison avec celui des portraits verbaux testés et relevant de types de verbalisations attendus dans chacune de ces consignes.

#### **III.1.1. Longueur des descriptions**

Compte tenu des remarques des sujets (cf. III.4) et de l'importance des descriptions de sources sonores lors de l'expérience 1, avec une consigne qui n'induisait pourtant pas *a priori* ce genre de descriptions, nous avons souhaité tester l'hypothèse suivante : Il est plus facile de nommer la source que de décrire des qualités sonores. Pour cela, nous voulions comparer la longueur des descriptions obtenues avec les différentes consignes. Une description serait plus longue si elle est plus difficile à formaliser. On s'attend donc par exemple à ce que la description sous forme de nom d'instrument de musique soit beaucoup plus courte que celle sous forme de qualités des sons.

Nous avons testé deux indicateurs de la longueur de chaque description : le nombre de caractères de la transcription de ces verbalisations dans un premier temps, et le nombre d'"unités verbales" différentes dans chaque description dans un deuxième temps. Par "unité verbale", nous entendons, à la manière de Samoylenko et al. (1996), les plus petites locutions porteuses de sens différents extraites des verbalisations.

#### Mesurée en nombre de caractères

Nous avons donc analysé la longueur, mesurée par le nombre de caractères dans la transcription, de chaque verbalisation spontanée produite par les locuteurs dans l'expérience 3, en fonction, de la consigne de verbalisation : C1 : "Décrire chaque son en donnant si possible un nom d'instrument ou de famille d'instruments ", C2 : "Décrire chaque son en donnant si possible un nom de source sonore", C3 : "Décrire chaque son en donnant si possible un nom de qualité" ou C4 : "Décrire chacun des sons comme vous le voulez".

On observe des longueurs de verbalisations libres (mesurées en nombres de caractères moyens) différentes en fonction des consignes données aux sujets. En effet, les verbalisations les plus longues sont obtenues dans les cas de consignes C3 et C4 (en moyenne respectivement 151 et 141 caractères). Ces deux moyennes ne sont pas significativement différentes ( $t(454)=1.08$ ,  $p=.28$ ). Par contre, quand on oriente les locuteurs vers la source sonore (consignes C1 et C2), la longueur des verbalisations produites est plus courte (en moyenne respectivement 51 et 87). Ces deux moyennes sont significativement différentes ( $t=6.17$ ,  $p<.0001$ ,  $dl=442$ ). Les verbalisations obtenues avec une des deux consignes orientant la description vers la source sonore (C1 ou C2) donnent toutes des nombres de caractères moyens significativement plus courts que dans les deux autres cas (C1/C4 :  $t(466)=11.60$ ,  $p<.0001$  ; C2/C4 :  $t(430)=6.12$ ,  $p<.0001$  ; C1/C3 :  $t(466)=13.08$ ,  $p<.0001$  ; C2/C3 :  $t(430)=7.39$ ,  $p<.0001$ ). De plus, si on regroupe les deux consignes sources sonores (instrument et source : C1 et C2) et les deux autres (libres et qualités : C4 et C3), les différences entre leurs nombres moyens de caractère verbalisés (respectivement égaux à 67 et 146) sont significativement différents ( $t(898)=13.41$ ,  $p<.0001$ ).

#### Mesurée en nombre d'unités verbales

Nous avons analysé la longueur, mesurée par le nombre d'unités verbales (nombres de locutions de sens différents) de chaque verbalisation spontanée produite par les sujets dans l'expérience 3, en fonction, de la consigne, du son écouté, de la position de la description dans les 12 demandées et de chaque participant. En effet, ceci nous permettrait de tester un éventuel effet de la consigne (on s'attend à ce qu'il soit plus facile de nommer la source que de décrire des qualités sonores), mais aussi du son décrit (on s'attend à ce que si un son est vraisemblable, il sera rapidement identifié, s'il est plus difficile à identifier, il sera plus longuement décrit).

#### *Consigne*

On retrouve l'idée qu'il y a plus d'unités verbales (U.V.) différentes produites lors des consignes libres et qualités (C4 et C3) (respectivement en moyenne  $4.3\pm 6.7$  et  $5.1\pm 7.4$  U.V.) que lors des consignes instrument et source (C1 et C2) (respectivement en moyenne  $2.2\pm 2.8$  et  $3.3\pm 3.8$  U.V.). Cette fois-ci, toutes les différences, en terme de nombre d'unités verbales produites, entre consignes prises 2 à 2 sont significatives ( $t>2.9$ ,  $p<.004$ ,  $dl$  entre 453 et 514).

### *Sons*

En moyenne, le nombre d'unités verbales produites pour décrire chaque son est de 3.7, on observe plus d'unités verbales produites quand les locuteurs décrivent le son VBN ( $t(80)=2.22$ ,  $p=0.03$ ) et moins d'U.V. quand ils décrivent le CNT ( $t(80)=2.05$ ,  $p=0.04$ ). Ceci pourrait être un signe de spécificité pour le VBN, peut-être de simplicité (facilité d'identification ?) pour le CNT.

On retrouve des différences significatives entre les sons seulement dans 3 des 4 consignes. En effet, dans la consigne C4 ("libre"  $M=4.4$ ), on observe aucune différence significative du nombre d'U.V. produites liée au son décrit. Par contre, on observe une différence significative dans les consignes C3 ("qualité"  $M=5.1$ ) et C2 ("source"  $M=3.3$ ). Il s'agit, du TBN, décrit avec moins d'U.V. que la moyenne dans la consigne "qualité" (4.0 U.V.,  $t=2.94$ ,  $p=.008$ ,  $dl=19$ ) et du CVC, décrit avec moins d'U.V. que la moyenne dans la consigne "source" (2.3 U.V.,  $t(17)=2.52$ ,  $p=0.02$ ). Ceci pourrait être dû à une facilité d'identifier une source sonore pour le CVC et à la facilité de trouver une qualité saillante dans le contexte pour le TBN. Enfin, dans la consigne "instrument" (C1), pour laquelle le nombre moyen d'U.V. produites pour décrire chaque son est de 2.2, on observe 2 différences significatives. On utilise moins d'U.V. pour décrire le CNT (1.5 U.V.,  $t(22)=4.71$ ,  $p=0.0001$ ) et pour décrire le son TBN (1.7 U.V.,  $t(22)=2.36$ ,  $p=.03$ ).

### *Ordre de description*

On n'observe globalement pas de différence significative de nombre d'U.V. produites en fonction de l'ordre de description des sons. On retrouve ce résultat à l'intérieur de chacune des consignes, sauf pour la consigne C1 ("instrument",  $M=2.24$ ), pour laquelle on obtient moins d'U.V. lors de la description du deuxième et du dernier son (respectivement : 1.7 et 1.6 U.V.,  $t(22)=2.27$  et  $2.74$ ,  $p=0.03$  et  $0.01$ ).

### *Locuteurs*

On observe par contre de grandes différences de nombre d'U.V. produites en fonction des locuteurs, qui se retrouvent à l'intérieur de chaque consigne. En effet, on obtient pour les consignes C1, C2, C3 et C4 respectivement 48, 53, 60 et 55% de sujets qui présentent des différences significatives de nombre d'U.V. produites par rapport à la moyenne obtenue dans la consigne qu'ils ont suivie. Ces différences semblent donc essentiellement dues à la personnalité de chaque locuteur.

### Conclusion

Il apparaît donc des descriptions plus longues quand les locuteurs doivent décrire les sons en donnant si possible une qualité des sons, ou en le faisant librement sans consigne précise que quand les locuteurs sont orientés vers le fait de donner un nom d'instrument de musique ou de famille d'instruments, ou quand on leur demande de décrire une source sonore possible. Ceci pourrait aller dans le sens d'une plus grande facilité à identifier la source d'un son plutôt qu'à décrire ses qualités.

De plus, le son VBN est globalement décrit avec plus d'U.V. que la moyenne, ce qui pourrait être un signe de sa spécificité (c'est un hybride très souvent reconnu pour son caractère *bizarre*, *raté*, ou

*double*). D'autres sons sont décrits avec moins d'U.V. que la moyenne : le CNT globalement, mais aussi les sons TBN dans la consigne "qualité", CVC dans la consigne "source", TBN et CNT dans la consigne "instrument". Ceci pourrait être lié à la saillance des concepts à décrire pour ces sons dans chacun de ces contextes : il serait alors facile de donner un nom d'instrument ou de famille d'instruments pour les sons de CNT et TBN, de donner un nom de source sonore pour le son CVC et de qualité sonore pour le son TBN.

### III.1.2. Locutions utilisées comme portrait

Comme le codage de l'intégralité des verbalisations correspondant aux quatre différentes consignes de verbalisations n'a pas encore été effectué, la comparaison ne pourra se faire que sur les locutions dont ont été extraits les portraits de l'expérience 3. Nous allons donc considérer ici tous les portraits qui ont été testés dans l'expérience 3 (36 noms d'instruments, 6 noms de familles d'instruments, 12 noms de fichiers sons, 12 noms de matériaux, 17 noms de modes d'excitation, 26 noms de sources sonores et au total 41 adjectifs différents, utilisés soit seuls, soit associés par deux ou par trois.

Ce vocabulaire avait été extrait des verbalisations libres de l'expérience 1. Nous avons cherché s'il était présent dans les verbalisations libres de l'expérience 3, et si sa présence était fonction de chacune des 4 consignes de descriptions demandées aux participants. En effet, on s'attend par exemple à trouver plus de noms d'instruments de musique produits dans la consigne "instrument" que dans les 3 autres. Au contraire, on s'attend à trouver plus d'adjectifs produits dans la consigne "qualité" que dans les trois autres. Nous allons donc analyser les locutions testées ci-dessous, réparties en fonction du type de portrait auxquelles elles renvoient. Les tableaux des données correspondantes sont en annexe A.6.3 à A.6.11.

#### Noms d'instruments de musique ou de familles d'instruments

Sur l'ensemble des 40 noms d'instruments dont la présence a été recherchée dans les verbalisations libres de l'expérience 3, 31 ont été produits, ce qui représente au total 489 occurrences, dont la répartition est significativement différente en fonction de la consigne ( $\chi^2(3)=22.95$ ,  $p<.05$ ). On observe en effet une sur-représentation de noms d'instruments verbalisés dans le cadre de la consigne "instrument". Ceci correspond donc totalement à l'effet attendu. Pourtant, si on considère chacun des 31 noms d'instruments produits ici (sur les 40 dont la présence a été testée), on ne retrouve de différences significatives (en terme de différences de nombres d'occurrences en fonction de la consigne suivie) seulement pour deux noms d'instruments (soit 5% d'entre eux). Il s'agit des noms *piano* et *guitare*, pour lesquels on observe à chaque fois une sur-représentation lors de la consigne "instrument", les autres noms d'instruments apparaissant avec la même fréquence dans les 4 consignes différentes.

Tous les noms de familles d'instruments (6) recherchés ont été produits lors des verbalisations libres de l'expérience 3, ce qui représente au total 365 occurrences, dont la répartition est significativement

différente en fonction de la consigne ( $\chi^2(3)=26.06$ ,  $p<.01$ ). On observe en effet une sur-représentation de noms de famille d'instruments verbalisés dans le cadre de la consigne "instrument". Ceci correspond donc ici aussi à l'effet attendu. Pourtant, si on considère chacun des 6 noms de famille d'instruments produits, ainsi que 4 autres sous-groupes de locutions qui peuvent leur être associés (voir tableau A6.III.2.), 5 de ces 10 catégories de locutions présentent des différences significatives en fonction de la consigne suivie ( $\chi^2(3)>9.06$ ,  $p<.01$ ), toutes dues à des sur-représentations de ces locutions dans la consigne "instrument". Ceci représente 50% des locutions testées dans ce sous-groupe, ce qui sera la part la plus importante observée.

Parmi les 12 noms "d'instruments visés" par leur concepteur, 6 reprennent des noms d'instruments déjà testés plus haut. Nous n'allons donc considérer que les 3 autres noms d'instruments (corde frottée, vibraphone et piano échantillonné) tous produits spontanément dans l'expérience 3, ainsi que les trois noms suggérant des mélanges (mélange hautbois / clavecin, vibraphone / trombone, corde frottée / piano). Un seul des noms de mélanges ayant été produit spontanément (avec une occurrence), nous y avons ajouté trois catégories de locutions que nous avons jugées proches (voir tableau A6.III.3.) qui associaient par exemple la notion de frappe et de souffle. Enfin, nous avons testé aussi les 3 autres noms d'instruments de musique, produits uniquement dans l'expérience 3 (non présents dans l'expérience 1). Pour l'ensemble de ces 11 catégories de locutions produites, qui représentent au total 61 occurrences, la répartition globale n'est pas significativement différente en fonction de la consigne ( $\chi^2(3)=1.62$ ,  $p>0.05$ ). De plus, une seule des 11 locutions mène à une différence de production significative en fonction de la consigne ( $\chi^2(3)=9.10$ ,  $p<0.05$ ), due à une sur-représentation du nom de vibraphone produit dans la consigne "qualité", ce qui ne correspond pas du tout aux résultats attendus.

#### Noms de mode d'excitation, de matériaux et d'association des deux

Parmi les 12 locutions de mode d'excitation recherchées dans les verbalisations, ainsi que les 5 catégories de locutions utilisant un des verbes testés de manière plus générale que sous la forme initiale, aucune des différences de nombre d'occurrences produites en fonction de la consigne suivie ne sont significatives ( $\chi^2(3)<3.6$ ,  $p>0.05$ ). On n'observe donc pas d'effet de la consigne sur la production de ce type de verbalisations.

Sur l'ensemble des 17 noms de matériaux correspondant aux portraits de l'expérience 3, 11 ont été produits ici. Nous y avons associé 16 locutions proches de ces portraits (voir tableau A6.III.5). Ceci représente au total 200 occurrences, dont la répartition est significativement différente en fonction de la consigne ( $\chi^2(3)=8.52$ ,  $p<.05$ ). On observe en effet une sur-représentation de noms d'instruments verbalisés dans le cadre de la consigne "instrument". Nous attendions une sur-représentation de ces verbalisations lors de la consigne "source", ce qui n'est pas observé. On peut noter tout de même que la consigne "instrument" oriente, elle aussi, vers une manière de désigner la source sonore. Pourtant, si on considère chacun des 17 noms de matériaux, aucun ne présente de différences de productions significatives en fonction de la consigne. Ces différences n'apparaissent en effet que pour 5 des 16



locutions associées. Ceci correspond notamment à une sur-représentation des locutions contenant *du cuivre* ou l'adjectif *cuivré* dans la consigne "qualité", *du cuivre* ou *famille cuivre* dans la consigne "instrument", *du souffle* (éventuellement associé à *du vent*) dans la consigne "libre" et *un son synthétique* dans la consigne "qualité". Ces différences correspondent donc à nos attentes concernant les consignes "qualité" et "instrument".

Sur l'ensemble des 25 noms de sources sonores correspondant aux portraits de l'expérience 3, 13 seulement ont été produits ici (ce qui est dû à leur spécificité). Nous y avons associé 7 locutions proches ou composant ces portraits (voir tableau A6.III.6). Ceci représente au total 161 occurrences, dont la répartition n'est pas significativement différente en fonction de la consigne ( $\chi^2(3)=5.19$ ,  $p>.05$ ). En fait, sur les 20 catégories de locutions produites, 3 seulement aboutissent à des différences de production significatives en fonction de la consigne. *On étouffe le son* est produit plus souvent dans la consigne "libre" ( $\chi^2(3)=14.00$ ,  $p<.01$ ), l'association des catégories *deux* et *taper/frapper*, associée au portrait *double percussion* est produite plus souvent dans la consigne "qualité" ( $\chi^2(3)=11.40$ ,  $p<.01$ ) et *on pince une corde* est produite moins souvent dans la condition "libre" ( $\chi^2(3)=9.20$ ,  $p<.05$ ). Ces différences ne sont pas facilement interprétables.

#### Adjectifs

Plusieurs groupes d'adjectifs ont été testés, correspondant aux différents sous-groupes de portraits de qualités : qualités simples construites à partir d'adjectifs produits dans l'expérience 1 pour comparer les sons 2 à 2, autres adjectifs simples, associations de 2 qualités et de 3 qualités.

Pour ce qui est de l'ensemble des 23 adjectifs produits dans l'expérience 1 pour comparer les sons 2 à 2, seul *diffus* n'a pas été ici produit spontanément. Nous y avons associé 8 locutions proches (voir tableau A6.III.7) ainsi que 7 de ces adjectifs présents sous forme négative (par exemple *pas clair*). Ceci représente au total 604 occurrences, dont la répartition est significativement différente en fonction de la consigne ( $\chi^2=141.95$ ,  $p<.01$ ,  $df=3$ ). On observe en effet une importante sur-représentation de production de ces adjectifs dans le cadre de la consigne "qualité", et une sous-représentation de cette production dans le cadre des deux consignes "instrument" et "source". Ceci correspond tout à fait à nos attentes. Pourtant, si on considère chacune des 37 locutions testées, seules 10 d'entre elles présentent des différences de productions significatives en fonction de la consigne. Il s'agit des adjectifs *doux*, *rond*, *sec*, *sourd*, *nasillard* (associé à *nasal*) et *résonnant* (associé à *résonne*, *résonance*...) et des locutions *une attaque*, *pas d'attaque* ainsi que l'ensemble des locutions parlant de *l'attaque*, toutes ces différences étant dues à une sur-représentation de ces locutions dans la consigne "qualité".

Pour ce qui est de l'ensemble des 17 autres adjectifs simples produits dans l'expérience 1 pour catégoriser les sons, 14 ont été produits ici spontanément. Nous y avons associé 4 locutions proches (voir tableau A6.III.8) ainsi que 5 de ces adjectifs présents sous forme négative (par exemple *pas clair*). Ceci représente au total 185 occurrences, dont la répartition est significativement différente en

fonction de la consigne ( $\chi^2(3)=69.08, p<.01$ ). On observe ici aussi une importante sur-représentation de production de ces adjectifs dans le cadre de la consigne "qualité", et une sous-représentation de cette production dans le cadre de la consigne "source". Pourtant, si on considère chacune des 26 locutions testées, seules 5 d'entre elles (soit 19%) présentent des différences de productions significatives en fonction de la consigne. Il s'agit des locutions *cuivré* (ainsi que toutes les locutions contenant l'idée de *cuivre*), *lointain*, *DEUX (double)*, dont les différences sont dues à une sur-représentation dans la consigne "qualité", et de *étouffé*, dont la différence est due à une sur-représentation dans la consigne "libre".

Les adjectifs qui ont été testés dans les portraits doubles (construits avec 2 adjectifs) reprennent pour la plupart des adjectifs testés dans les 2 sous-groupes précédents. Il faut tout de même y ajouter 6 autres, dont 4 ont été produits ici. De plus, sur les 22 associations de 2 adjectifs, 7 ont été produites spontanément ici. L'ensemble de ces locutions représente 36 occurrences, dont la répartition est significativement différente en fonction de la consigne ( $\chi^2(3)=36.00, p<.01$ ), ce qui est à nouveau dû à une sur-représentation de production de ces adjectifs dans le cadre de la consigne "qualité". Pourtant, aucune de ces locutions prises séparément n'aboutit à une différence de production significative en fonction de la consigne. Les portraits construits à partir de 3 qualités reprennent les adjectifs simples du premier sous-groupe et n'ont jamais été produits ensemble spontanément.

#### Conclusion

On observe donc globalement les résultats attendus. La consigne de décrire les sons en donnant si possible un nom d'instrument induit la production de plus de noms d'instruments et de noms de famille d'instruments (mais aussi de noms de matériaux) qu'avec les autres consignes de descriptions. De même la consigne de décrire en donnant des qualités des sons induit plus de productions d'adjectifs qualifiant les sons que les autres consignes. Par contre, les descriptions des sons sous forme de mode d'excitation et de source sonore ne dépendent pas de la consigne. Ces effets, significatifs au niveau de l'ensemble des locutions de ces groupes, ne se retrouvent qu'au maximum dans 50% des locutions prises séparément.

### III.2. Comparaison de la consigne "Instruments" avec l'expérience 1

Le codage dans Stone des descriptions de chacun des 12 sons par les participants de l'expérience 3 n'a donc été effectué de manière exhaustive que pour une de 4 consignes, celle correspondant au fait de "décrire les sons en donnant si possible un nom d'instrument". Ceci représente 240 entrées (12 sons décrits par 20 sujets) qui ont été codées selon le même principe que celui utilisé pour les verbalisations de l'expérience 1 (voir §I.2). La seule différence est qu'une production (un objet dans Stone) correspond à la description d'une paire de sons dans l'expérience 1 et d'un seul son dans l'expérience 3.

On retrouve sur le premier niveau de description une répartition des locutions produites par les participants en 4 grandes catégories, selon qu'ils décrivent : le son (présent dans 27% des descriptions), une source possible (dans 95% d'entre elles), qu'ils parlent d'eux-mêmes (dans 55% des descriptions) ou qu'ils effectuent des comparaisons avec les autres sons du groupe (dans 21% des descriptions). On constate donc ici que les locuteurs parlent principalement de ce qui a produit le son (le geste ou l'objet vibrant), beaucoup plus que lors de l'expérience 1 (dans 60% des productions), ce qui est en partie lié à la consigne analysée ici (voir §III.1.2). On peut noter par contre que même dans le cas où on oriente les locuteurs vers la description d'une source sonore des sons (sous forme de noms d'instruments de musique ici), on obtient des descriptions des sons dans 27% des productions. Malgré le fait que les locuteurs devaient cette fois-ci décrire chacun des 12 sons séparément, ils effectuent des comparaisons avec d'autres sons, soit ceux qu'ils ont précédemment écoutés, soit des sons qu'ils connaissent et qu'ils considèrent comme référence, dans 21% de leurs descriptions. Enfin, nous avons noté ici que dans 55% des descriptions, la formulation du locuteur le met lui-même en scène. Ce pourcentage est plus important que lors de l'expérience 1 (23%), mais c'est sans doute dû à une plus grande finesse de codage ici. En effet, nous avons tenu compte ici des analyses linguistiques effectuées par David (1997) sur les marques de la personne dans les descriptions d'odeurs et de sons comme une mesure de la relation que les locuteurs entretiennent avec l'objet qu'ils décrivent. Ceci sera détaillé plus loin.

### III.2.1. Ce qui a produit le son

Des descriptions de ce qui aurait pu produire le son sont donc présentes dans 229 des 240 descriptions analysées. Il s'agit, pour 93% d'entre elles, de descriptions d'objets (ce qui est sans doute induit par la consigne de décrire en donnant si possible un nom d'instruments ou de familles d'instruments), mais aussi de descriptions d'actions effectuées qui auraient pu produire le son (dans 19% des cas) et de noms de matériaux dans 10% des cas.

Sur les 43 actions décrites verbalement par les participants, 38 décrivent des actions servant à l'émission du son, 17 de manière soutenue (avec notamment *on souffle* dans 11 essais) et 25 de manière "événementielle" (12 *frappés*, 9 *on pince*). Par contre, sur 22 noms de matériaux cités, on trouve 13 noms de matériaux fabriqués (4 *électronique*, les autres sont très variés : *synthétique*, *artificiel*, *plastique*...) contre 9 naturel (7 *métal*, 3 *végétal*, 1 *air*). Il pourrait être intéressant de confronter ces verbalisations aux sons décrits.

Nous avons aussi noté des gestes effectués par les participants quand ils ne peuvent pas exprimer leur pensée (souvent quand ils cherchent un mot ou se sentent en difficulté). Ces gestes correspondent souvent à des actions qui auraient pu produire le son : geste d'archet qui frotte, geste de pincer une corde, geste de percussion, geste de scie, mouvement des doigts comme sur un clavier. Bien qu'ils ne soient présents ici que dans 2% des descriptions, nous analyserons tous ceux produits dans l'ensemble des 4 consignes de descriptions dans le paragraphe suivant (§III.4). En effet, leur

présence souligne la primauté d'une identification d'une source sonore possible lorsqu'on écoute un son.

Le codage effectué ici nous a posé deux questions. D'une part, on s'interroge sur le maintien de la distinction objet / matériaux, notamment pour les locutions *vent*, *bois* ou *cuivre*. En effet, vu la consigne analysée ici, ces locutions renvoient plutôt à une famille d'instruments qu'à un matériau, même si dans l'absolu (et notamment pour d'autres types de consignes), l'ambiguïté reste. D'autre part, se pose aussi le problème de distinguer les qualités de l'objet qui a produit le son des qualités du son lui-même, notamment avec des locutions comme *artificiel* ou *électronique*.

Nous allons maintenant détailler la partie de l'arborescence qui rassemble les noms d'objets et qui représente la branche majoritaire ici. En effet, sur les 212 descriptions pour lesquelles un nom d'objet a été verbalisé, 199 concernent des noms d'instruments, contre 44 des noms d'objets indéfinis et 15 des objets courants. Nous avons regroupé sous la catégorie *objets indéfinis* des locutions comme *ça* (26 occurrences), produites quand le locuteur essaye de définir un objet sans forcément y arriver (*ça, c'est un...*), mais aussi 7 locutions de *trucs*, 3 de *choses*, 5 de *c'est quoi ?*, qui réfèrent toutes à une description sous forme d'objet, bien qu'il ne soit pas précisé. La catégorie *objets courants* rassemble notamment les *sonneries* ainsi qu'un amas très hétéroclite d'*ordinateur*, *tubes*, *carafe*, *règle*, *lattes*, *barres*, *planches*... On retrouve ici le problème de savoir où placer les locutions de *corde*, *une corde*, *la corde*. En effet, selon le contexte, elles peuvent désigner un objet courant ou une famille d'instruments de musique.

Pour ce qui est des locutions catégorisées comme *instruments*, elles sont présentes dans la majorité des verbalisations d'objets ayant pu produire le son et 83% des verbalisations totales, ce qui est normal compte tenu de la consigne. Pour construire le tableau suivant, nous avons regroupé les verbalisations implicites et explicites des principales catégories de noms d'instruments de musique (par exemple dans la catégorie *instrument à corde*, les locutions littérales d'*instrument à cordes* sont considérées comme explicites, alors que celle de *clavecin*, incluse dans une branche sous cette catégorie, est considérée comme implicite)

**Tableau 6.7.** Répartition des catégories de noms d'instruments à vent en fonction du son décrit.

	nb	TPT	TBN	CLA	CRA	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO	2 p
Total	199	17	17	15	14	16	18	17	16	19	19	17	14	1.86
I. électronique/électrique	20	2	2	1	3	1	3	0	2	0	2	2	2	6.40
Instrument à cordes	106	2	<b>0</b>	<b>0</b>	1	12	11	17	14	<b>18</b>	16	9	6	58.38 *
Instrument à touche	43	1	1	0	2	1	5	<b>9</b>	8	6	2	4	4	26.49 *
Instrument à vent	76	<b>15</b>	<b>16</b>	<b>15</b>	9	5	0	0	1	0	2	6	7	66.42 *
Instrument indéfini	12	2	0	0	3	2	0	0	1	0	1	1	2	12.00
Instrument à percussion	8	0	0	0	0	0	<b>6</b>	0	0	0	0	1	1	48.83 *

avec \* :  $p < .01$ , † :  $p < .05$ , dl=11. En gras, les effectifs donnant les plus grandes participations au 2.

Ainsi, l'ensemble des *noms d'instruments* est également réparti sur l'ensemble des sons décrits, par contre, si on regarde les sous-catégories, il n'y a plus que les noms d'*instruments électriques ou électroniques* ainsi que les noms d'*instruments indéfinis* qui présentent une répartition égale sur l'ensemble des 12 sons. En effet, les noms d'instruments à cordes sont plus souvent produits pour décrire le son CVC et jamais pour décrire les sons TBN et CLA ( $\chi^2(11)=53.38, p<.01$ ), les noms d'instruments à touches sont plus souvent produits pour décrire le son PNE ( $\chi^2(11)=26.49, p<.01$ ), les noms d'instruments à vent sont plus souvent produits pour décrire les sons TBN, TPT et CLA (mais pas le son CRA) ( $\chi^2(11)=66.42 p<.01$ ). Enfin, les noms d'instruments à percussion sont plus souvent produits pour décrire le son VBR ( $\chi^2(11)=48.83, p<.01$ ). Tout ceci correspond bien aux instruments que ces sons imitent.

Pour ce qui est de la répartition de la production de ces catégories de noms d'instrument en fonction de la compétence musicale des locuteurs, on obtient en général pas de différence significative entre musiciens et non-musiciens ( $\chi^2(1)<2.22 ; p>.05$ ) sauf pour les noms d'instruments à vent et d'instruments à percussion (respectivement  $\chi^2(1)=6.37$  et  $4.50 p<.05$ ), produits plus souvent par des locuteurs musiciens (respectivement 64% et 88% de ces locutions sont dues aux musiciens).

Il faut noter que ces regroupements en grandes catégories posent parfois des problèmes. Par exemple un piano est à la fois un instrument à corde, à touche et à percussion. Les locutions de *piano* n'ont été ici indexées que dans les deux premières grandes catégories, bien qu'il ne soit pas du tout évident que tout cela soit à l'esprit de toute personne utilisant le mot *piano* ! De même, un orgue est un instrument à touche et à vent, mais pas l'orgue électronique est un instrument à touche et électronique. Il est difficile d'être sûr de ce qu'un locuteur sous-entend dans notre contexte quand il dit *orgue*.

Par contre, les différences de niveau dans l'arborescence des locutions utilisées pour décrire un son sont intéressantes à étudier en fonction de la compétence des sujets et à comparer à la notion de niveau de base, sur et sous ordonnés de Rosch (1976). Par exemple, le tableau ci-dessous étudie les deux sons TBN et TPT en fonction de leur niveau de description.

Tableau 6.8. Locutions utilisées pour décrire les sons TBN et TPT en fonction de la compétence des sujets (M : musiciens / nm : non-musiciens).

SONS Compétence	TPT			TBN		
	tous	M	nm	tous	M	nm
TOTAL VENT	45	24	21	54	26	28
famille des vent	17	6	11	16	4	12
TOTAL CUIVRE	32	18	14	40	23	17
famille des cuivres	7	7	0	16	13	3
TOTAL INSTRUMENTS CUIVRES	32	17	15	32	17	15
trompette	25	12	13	16	6	10
trombone	4	4	0	10	7	3

Avec en Majuscule les catégories de locutions, en minuscule les occurrences).

Si on considère les locutions décrivant directement les 2 sons concernés soit comme (*instrument de la famille des vents*, soit comme (*un*) *cuivre*, soit à partir d'un des noms d'instruments appartenant à la famille des cuivres (*trompette, trombone*, mais aussi *tuba, cor, clairon et corne de brume*), on remarque que la différence de répartition sur l'ensemble des sujets ne dépend pas du son concerné : TPT ou TBN ( $\chi^2(2)=3.03$   $p=.22$ ). Par contre les différences de comportement entre musiciens et non-musiciens sont très significatives. Les non-musiciens décrivant moins souvent le son TPT comme un *cuivre* ( $\chi^2(1)=8.35$   $p=.01$ ) et les musiciens décrivant beaucoup plus souvent le TBN comme un *cuivre* et moins souvent comme un *vent* ( $\chi^2(1)=23.46$   $p=7.10^{-6}$ ). Les noms de *trompette* et de *trombone* ne sont pas utilisés de manière symétrique. Si l'ensemble des sujets utilise plus souvent le nom de *trompette* pour le TPT et celui de *trombone* pour le TBN, ce dernier est beaucoup plus souvent appelé *trompette* que l'inverse n'est vrai. Cette différence dépend beaucoup de la compétence musicale : les non-musiciens utilisant moins souvent le nom de *trombone* que les musiciens ( $\chi^2(1)=5.75$   $p=.01$ ). Ils semblent se servir donc du nom de *trompette* comme de celui d'un instrument générique.

### III.2.2. Concerne le sujet

Nous avons vu que dans 132 descriptions de sons (soit 55% d'entre elles), le sujet dit quelque chose de lui ou de son attitude vis à vis de sa perception quand il décrit les sons. Nous avons en effet noté dans les locutions utilisées par les sujets pour décrire les sons d'une part, les expressions de *difficultés* rencontrées par les participants dans cette tâche (présentes dans 52% des descriptions), ou au contraire de *facilité* (dans 3% des cas), ainsi que la manière dont ils s'impliquaient ou non dans la description qu'ils donnaient (nous étudierons cette partie dans le paragraphe suivant). Ceci nous permet en effet d'une part, de donner un ordre de grandeur du type de difficultés rencontrées par les sujets et d'autre part, de comparer nos données à celles étudiées par David (1997) sur les marques de la personne présentes dans les descriptions comparées d'odeurs et de sons.

#### Difficultés

L'ensemble des 126 descriptions présentant des expressions de difficultés des locuteurs, qui représentent donc 52% des réponses totales, a été répartie en plusieurs catégories. Nous avons en effet séparé les locutions exprimant des difficultés de manière explicite (*c'est difficile*) d'expressions plus implicites laissant entrevoir la notion d'arbitraire (*on va dire..*), d'hésitation (*euh...*) d'ignorance (*je ne sais pas*), de prudence (*on dirait que...*), des impressions (*j'ai l'impression que...*), mais aussi des questionnements (*c'est quoi ?*) ou des exclamations (*ça alors !*). Le tableau 6.9 reprend la répartition des locutions dans ces différentes sous-catégories.

Tableau 6.9. Répartition des sous-catégories de locutions d'expression de difficulté des locuteurs

	nb réponses	% (des difficultés)	% (des descriptions)
Arbitraire ( <i>on va dire..., mettons</i> )	21	17	9
Difficile ( <i>c'est difficile</i> )	18	14	7
Exclamation ( <i>ça alors !</i> )	18	14	7
Hésitation ( <i>euh...</i> )	21	17	9
Ignorance ( <i>je ne sais pas</i> )	17	13	7
Impression ( <i>j'ai l'impression que</i> )	4	3	2
Négation ( <i>c'est pas ça !</i> )	12	9	5
Prudence ( <i>on dirait que...</i> )	44	35	18
Questionnement ( <i>c'est quoi ?</i> )	38	30	16
Percevoir ( <i>ça évoque</i> )	6	5	2
Non trié	8	6	3

On note en particulier beaucoup d'expressions de prudence et plusieurs de questionnement (respectivement dans 18 et 16% des verbalisations totales). Apparaissent ensuite les impressions d'arbitraire, les hésitations, puis la difficulté et les exclamations, enfin les aveux d'ignorance. Ceci nous permet de souligner le fait que décrire des sons est une tâche difficile, déroutante.

#### Marques de la personne

L'intérêt de cette partie de l'arborescence était de comparer nos données à celles recueillies par David (1997) sur les marques de la personne présentes dans les descriptions comparées d'odeurs et de sons. Elle note en effet que le sujet donne une idée différente de sa notion de l'objectivité de ce qu'il dit selon qu'il s'exprime en disant : *c'est un violon*, *on dirait que c'est un violon*, *je dirais que c'est un violon*, *ils appellent ça un violon*. En effet, s'il dit *c'est un violon*, l'objectivité de ce qu'il dit ne présente pas de doute pour lui, alors que dire *je dirais que c'est un violon*, exprime aussi le fait que c'est son point de vue personnel et qu'il n'est peut-être pas partagé par d'autres personnes. Nous avons donc regroupé dans la catégorie *mode d'expression*, des marques de la personne que nous avons séparées en *individuel / collectif* suivant que le locuteur s'impliquait personnellement dans la description (*je, mon, mes...*) ou qu'il s'impliquait de manière collective, (*on, nous, vous, notre, votre, nos...*). Nous avons d'autre part séparés les pronoms renvoyant au locuteur par rapport à ce dont "on" parle (*je, tu, nous, vous, on, moi, toi,...* / *il, elle, lui, le, la, ...*). Au niveau des pronoms, on observe la présence de 43 *je* (soit dans 18% des verbalisations) 7 *me* ou *moi* (*pour moi*), 19 *on* (mélange de *ce qu'on appelle* et *on pince la corde*, qui ne sont sans doute pas à mélanger), 14 *il(s)* (mélange de *il y a*, de *il faut que* et de *il sonne comme ça*), 1 *elle*, 1 *nous*, 3 *vous*, 2 *quelqu'un*.

Les verbalisations où le sujet s'implique personnellement sont donc assez présentes (dans environ 30% des descriptions), ce qui est interprété comme un manque d'objectivité ou un doute sur ce qu'on est en train de dire. Cela introduit la notion de relativité de ce qu'on est en train d'exprimer et rejoint les résultats de David (1997).

#### **III.2.3. Concerne le son**

Les verbalisations décrivant directement le son entendu ne sont présentes ici que dans 66 descriptions (27% des cas, contre dans 98% des cas dans l'expérience 1). On observe ainsi la présence de mots

*lexicalisés* (mots désignant un son) dans 15% des verbalisations : 3 onomatopées, 34 noms de sons (2 *bruit*, 31 *son*, 2 *timbre*, 1 *harmonique*, 1 *onde*) et une production contenant des noms déverbaux (*frottement* et *pincement*). Sachant qu'on considère en général qu'il y a peu de mots lexicalisés pour décrire les sons, leur présence dans 15% des descriptions est tout de même notable, bien que dans 31 cas sur 35 il y a au moins une occurrence de *son*.

On observe relativement peu de locutions de *qualités* "intrinsèques" (dans 19 descriptions soit 8% de l'ensemble des cas) : 13 dans le domaine de l'audition (5 *aigu*, 1 *grave*, 2 *faible*, 2 *fort.*), 5 métaphores faisant appel au toucher ou à la vision (2 *doux*, 1 *son très fin*, 1 *souple*, 1 *creux*, 1 *la première image...*) et 2 lieux d'origine (*extra européen* et *japonais*).

De même, on observe relativement peu de descriptions de *l'évolution du son* (dans 17 descriptions soit 7% des cas) dont essentiellement des mentions des limites temporelles (13 sur 17, en général *l'attaque*) 2 de *direction* et 3 de *durée*.

On observe enfin quelques jugements de valeur (dans 20 descriptions soit 8% des cas), surtout *bizarre* (dans 75% d'entre elles) 4 *agréables*, 2 *désagréables* et peu de mention du résultat entendu (*ça fait comme ça*, *ce qu'on entend à la fin*) et une seule imitation dans ce corpus.

La consigne de donner un nom d'instrument ou de famille d'instruments à donc fortement limité ce type de productions, sans toutefois le faire disparaître complètement.

#### **III.2.4. Comparaison avec un autre son**

Le son décrit est comparé à un autre (de façon plus ou moins explicite) dans 21 % des verbalisations. En général, il est comparé à un autre son de l'expérience (70% des cas de comparaison). Sur les 35 cas, il s'agit de comparaisons notant des différences dans 51% des cas (*plus quelque chose que l'autre*), d'expressions de ressemblances dans 60% des cas (*aussi... que l'autre*). Quand il est fait référence à des sons mémorisés (dans 32% des comparaisons avec un autre son), c'est souvent de manière plus implicite (*comme un son de piano*, *ce son est connu*, *ça ressemble à*).

### III.3. Des gestes quand on ne trouve pas comment dire

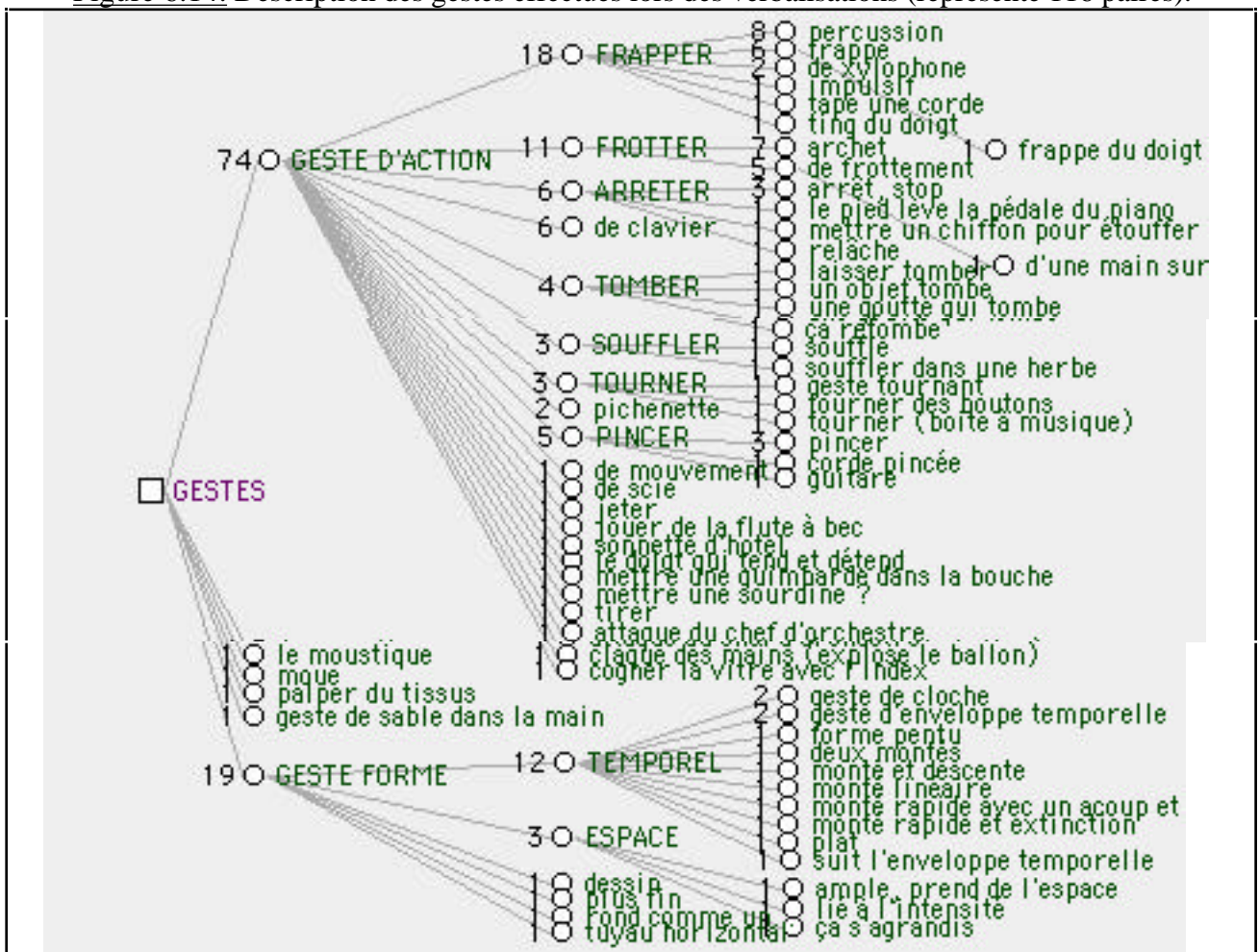
On retrouve donc les grandes catégories de verbalisations découvertes dans les verbalisations libres de l'expérience 1, avec tout de même une grande nouveauté due au mode de recueil des données. Les sujets ne sont plus seuls dans une chambre insonorisée face à un ordinateur et parlant dans un micro. Cette fois-ci il y a un expérimentateur à côté d'eux qui note leurs verbalisations pendant la première phase de l'expérience. Et on observe alors de nombreux moments où les participants interrogent l'expérimentateur et d'autres où ils accompagnent ou remplacent leurs verbalisations par des gestes (présents dans 116 descriptions sur 960 sons décrits, soit dans 12% des cas) qui ont été notés et codés pour l'ensemble des 4 consignes.

#### **III.3.1. Quels gestes ?**



L'ensemble des gestes qui ont été notés peuvent être regroupés dans deux grandes catégories selon qu'ils miment une action qui aurait pu produire le son, ou qu'ils décrivent la forme du son (souvent l'enveloppe temporelle). L'arborescence construite est représentée figure 6.14. Les gestes mimant des actions sont principalement constitués de mimes d'actions qui sont des modes de productions possibles de sons (frapper, frotter, faire tomber, souffler, pincer...). Ils pourront donc être comparés aux sons à propos desquels ils ont été effectués. Les gestes qui décrivent la forme du son sont principalement constitués de descriptions temporelles comparables à l'enveloppe temporelle des sons, par exemple les gestes décrits par "un geste de cloche", "2 montées", "montée et descente", "plat"...

Figure 6.14. Description des gestes effectués lors des verbalisations (représente 116 paires).



### III.3.2. Les gestes dépendent-ils des sons ?

Nous avons analysé la répartition sur les 12 sons décrits des gestes notés, afin de voir si on pouvait trouver un lien entre le type de gestes et le type de sons écoutés (voir tableau ci-dessous).

Globalement, l'ensemble des gestes présents lors des descriptions des participants sont plus nombreux quand le son décrit est le son CFT et moins nombreux quand il s'agit du TPT ( $\chi^2(11)=21.34, p<.05$ ).

Tableau 6.10. Nombre de gestes notés en fonction des sons décrits et de leur "contenu".

Sons	Total	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO	$\chi^2$	p
GESTES	113	3	6	6	10	20	10	7	10	11	9	13	8	21.34	†

GESTES de forme	19	1	3		2	4	1		2		2	2	2	5.93
Forme temporelle	12		3		1	3	1		1		1	1	1	8.00
GESTES d'action	74	3	3	4	4	12	8	6	6	9	6	8	5	12.92
Geste Frapper	18		1				<b>6</b>	3	2		1	4	1	19.83 †
Geste Frotter	11	1				<b>10</b>								99.18 *
Geste Arrêter	6				2	1		1		2				10.00
Geste Pincer	5							1		2	2			12.85
Geste Souffler	3	2											1	14.50
Geste Faire tomber	4								1	1	1	1		5.33
Geste Clavier	6			2					1	1			2	10.00

Avec \* :  $p < .01$ , † :  $p < .05$ .

Les gestes de forme étant peu nombreux (19 en tout) et très variés, il est difficile d'analyser leur lien avec les sons décrits de façon générale. Par contre, on remarque que certains gestes d'actions sont produits significativement plus souvent en présence d'un des 12 sons. C'est particulièrement le cas des gestes de "frotter", notés quasiment essentiellement en présence du CFT ( $\chi^2(11)=99.18$ ,  $p < .01$ ), et des gestes de "frapper", présents plus souvent lors de l'écoute du VBR ( $\chi^2(11)=19.83$ ,  $p < .05$ ). Ceci correspond à chaque fois au mode de production de ce son s'il n'était pas synthétisé.

### III.3.3. Les gestes dépendent-ils de la consigne ?

Sur l'ensemble des 113 occurrences de gestes notées lors de cette expérience, la répartition est significativement différente en fonction de la consigne ( $\chi^2(3)=53.97$ ,  $p < .01$  voir tableau 6.11). On observe en effet une sur-représentation de gestes dans le cadre de la consigne "libre" et une sous-représentation de ces gestes dans le cadre de la consigne "instrument". De plus, parmi les 10 sous-catégories de gestes testés, 4 présentent des différences de production significatives en fonction de la consigne suivie ( $\chi^2(3) > 10.89$ ,  $p < .05$ ). En effet, l'ensemble des gestes de forme (et en particulier ceux décrivant une forme temporelle) sont sur-représentés lors de la consigne "qualité", l'ensemble des gestes d'actions (et en particulier le geste de frapper) étant par contre sur-représentés lors de la consigne "libre".

Tableau 6.11. Effet de la consigne de description sur les gestes produits (nombre d'occurrences produits dans chaque consigne et test de la différence)

Consigne	Total	instrument	libre	qualité	source	$\chi^2$ p
GESTES	113	<b>6</b>	<b>51</b>	44	12	53.97 *
Gestes de FORME	19		7	<b>12</b>		12.13 *
Forme temporelle	12		3	<b>9</b>		12.00 *
Gestes d'ACTION	74	7	<b>37</b>	19	11	28.55 *
Geste frapper	18	1	<b>10</b>	5	2	10.89 †
Geste frotter	11	1	5	3	2	3.18
Geste arrêter	6		4	2		4.33
Geste pincer	5	3	2			2.90
Geste souffler	3		1	1	1	0.25
Geste faire tomber	4			2	2	2.00
Geste clavier	6	1	2	1	2	0.67

avec \* :  $p < .01$ , † :  $p < .05$ , rien si la différence de répartition n'est pas significative

On peut donc conclure que lors de la consigne de description "libre", les locuteurs étant moins dirigés ou guidés vers un type de description des sons font plus souvent appel à des gestes mimant l'action effectuée pour produire le son, afin de compléter ou d'élaborer leurs descriptions. Lors de la consigne de description de qualités sonores, ils font par contre plus souvent appel à des gestes décrivant la forme des sons, en particulier leur forme temporelle.

#### **III.3.4. Conclusion**

La présence de ces gestes, produits pour accompagner ou faciliter la description verbale quand le locuteur est en difficulté, nous permet de souligner d'une part l'importance du fait qu'une source sonore possible des sons est associée directement à la perception auditive, notamment sous forme de gestes pouvant permettre l'émission du son, et d'autre part l'importance du déroulement temporel comme moyen de décrire un son. En effet, on retrouve ces deux points à la fois dans les descriptions verbales et non verbales des sons. Ceci nous renvoie à la nature même de ce qu'est un son : une vibration produite par une action sur un matériau.

#### III.4. Quelques remarques faites par les sujets

Les locuteurs ont parfois donné des remarques générales sur l'ensemble des sons ou de la tâche qu'ils avaient à effectuer. Mais ils ont aussi parfois essayé de comprendre comment ils avaient décrit les sons, faisant des remarques sur d'éventuelles différences de facilité à identifier ou à donner des qualités, ou sur le fait que les descriptions proposées pouvaient modifier leur perception. Nous allons donc décrire brièvement quelques remarques recueillies. Les numéros des sujets sont entre parenthèses.

##### **III.4.1. Sur les sons en général**

On trouve ainsi beaucoup de remarques sur le caractère synthétique ou très court des sons, comme première impression générale, après avoir écouté les 12 sons et avant même de commencer à les décrire. Par exemple : "je ne sais pas comment tu les a enregistré mais les sons me semblent tous synthétiques" (nd6), "ils passent par un filtre qui les rend à la fois... coupe la vibration et donne du corps, si je puis dire" (ns11) ; "ça résonne pas : j'ai l'impression que ça s'arrête net (imite)" (nd7) ou "ça dure pas assez longtemps vos sons" (ns8). Cet aspect synthétique les rend souvent assez désagréables à écouter. Ainsi, après avoir décrit les 12 sons, en lisant les portraits qui lui étaient proposés, le sujet nd7 s'arrête sur *agréable* et réagit : "il y en a pas beaucoup !". Ces remarques générales sur les sons peuvent aussi amener les participants à les séparer en deux groupes, par exemple : "j'ai l'impression de séparer les sons en deux groupes : ceux qui sont des cordes, assez vraisemblables, qui sont purs, francs... et les autres assez indéfinissables (moue) qui ressemblent à des cuivres mais "étouffé" ? pas vraiment..." (nd6). Enfin, certains sons apparaissent déjà comme particuliers, ainsi, en lisant les portraits d'adjectifs qui lui sont proposés, le sujet nd3 s'arrête à *double*

: "ah, celui là, je sais déjà !". La notion de spécificité ajoutée pour comprendre la représentation perceptive des sons semble donc fondée.

#### **III.4.2. Sur la difficulté de la tâche**

On trouve aussi beaucoup de remarques sur la difficulté de la tâche, par exemple "c'est pas évident", "c'est très difficile" (nq10), "c'est une guitare ? c'est toujours la même note ? ça va être dur !" (nd3). Mais aussi sur la façon dont la personne procède ou pense procéder. Par exemple, le participant ns8, travaillant dans l'art graphique, dit "décrire son image mentale" pour chaque son écouté, par exemple "issus d'un tube, me paraît sortir d'un tuyau". Le participant mq5 dit : "je me base essentiellement sur l'évolution dans le temps. Ce qui me ressort de plus c'est la partie enveloppe (ils ont tous des enveloppes différentes en fait)", et conclut plus loin : "donc je dirais abrupte mais je qualifie toujours l'évolution dans le temps et jamais le timbre en fait (j'ai vraiment du mal à qualifier un timbre)".

#### **III.4.3. Sur la description de qualités / identification de la source**

Plusieurs personnes ont été étonnées de leurs réactions vis à vis de l'identification des sons. Par exemple, le sujet md8, musicien professionnel soumis à la consigne de décrire librement chacun des sons remarque : "ça ressemble à des instruments de musique... je le dis en terme... d'instruments. C'est le premier truc qui me vient, l'instrument". Un autre musicien professionnel, le seul à avoir participé à deux des sous-groupes de portraits (mq2 (qualité simple) et ms7 (source)), déclare la deuxième fois, en répondant à la consigne de donner des qualités des sons : "c'est beaucoup plus intéressant, beaucoup plus subjectif, sensuel (que de décrire des modes de production), je m'étais laissé piéger à donner des noms d'instruments". Enfin, une musicienne - psychoacousticienne (mq4), soumise à la consigne de décrire des qualités remarque : "en fait, quand je les reconnaît, j'ai plus de mal à donner des adjectifs", ce qui se retrouve en effet dans ses verbalisations (CRA : "bref, nasillard aussi, rugueux ?" ; CFT : "là c'est les cordes frottées, donc euh..., il y a du mouvement ("geste")...., ça bouge... enfin...")

#### **III.4.4. Sur les liens langage-perception**

Enfin, nous voulions noter une remarque du sujet nq10, néophyte, qui remarque après avoir associé les sons aux différents portraits constitués d'adjectifs simples : "c'est curieux, j'ai vraiment l'impression que les sons ne sont pas les mêmes en fonction du mot donné, comme si le mot imposait un éclairage particulier, et faisait scintiller une partie différente du même son". Cette remarque renvoie en effet directement aux questions sur la possibilité d'une perception pure, indépendante du langage et des interactions entre décrire vraiment le contenu de nos perceptions ou décrire les concepts que nous y avons associés.

## **IV. Discussion**

### **IV.1. Apports de l'expérience 1**

Les verbalisations produites par les participants pour décrire les sons se répartissent donc en quatre grandes catégories selon que les participants parlent des dissemblances entre les sons, d'eux-mêmes, des sons ou de la source sonore qui aurait pu les produire. Ces catégories de verbalisations sont produites de manière égale par les musiciens et les non-musiciens, même si ces derniers sont plus nombreux à parler d'eux-mêmes.

Par contre, on peut noter des différences inter-participants qui semblent plus liées à la personnalité de chacun qu'à sa compétence musicale et qui sont très significatives. On observe une tendance à parler plus de la source quand un des sons de la paire est très vraisemblablement une imitation de son d'instrument de musique connu et au contraire à en parler moins quand un des sons de la paire est un son hybride. On observe aussi la tendance à parler plus de dissemblance quand les sons sont jugés proches (jugement de dissemblance de 1 sur l'échelle de 1 à 9). De plus, les verbalisations sur le caractère semblable ou différent des sons sont significativement liées, de manière redondante, aux jugements de dissemblance perceptive donnés par les participants. On retrouve ainsi un des résultats de Samoylenko et al. (1996).

#### **IV.1.1. Des verbalisations libres comme données ?**

La présence de verbalisations où le participant parle d'autre chose que de la tâche demandée, en particulier de lui-même a souvent servi à critiquer l'utilisation de verbalisations libres (Caverni, 1988). Cela concerne ici 23% des verbalisations. Pourtant, il est important de noter d'une part que c'est le fait de certains participants, sans doute d'autant plus qu'ils se sentent en difficulté par rapport à la tâche demandée ; d'autre part que ce n'est que dans quelques cas exceptionnels que les verbalisations d'un participant pour une paire donnée ne sont que de ce type. En général, elles ne représentent qu'une faible partie de chacune des verbalisations.

De plus, alors que les non-musiciens verbalisent plus leurs jugements de valeur sur les sons, ou des signes de répétition, ou encore des activités autres que celles demandées, les verbalisations exprimant des explications sur les termes utilisés ou sur des difficultés rencontrées se répartissent plus équitablement entre musiciens et non-musiciens. D'autant plus qu'il s'agit de mentionner des erreurs effectuées (67% par des musiciens) ou des difficultés d'expression (69% par des musiciens).

En fait, ces verbalisations sont souvent de l'ordre du métalangage et peuvent donc être utiles pour l'expérimentateur qui doit coder le sens des expressions des participants. Ceci est particulièrement vrai dans le cas de verbalisations d'explications sur l'utilisation que le participant fait de certains termes. En effet, nous verrons que les mots que les participants emploient spontanément le plus souvent, sont aussi des mots qu'ils considèrent eux-mêmes comme impropres ou inadaptés aux sons écoutés (Faure et al., en préparation). Les précisions données par les locuteurs visent donc souvent à expliquer pourquoi ou dans quel cas ils peuvent tout de même utiliser ce vocabulaire (le seul à leur disposition) pour parler de ces sons en particulier. Ces verbalisations nous sont donc particulièrement utiles pour comprendre pourquoi et comment les locuteurs ont utilisé les termes choisis.

Enfin, l'analyse des verbalisations extérieures à la consigne pourrait être un bon indice de la difficulté de la tâche ou du niveau d'attention du participant. En effet, la tâche de descriptions de nos perceptions auditives est inhabituelle et difficile : on n'a pas de vocabulaire conventionnel pour cela, mais celle d'effectuer 66 jugements de dissemblance de suite est aussi fastidieuse. Le fait de révéler ces difficultés peut être intéressant.

#### IV.1.2. Les locuteurs parlent du son

Les verbalisations où le participant parle des sons sont à la fois systématiques et très variées. On y trouve assez systématiquement des descriptions d'ordre temporel, de la forme des sons ainsi que des qualités relevant de quatre de nos sens (audition, toucher, vision et goût). Ces verbalisations sont produites par tous les participants sans distinction de compétence musicale. On trouve aussi citées, de façon moins systématique et plus dépendante de la compétence musicale ou de la personnalité de chacun des participants, des qualités d'ordre émotionnel ainsi que d'autres qualités plus générales, des noms de sons et des onomatopées. Les verbalisations sur le déroulement temporel des sons occupent une place importante dans la description des sons (79%). Elles dépendent peu de la compétence musicale des participants qui les produisent. Les analyses que nous avons faites montrent que ces verbalisations sont souvent reliées très finement avec la réalité acoustique des sons et décrivent des successions d'événements ayant des durées inférieures à une seconde. Les participants décrivent aussi une forme des sons qu'ils perçoivent (dans 46% de leurs verbalisations) et celle-ci semble d'ordre temporel. Elle est en effet, souvent assez proche de la représentation graphique donnée par l'enveloppe temporelle de ces sons.

La description des sons fait aussi appel à des mots d'un des domaines sensoriels comme le *toucher*, la *vision* et le *goût*, et apparaît dans 33% des verbalisations. Tous les participants y font appel au moins une fois dans leurs verbalisations même si cela représente des pourcentages de leurs verbalisations très variés, entre 2 et 70% ( $M=34\pm 20\%$ ). La part due aux participants musiciens est égale à celle due aux non-musiciens. Trois participants utilisent ce genre de verbalisations plus souvent que les autres (entre 56 et 70%), quatre moins souvent que les autres (entre 2 et 6%) parmi lesquels on trouve notamment deux des participants qui ont produit la plus petite quantité de verbalisations. Les professionnels du domaine sonore (par exemple assistant de composition, ou producteur de radio) affirment en général que l'essentiel de la façon dont ils verbalisent les qualités du timbre des sons relève du domaine du toucher. Les mots les plus fréquemment cités étant *rugueux*, et *doux*. Il est par contre intéressant de constater que les verbalisations du caractère *brillant* des sons, qui fait pourtant partie du vocabulaire que les musiciens considèrent comme propre à décrire les sons n'est en fait verbalisé spontanément ici que par 4 participants et principalement par un participant néophyte en musique.

Cette façon de décrire les sons en utilisant du vocabulaire relevant *a priori* d'autres modalités sensorielles renvoie au phénomène de synesthésie, l'expérience du goût lors de la stimulation auditive

par exemple. Si on considère les verbalisations de "forme" comme relevant elles aussi du phénomène synesthésique, les verbalisations de ce type sont présentes ici dans 82% des verbalisations des participants. L'importance et la nature de ce phénomène seront discutées dans la partie III.

L'ensemble de ces analyses nous permet de dégager quelques qualités spécifiques de chacun des 12 sons de notre corpus. Le tableau suivant donne ainsi une ébauche de portraits verbaux dont la reconnaissance sera testée ultérieurement afin de savoir si les sons les plus souvent associés par de nouveaux participants à certaines descriptions seront les mêmes que ceux qui ont inspiré ces descriptions spontanément.

Tableau 6.12. Mots ou catégories spécifiques de chacun des sons écoutés (extrait des tableaux de répartition des verbalisations en fonction des sons contenus dans les paires).

Timbres	Descriptions
TPT (trompette)	pas d'évolution, nasillard, désagréable, stérile, cuivré
TBN (trombone)	rond, rugueux, cuivré
VBR (vibraphone)	fin progressive, résonance, réverbération, rond, plat, doux, bleu, pur, limpide
SNO ( <i>striano</i> )	acide
PNO(piano échantillonné)	simple
CVC (clavecin)	fin arrêtée, rond, aigu, nasillard
CRA (cor anglais)	étouffé, pas beau, stérile, faible-lointain
CLA (clarinette)	pas d'évolution, plat, sourd, pas beau, simple
VBN ( <i>vibrone</i> )	double, désagréable, pas beau, raté, bizarre
OBC ( <i>obochord</i> )	double, aigu, nasillard, doux, synthétique
CFT (corde frottée)	vitesse/lent, début lent, progression, direction, rugueux, chaleureux
PNO (piano)	fin progressive, résonance

#### IV.1.3. Les locuteurs parlent de la source

Les verbalisations à propos de la source sonore, que ce soit à travers des noms d'*objets*, des verbes d'*actions* ou des *descriptions* plus complètes, sont à la fois fréquentes (dans 59% des cas), produites par un grand nombre de participants souvent sans distinction de compétence musicale et en rapport étroit avec le type de sons entendus. Ceci souligne la finesse de perception des participants et leur possibilité d'en parler. Le travail étant souvent d'arriver à rassembler des verbalisations de même type parmi la grande variété des locutions particulières à chacun. Les verbalisations parlant des sources permettent elles aussi d'envisager la construction de portraits verbaux dont on pourra tester la reconnaissance par d'autres participants.

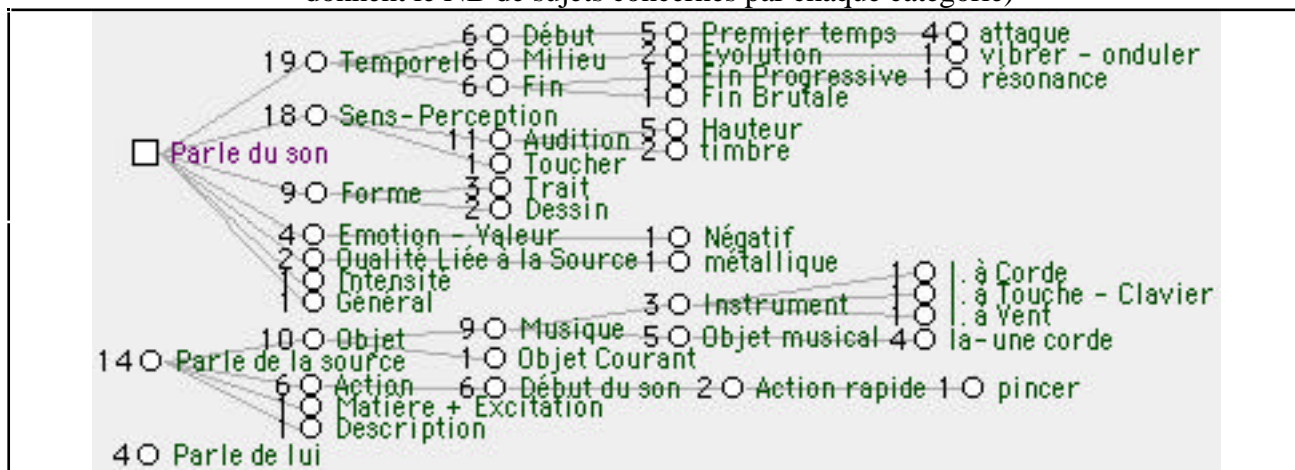
#### IV.1.4. Parlez-moi des sons, je vous dirai qui vous êtes...

Même si ce n'est qu'une boutade, nous avons constaté, tout au long de cette analyse, à quel point chacun des types de verbalisations étudié dépendait avant tout de la personnalité de chacun des participants. La figure 6.15 montre l'ensemble des catégories de verbalisations utilisées par les participants dans plus de 50% de leurs verbalisations.

Nous avons effectué une analyse de cluster à partir du nombre de paires pour lesquelles chacun des participants a produit des verbalisations correspondant à l'ensemble de catégories sélectionnées ci-

dessus, c'est-à-dire les catégories de verbalisations qu'au moins un des participants a utilisé dans plus de 50% de ses verbalisations. On obtient deux groupes de participants qui ne sont pas constitués sur des critères de compétence musicale : on trouve des participants appartenant aux groupes de compétence, notés am1, am2, am3 et néo, dans les deux grands groupes de participants formés à partir de la fréquence de certaines catégories de verbalisations. Seuls les 3 participants professionnels (pro) se retrouvent dans le même groupe.

Figure 6.15. Catégories utilisées par les sujets dans plus de 50% de leurs verbalisations (Les chiffres donnent le NB de sujets concernés par chaque catégorie)



Le premier groupe est formé des participants qui ne parlent systématiquement que des sons (dans plus de 50% de leurs verbalisations), et toujours plus souvent que des sources. Il se scinde en deux sous-groupes : le premier se caractérise surtout par un faible nombre général de verbalisations différentes à propos de chacune des paires et en particulier très peu de verbalisations sur la source sonore. Ceci qui est particulièrement vrai pour 3 des 5 participants. Le deuxième sous-groupe regroupe 2 participants qui produisent des verbalisations sur la source sonore plus souvent, ainsi que les trois participants ayant le plus parlé d'eux-mêmes ainsi que de qualités des sons d'ordre émotionnel.

Le deuxième grand groupe rassemble les participants qui ont tous décrit les sons en terme d'objets, et en terme de déroulement temporel, dans plus de 50% de leurs verbalisations. Ceci est particulièrement vrai pour les 3 participants d'un des 2 sous-groupes qui ont notamment donné des noms d'objets musicaux (et d'instruments pour deux d'entre eux) dans plus de 75% de leurs verbalisations, ainsi que les 3 participants qui ont eux aussi parlé très souvent de la source mais plutôt sous forme d'occurrences de *la-une corde*. Le deuxième sous-groupe est composé de 4 participants dont les productions à propos de la source sonore semblent moins systématiques et moins variées. On pourrait ainsi tenter de définir des portraits de chacun des participants à partir de leurs descriptions.

#### IV.2. Comparaison avec l'approche de Samoylenko et Nosulenko (1996, 1997)



De manière générale l'essentiel des différences vient du fait que nous construisons beaucoup plus de catégories différentes et que cette construction est motivée par les données : elle n'est pas construite *a priori*.

#### **IV.2.1. Le mode d'élaboration et le principe sont très différents.**

La base de données de Samoylenko et Nosulenko est conçue en trois parties : logique, référence au stimulus (*stimulus relatedness*) et sémantique, et 10 niveaux de codage. L'arborescence de codage est conçue a priori et on doit coder chaque locution (définie comme unité verbale) par rapport à ces 10 niveaux de codage de manière systématique.

Notre arborescence se construit en partant des locutions des sujets qu'on regroupe petit à petit dans des catégories qui se ramifient. On aboutit à plus de 1000 catégories différentes, ce qui permet une finesse beaucoup plus grande dans le codage des données, et donc un panorama plus fidèle du champ sémantique révélé. Il serait impossible de coder de manière systématique chaque objet par rapport à chacune de ces catégories, mais le logiciel Stone considère par défaut qu'un objet n'a pas les propriétés que l'on ne lui attribue pas. On se contente donc d'attribuer à chaque objet les catégories pertinentes pour lui. De plus, la même locution peut être considérée à plusieurs endroits de l'arborescence.

Nous n'avons pas repris le concept d'unité verbale, fondamental chez Samoylenko et Nosulenko, qui consiste à segmenter le matériel verbal préalablement à toute analyse en locutions de sens différents. Ce n'est pas que nous le remettons en question, mais à cause de problèmes techniques, notamment de gestion de taille de fichier. En effet, avec 14 sujets décrivant 48 paires de sons, Samoylenko et al. (1996) obtiennent 7300 entrées codées sous 10 niveaux. Nous avons choisi de coder sous environ 1000 niveaux les données de nos 20 sujets décrivant chacun 66 paires correspondant déjà à 1300 entrées sans les avoir subdivisé en unités verbales. Il nous était impossible, considérant les capacités actuelles du logiciel et de l'ordinateur utilisé, de travailler avec de l'ordre de 10000 entrées codées sous 1000 niveaux. De ceci résulte l'inconvénient majeur de notre codage, dû au fait que nous ne séparons pas dans chaque verbalisation des sujets, ce qui concerne les deux sons dans leur ensemble ou chacun des deux sons individuellement. Ceci réduit nos possibilités de comparaison avec les paramètres acoustiques de chacun des sons, puisqu'on sait seulement que cette verbalisation a été produite à propos d'une paire contenant T1 et T2 mais pas spécifiquement T1 ou T2.

Une remarque qui en découle correspond au fait que nous n'avons pas doublé les entrées correspondant aux similarités et aux différences graduelles comme Samoylenko et al. Il faut souligner à ce propos que des psychologues (Tijus, pers. com.) ont montré que des locutions du type "A est plus grand que B" ne sont pas forcément symétriques de "B est plus petit que A" voir même de "B est moins grand que A", car dans la première de ces locutions, c'est une façon d'attribuer la propriété "grand" à A.

Enfin, Samoylenko et al. (1996) définissent un coefficient par sujet permettant de tenir compte, dans les verbalisations de chacun, du nombre total de verbalisations que ce sujet a produits. Ceci afin de ne pas avantager dans le dépouillement les sujets très bavards par rapport aux autres. Cette nécessité est liée au découpage initial du corpus en unités verbales (U.V.), le nombre d'U.V. produits par chaque sujet variant quasiment du simple au quintuple. Nous n'avons donc pas mis en place de choses équivalentes, chacun de nos sujets ayant produit le même nombre de verbalisations, même si elles sont de longueurs différentes. Nos résultats sont exprimés sous la forme du pourcentage de sujets (éventuellement en fonction de sa compétence musicale) qui ont utilisé chacune des catégories de verbalisations construites, ou de pourcentage de l'ensemble des verbalisations que représente une catégorie donnée.

#### **IV.2.2. Comparaison des deux arborescences.**

Ces deux arborescences ont été conçues sans influence réciproque l'une de l'autre pour analyser des données équivalentes (voir similaires...). Il est donc intéressant de comparer ces deux arborescences pour voir ce qu'elles ont en commun et comment elles diffèrent.

##### Niveau logique

La partie logique de la base de données de Samoylenko et al. (1996) est absente de notre arborescence. Nous n'avons donné qu'une "pale" imitation avec notre catégorie *dissemblance / semblable ou différent*, qui ne rassemble que les locutions où le participant parle explicitement des éventuelles ressemblances ou différences qu'il a remarquées. Il ne s'agit donc pas de coder systématiquement toute locution pour savoir si elle concerne des ressemblances ou des différences entre les deux sons écoutés. Par contre le travail de maîtrise de Leslie Ganet, sous la direction de Charles Tijus à Paris 8, traite à fond le problème du mode de comparaison de deux sons dans une paire avec un point de vue différent de celui de Samoylenko et al. (1996). En effet, les auteurs considèrent que lors de la comparaison de deux sons dans une paire, les sujets se focalisent de manière séquentielle : d'abord sur ce qu'il y a de commun aux deux sons (équivalent de /dif/gen/ et /sim/ dans Samoylenko et al.), puis sur le premier, ensuite sur le deuxième (équivalent de /dif/con/cla/), puis sur des propriétés relationnelles entre les deux sons (équivalent de /dif/con/gra/). Ces résultats montrent que 60% des verbalisations des sujets suivent ce modèle de verbalisations.

##### Niveau de référence au stimulus

Nous n'avons pas séparé dans notre arborescence ce que Samoylenko et al. ont séparé en deux parties *stimulus relatedness* et sémantique. Il est à noter que leur partie *stimulus relatedness* qui correspond aux niveaux 4 et 5 de leur codage, est aussi celle qui correspond au plus grand nombre de différences inter-experts lors du codage. De plus, leur niveau .../ssp/tem/ qui code les propriétés spécifiques d'ordre temporel correspond chez nous à une des parties les plus complexes de notre arborescence (très ramifiée en sous-parties). En effet, les sujets verbalisent des propriétés d'ordre temporel dans 80% de leurs verbalisations. La partie .../ssp/int/ (intensité) se retrouve dans notre arborescence mais nous y avons associé des qualités comme *lointain* que Samoylenko et al. ont

codées dans .../ssp/spa/ (spatiales). Nous ne comprenons d'ailleurs pas très bien quelles sont les locutions qui peuvent être rassemblées sous ce code spatial (caractère spatial des sons ? localisation de la source ?). La partie ..ssp/spe/ d'ordre spectral a été par contre chez nous considérée comme faisant partie des qualités liées au domaine de l'audition.

#### Niveau sémantique

Ce que Samoylenko et al. (1996) ont appelé le niveau sémantique représente la partie que notre arborescence développe. Nous n'opérons pas la distinction Analytique/Globale faite au premier niveau de cette partie de l'arborescence par Samoylenko et al. Par contre la différence fondamentale que nous effectuons répond à la question "de quoi parle le sujet ?", de lui-même, du son ou d'une source sonore possible du son. Cette distinction n'apparaît pas dans l'arborescence de Samoylenko et al. et, alors que la plupart des sous-catégories de l'un apparaissent chez l'autre, on trouve une catégorie uniquement dans notre arborescence : les descriptions d'actions effectuées pour produire les sons.

On retrouve ce que Samoylenko et al. avaient codé sous /fea/afe/ (attitudinal features) dans ce que nous avons codé *parle du son/émotion-valeur*. Mais nous ne séparons pas au niveau suivant les verbalisations de type émotionnel de celles concernant le caractère artificiel ou naturel des sons. Les verbalisations de type *artificiel* ont été considérées à la fois comme des jugements de valeur négatifs et placées dans la catégorie *parle du son/qualités liées à la source*, qui représente un moyen intermédiaire entre parler du son et parler de la source.

On retrouve ce que Samoylenko et al. ont codé sous fea/dfe/und-pmd, c'est à dire les traits descriptifs unimodaux (liés uniquement à l'audition) de ceux plurimodaux liés à différentes modalités sensorielles dans notre partie *parle du son / sens-perception/* qui est subdivisé en audition, toucher, vision et goût... Cette partie comporte de nombreuses sous-catégories dans notre arborescence. Elle contient notamment les descriptions de type spectral des sons (codés /ssp/spe/ chez Samoylenko et al.). Nous avons par contre considéré le terme "aigu" (traduction possible de "sharp" comme relevant de l'audition, alors que Samoylenko et al. semblent coder la locution "more sharp" comme /fea/dfe/pmd/ donc plurimodale).

Pour ce qui est de la partie que Samoylenko et al. ont codée sous *holistic meaningful entities (/hen/)*, nous ne faisons pas de différence du type : relié au son / pas relié au son. Nous mettons tous les noms d'objets sous la même sous-catégorie de *parle de la source*, qui se subdivise ensuite dans notre arborescence en objets musicaux, courants, indéfinis ou matériaux. De la même façon, ici aussi, le nombre de nos sous-catégories est beaucoup plus important. Nous re-soulignons ici le fait que notre catégorie *parle de la source/action/* n'a pas d'équivalent dans la base de Samoylenko et al. De la même façon, dans notre base de donnée, la partie codée chez Samoylenko et al. sous /hen/rel/rob/act/mus, objets ayant trait à la musique, est très subdivisée chez nous. Malgré des *a priori* et des logiciels dont le principe est très différent, on retrouve la plupart des catégories considérées par Samoylenko et Nosulenko dans notre arborescence.

### IV.3. Apport de l'expérience 3

L'essentiel des apports des verbalisations libres enregistrées lors de la première partie de l'expérience 3 consiste à pouvoir faire une comparaison plus quantitative entre les productions spontanées des participants et les sons qu'ils associent aux portraits proposés. Ceci sera analysé dans le chapitre suivant.

Pour ce qui est d'un panorama de la manière dont on parle des sons, l'intérêt se situe surtout dans les différences de modes de descriptions en fonction de la consigne donnée aux locuteurs. Il est intéressant de remarquer que, de même que lorsqu'on demandait (dans l'expérience 1) de décrire les différences et les ressemblances entre les sons, on obtenait des descriptions des sons (dans 98% des productions) mais aussi des sources sonores possibles (dans 60% des productions). Si on oriente les locuteurs vers la production de noms de sources sonores, on obtient aussi des descriptions des sons, bien que seulement dans 21% des productions.

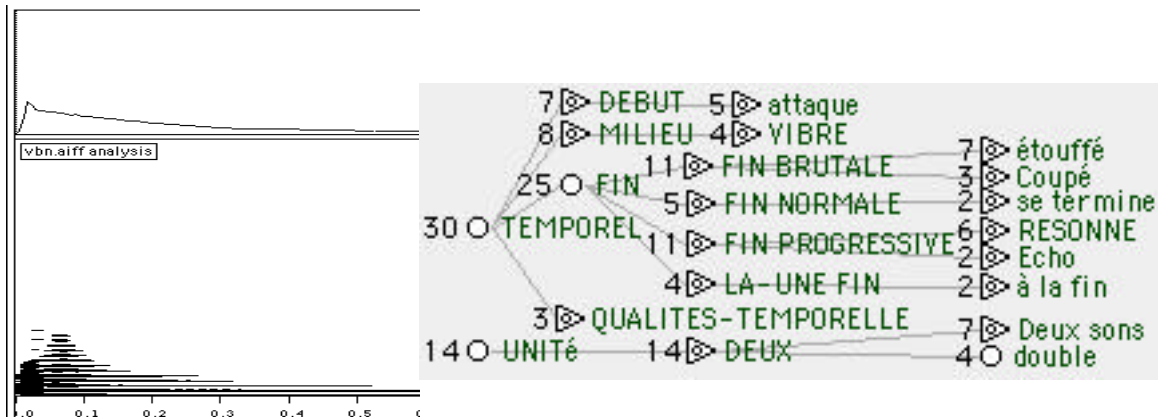
Les différences d'enregistrement des verbalisations nous ont permis de noter les gestes effectués par les locuteurs au cours de leur tentative de description des sons. L'analyse de ces gestes nous a apporté un argument supplémentaire pour souligner l'importance de la référence à la source sonore, notamment au mode de production des sons, lorsqu'on les entend.

Enfin, une analyse exhaustive de ces verbalisations nous permettrait des comparaisons plus fiables que celles qu'on peut effectuer à partir des verbalisations de l'expérience 1, entre les verbalisations d'une part et les caractéristiques physiques des sons entendus d'autre part. En effet, le fait que chaque production soit la description d'un seul des 12 sons et non pas d'une des 66 paires possibles, les capacités d'analyse du logiciel que nous avons choisi sont plus optimisées. Nous pouvons obtenir un aperçu de ce que ces comparaisons peuvent être dans les figure 6.16 et 6.17.

Figure 6.16. Comparaison des profils temporels (enveloppes temporelles et sonogrammes) du son CVC et des locutions d'ordre temporel utilisées pour les décrire.



Figure 6.17. Comparaison des profils temporels (enveloppes temporelles et sonogrammes) du son VBN et des locutions d'ordre temporel utilisées pour les décrire.



Les noms en minuscules sont des occurrences de verbalisations, quand on ajoute une majuscule, ils regroupent des occurrences de formes proches (par exemple conjugaison du verbe) ; les noms en majuscules sont des catégories qui regroupent des occurrences de sens jugés proche.

#### IV.4. Conclusion

Les verbalisations libres, souvent accusées de gêner la réalisation d'expériences et de relater des choses extérieures à celles-ci, apparaissent ici au contraire comme des révélateurs extrêmement fins de perceptions des participants, elles-mêmes très fines à propos de sons perçus. En effet les verbalisations des participants peuvent, par exemple, décrire précisément la succession de phénomènes se déroulant dans un délai de moins d'une seconde, comme c'est le cas des verbalisations d'ordre temporel, ou attester de la distinction perçue entre des sons imitant une production de corde pincée ou frappée, voir même de décrire un mode de fabrication qui correspond à la manière dont le compositeur a conçu certains des sons hybrides proposés. Par exemple, le participant S2.pro décrit le son VBN, conçu comme mélange d'un son de trombone avec un son de vibraphone, par *un tromboniste qui fait une bêtise sur une paire de timbales*. Mais ce n'est pas l'apanage des musiciens professionnels puisque le participant S16.néo décrit le son OBC, conçu comme mélange de clavecin et de hautbois, par *mélange de percussion et de corde pincée*. Ces verbalisations permettent aussi de détecter d'éventuelles difficultés des participants, de mieux comprendre le sens qu'ils donnent à certains mots, voir de révéler leur personnalité !

Ce travail présente donc la manière dont des participants de compétence musicale variée décrivent un ensemble de sons qui ne diffèrent que par leur timbre. Pour être représentatif du timbre musical, il devrait être étendu à un ensemble de sons plus grands et plus variés (notamment contenant des sons naturels). Pourtant, il souligne déjà des tendances intéressantes qui semblent refléter le sens que nous donnons aux sons que nous percevons.

Nous avons construit, à partir de ces verbalisations, différents types de portraits verbaux de chacun de nos 12 sons. Nous testerons dans le chapitre 8 la capacité des participants à reconnaître un de ces 12 sons s'ils sont confronté à la lecture d'une de ces descriptions.

## Chapitre 7. Pertinence des échelles et sens des portraits

Nous avons utilisé deux types de démarche dans la partie expérimentale de cette thèse. Nous avons, d'une part, recueilli la manière dont les auditeurs parlaient spontanément des sons qu'ils entendaient (ceci à été analysé dans le chapitre 6), et d'autre part, nous avons proposé du vocabulaire à utiliser de manière systématique. Dans cette deuxième perspective, nous avons toujours voulu avoir malgré tout l'avis des auditeurs sur la pertinence qu'avait pour eux le vocabulaire qu'on leur demandait d'utiliser. Ainsi, dans l'expérience 2, où les auditeurs devaient positionner les 12 sons sur 23 échelles sémantiques, nous leur avons demandé un jugement global de la pertinence de chaque échelle pour l'ensemble des sons écoutés ("est-ce que cette échelle est globalement adaptée / pas adaptée dans ce contexte"), ainsi qu'un jugement de chaque échelle pour chacun des sons écoutés ("est-ce que cette échelle est adaptée / pas adaptée pour ce son"). Pour l'expérience 3, où les auditeurs devaient associer un ou plusieurs sons à chacun des portraits verbaux qui leur était proposé, on demandait à la fin, pour chacun des portraits, s'il présentait un sens, "précis", "vague" ou "pas de sens" dans ce contexte.

Nous avons été encouragés à demander ces jugements :

- d'une part à cause des critiques habituelles faites à l'utilisation des échelles sémantiques, c'est-à-dire que le vocabulaire testé est sélectionné préalablement à l'expérience par des spécialistes et n'a peut-être ni la même signification, ni réellement de sens pour les auditeurs qui doivent ensuite utiliser ces échelles ;

- d'autre part, du fait que, dans notre contexte où les sons écoutés étaient égalisés en hauteur, durée et sonie et que malgré cela, parmi les adjectifs les plus souvent utilisés pour comparer ces sons deux à deux, on pouvait trouver des adjectifs se référant à ces qualités égalisées, soit respectivement, *aigu / grave, long / court et fort*.

Il nous fallait comprendre pourquoi ces mots avaient été utilisés et quels sens leurs donnaient leurs locuteurs, et d'autres utilisateurs. Enfin, ils étaient peut-être pertinents pour seulement quelques sons, et il était intéressant de comprendre lesquels. L'intérêt des résultats dégagés de ces jugements de pertinence dans cette expérience 2, nous a encouragé à réitérer l'expérience lors de notre troisième expérience. Les analyses correspondant aux données recueillies dans les expériences 2 et 3 seront présentées dans les deux paragraphes suivants.

## **I. Expérience 2 : pertinence des échelles et commentaires des sujets**

Dans ce paragraphe vont être analysés les jugements de pertinence demandés aux sujets dans l'expérience 2, ainsi que les éventuels commentaires écrits qu'ils ont donné à la fin de la session sur leur utilisation de chacune des échelles.

### I.1. Jugements de pertinence

Nous avons jugé important que cette expérience, qui s'inspirait de la sémantique différentielle, actualise deux variantes de cette méthode. La première variante, discutée de manière plus approfondie dans l'introduction du chapitre 4 §II, est de choisir 23 échelles unipolaires. La deuxième variante consiste à ajouter aux jugements de positionnement des sons sur les échelles sémantiques, consigne classique en sémantique différentielle, des jugements de pertinences de ces échelles.

Ceci nous permettait d'assumer mieux l'idée d'une expérience contrainte par un choix de vocabulaire préalablement à l'expérience. En effet, si chaque sujet doit obligatoirement juger les 12 sons de notre expérience selon les 23 échelles sémantiques que nous avons choisies, il nous semblait important de pouvoir écouter son avis sur la pertinence de ces échelles (pour lui) et de pouvoir nuancer ses jugements numériques par ses jugements de pertinence. Nous avons donc fait intervenir deux types de jugements de la pertinence des échelles sémantiques proposées au cours de cette expérience. Un jugement global sur la pertinence de l'échelle par rapport à l'ensemble de sons écoutés et jugés, et un jugement individuel pour chaque échelle et pour chaque son (une échelle pouvant être globalement inappropriée mais convenir à quelques sons). Ceci rejoint les considérations que nous avons eu lors du choix de ces échelles, à partir d'un critère de nombre d'occurrence lors de la production spontanée et dans un cadre précis de celle-ci (comparaison graduelle entre deux sons).

Nous allons donc analyser ici successivement :

- Les jugements de pertinence globale, c'est à dire la réponse à la question "est-ce que cette échelle vous paraît adaptée / pas adaptée pour l'ensemble des sons écoutés".
- Nous considérerons alors les jugements de pertinence de chaque échelle pour chacun des sons afin de vérifier le fait que, même si les échelles choisies correspondaient toutes aux mots les plus souvent utilisés (lors de comparaisons graduelles entre les deux sons d'une paire), ils ne correspondaient pas forcément à une utilisation de comparaison pour l'ensemble des sons proposés et étaient peut être spécifiques de certaines paires de sons.

Ceci nous permettra aussi de reposer la question du jugement graduel versus absolu : Peut-on par exemple juger qu'un son est "peu soufflé" s'il est absolument "pas soufflé" ?



Enfin, nous essayerons de montrer le lien entre le jugement de pertinence d'une échelle effectué dans le cadre de cette expérience, et l'utilisation spontanée ou non de ce mot par les mêmes sujets lors de l'expérience précédente.

### I.1.1. Pertinence globale brute

Nous allons analyser ici les jugements de pertinence globale effectués sur chaque échelle par les sujets après l'avoir utilisée. Le tableau 7.1 donne le pourcentage de sujets ayant considéré que chaque échelle sémantique était pertinente globalement pour l'ensemble des sons écoutés.

Tableau 7.1. Pourcentage de sujets ayant considéré une échelle comme globalement adaptée pour l'ensemble des sons écoutés.

Pertinence globale	% sujets	% anciens	% nouveaux	% musiciens	% non-mus.
aigu	<b>39</b>	<b>40</b>	<b>38</b>	<b>31</b>	47
attaque	<b>91</b>	88	<b>94</b>	<b>94</b>	<b>88</b>
bas	<b>29</b>	<b>27</b>	<b>31</b>	<b>25</b>	<b>33</b>
brillant	62	54	69	53	71
clair	71	87	56	69	73
diffus	57	67	47	60	53
doux	<b>90</b>	<b>93</b>	<b>88</b>	81	<b>100</b>
fort	<b>23</b>	<b>21</b>	<b>25</b>	<b>27</b>	<b>20</b>
grave	<b>40</b>	<b>36</b>	<b>44</b>	<b>31</b>	50
haut	<b>26</b>	<b>27</b>	<b>25</b>	<b>19</b>	<b>33</b>
large	52	47	56	50	53
long	52	53	50	50	53
métallique	<b>84</b>	<b>93</b>	75	75	<b>93</b>
nasillard	78	88	69	<b>88</b>	69
net	81	75	<b>88</b>	81	81
pincé	68	87	50	50	<b>87</b>
résonnant	79	85	75	75	85
riche	55	44	67	67	44
rond	69	69	69	<b>88</b>	50
sec	74	80	69	63	<b>87</b>
soufflé	48	67	<b>31</b>	44	53
sourd	<b>87</b>	<b>93</b>	<b>81</b>	81	<b>93</b>
vibré	65	60	69	69	60
moyenne	62	64	59	60	64
écart type	21	24	21	22	22

En gras, les valeurs s'écartant de la moyenne de plus d'un écart-type.

En moyenne, les échelles sémantiques proposées sont considérées comme pertinentes par seulement 62% ( $\pm 21\%$ ) des sujets. Bien que ces échelles soient extraites de verbalisations libres d'un groupe de 20 sujets d'expérience musicale variée effectuées sur les mêmes sons que ceux testés ici, et que les mots choisis soient ceux qui apparaissent le plus fréquemment spontanément, leur pertinence n'est pas forcément une évidence pour les sujets auxquels on demande de les utiliser. Ceci est un résultat important, peu considéré lors d'utilisations de techniques similaires de sémantique différentielle ou de sa version VAME avec des échelles unipolaires.

De plus, si on considère l'ensemble des 32 sujets de cette expérience, les différences entre les échelles de nombre de sujets qui considèrent qu'une échelle est globalement pertinente pour l'ensemble des sons écoutés, sont très significatives ( $\chi^2(22)=48.6$ ,  $p<.0001$ ). Les différentes échelles proposées n'ont donc pas été jugées de la même façon par les sujets.

#### Jugements de l'ensemble des sujets

Si on considère les échelles pour lesquelles le pourcentage de sujets les ayant jugées pertinentes sort de l'écart-type, on observe que :

- deux échelles sémantiques sont considérées comme pertinentes par plus de 90% des sujets : celle liée à la notion d'*attaque rapide-attaque lente* (91% des sujets) et celle liée à la notion *peu doux-très doux* (90% des sujets). Les échelles liées à *sourd* et *métallique* sont considérées comme pertinentes respectivement par 87 et 84% des sujets. Ces quatre échelles remportent donc l'adhésion de la majorité des sujets. Elles seraient les plus adaptées pour juger de la qualité sonore de sons complexes égalisés en hauteur, sonie et durée.

- à l'autre extrême, 5 échelles sont jugées pertinentes par moins 41% des sujets :

L'échelle sémantique considérée comme pertinente par le moins de sujets est celle liée à *fort* (23%), ce qui peut se comprendre compte tenu du fait que *fort* correspond à la notion de sonie, égalisée perceptivement lors de la synthèse sonore. Pourtant ce terme faisait bien partie de ceux utilisés le plus fréquemment spontanément pour comparer des sons 2 à 2 dans notre expérience 1. C'est en prévoyant ce genre de réponses que nous avons choisis de demander un jugement de pertinence globale mais aussi un jugement de pertinence pour chacun des sons. En effet, une hypothèse permettant de recueillir le terme *fort* comme un des plus utilisés spontanément alors que seulement 23% des sujets le considèrent comme pertinent serait qu'il n'est pertinent que pour quelques sons très précis. Nous pourrions tester cette hypothèse dans le paragraphe suivant. Ensuite viennent les 4 termes liés habituellement à la notion de hauteur tonale : *haut* (26%), *bas* (29%), *aigu* (39%) et *grave* (40%), cette hauteur étant égalisée entre tous les sons dans l'expérience (fréquence fondamentale de 311Hz, Mib3 pour les 12 sons). Ces termes sont donc eux aussi jugés comme globalement non pertinents par la majorité des sujets alors qu'ils font partie des plus utilisés lors de la comparaison spontanée des sons 2 à 2 (*aigu* est le mot le plus utilisé dans ce contexte).

Nous reprendrons la comparaison de ces différents comportements entre les deux expériences (verbalisation spontanée versus échelles sémantiques) à la fin de ce paragraphe.

#### Comparaison entre différents groupes de sujets

A partir du moment où le vocabulaire a été extrait de verbalisations libres d'un groupe de 20 sujets, il pouvait être intéressant :

- d'une part de reposer systématiquement ce vocabulaire à ces mêmes sujets (groupes des "anciens") afin de tester une éventuelle différence de comportement entre une utilisation libre et spontanée d'un vocabulaire et une utilisation plus contrainte et systématisée de ce même vocabulaire ;

- d'autre part de tester une éventuelle différence d'utilisation de ce vocabulaire par rapport à un groupe de sujets "nouveaux". Une hypothèse serait que dans le cas où une différence significative de pertinence apparaîtrait entre ces deux groupes de sujets, la pertinence devrait être moindre dans le groupe des sujets "nouveaux" que dans le groupe des sujets "anciens". (Pour une définition de la répartition des sujets en différents groupes, voir l'Annexe A.4.2.)

Si on considère l'ensemble des réponses des deux groupes de sujets pour les 23 échelles, une analyse de  $\chi^2$  sur le nombre de sujets montre que ces résultats ne sont pas significativement différents ( $\chi^2(22)=7.3, p=.998$ ). Globalement, les mêmes échelles ont été jugées très pertinentes ou très peu pertinentes par les deux groupes de sujets : anciens ou nouveaux. De même, les différences entre les sujets musiciens et non-musiciens, ne sont pas significatives ( $\chi^2(22)=11.0, p=.975$ ).

Les différences de pertinence observées entre les échelles ne dépendent donc pas d'un groupe de sujets particulier.

### I.1.2. Pertinence son par son

Considérant qu'une échelle peut être globalement inadaptée sauf pour quelques sons (ou à l'inverse être globalement adaptée sauf pour quelques sons), nous avons demandé à chaque sujet, lors de la séance 2, de juger de la pertinence de chaque échelle proposée pour chacun des sons jugés après leur positionnement sur l'échelle sémantique. Nous avons donc recueilli pour chaque sujet, 12\*23 jugements de pertinence. Ils vont être analysés ci-dessous.

#### Corrélations inter-sujets sur les jugements de pertinence

Nous avons effectué des corrélations inter-sujets sur ces jugements de pertinence de chaque son pour chaque échelle proposée effectués par les sujets. Les résultats de deux d'entre eux n'étant pas analysables, cette analyse est effectuée sur 30 sujets et 12\*23=276 jugements par sujets.

Tableau 7.2. Nombre et pourcentage de corrélations inter-sujets (sur 439) correspondant aux différents seuils de significativité (corrigés pour 30 analyses),  $df=274$ .

Seuil	Nombre de corrélations I.S.	% de corrélations I.S.
$p<.003$	44	10
$.003<p<.01$	17	4
$.01<p<.05$	26	6
$.05<p<.1$	14	3
ns ( $p>.1$ )	334	77

Globalement ces corrélations sont mauvaises, et ce malgré le nombre élevé de degrés de liberté lors du calcul. Les sujets ne considèrent donc pas qu'une même échelle soit adaptée pour les mêmes sons.

#### Description générale

Nous avons retenu les cas où une échelle était considérée comme adaptée à un son par plus de 50% des sujets. Les résultats sont rassemblés dans le tableau de l'annexe A.7.1. Un résumé des principaux

résultats se trouve dans le tableau 7.3 ci-dessous. On peut remarquer que si on calcule la moyenne sur les 12 sons du pourcentage de sujets ayant trouvé une échelle adaptée (dernière colonne du tableau), le résultat est très fortement corrélé à la pertinence globale exprimée sous forme de pourcentage de sujets ayant considéré que cette échelle était globalement adaptée pour l'ensemble des sons ( $r(21)=.93$ ,  $p<.01$ ).

Tableau 7.3. Pourcentage de sujets qui ont trouvé une échelle adaptée pour un timbre considéré.

Nombre de timbres pour lesquels l'échelle est adaptée pour plus de 50% des sujets et pourcentage moyen sur les 12 sons (%MS)

Tous	TPT	TBN	CRA	CLA	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO	Nb de Timbres	% MS
haut													0	38
bas											•		1	32
fort							*						1	42
aigu									•	•			2	43
grave									•	•			2	43
soufflé	•	*		•									3	46
large					•	•		•			•		4	51
long				•	•	•		•			*		5	56
brillant						•	•		•	•	•		5	54
pincé						•	•	•	**	*			5	54
riche	•	•				•	•	•	•		•		7	49
nasillard	*	•	*		•					•	•	•	7	58
rond		•		•	•	•	•	•	•		•		8	56
diffus		•		•	*	•	•	•			•	•	8	58
sec	•			•		•	*	•	•	•	•	•	9	57
vibré		•			•	•	•	•	•	•	*	•	9	60
métallique	•	•	•			•	*	•	*	*	•		9	64
résonnant		•			•	*	•	*	•	•	*	•	9	64
clair	•		•	•		*	*	•	*	*		•	9	64
net	•	•		•	•	•	*	*	*	•	•	•	11	67
doux	•	•	•	*	•	*	•	*	*		•	•	11	68
sourd	•	*	•	*	•	•	•	•	•	•	*	•	12	66
attaque	*	•	•	•	*	*	*	*	*	*	*	*	12	81

Avec : \*\* (respectivement \*, •) : si 100% (respectivement plus de 75% ou plus de 50%) des sujets considèrent l'échelle appropriée pour ce son.

Parmi les échelles pertinentes pour plus de 50% des sujets et pour plus de 10 des 12 sons, on trouve les 4 échelles suivantes : celles liées à *sourd* et *attaque rapide*, pertinentes pour l'ensemble des 12 sons (particulièrement pour les sons TBN, CLA et VBN pour l'échelle liée à *sourd*) et celles liées à *net* et *doux*, pertinentes pour tous les sons sauf respectivement les sons CRA pour *net*, et OBC pour *doux*. A part pour *net* (considéré tout de même comme globalement adapté par 81% des sujets) ce sont les échelles qui étaient considérées comme globalement adaptées à l'ensemble des sons par le plus grand nombre de sujets à travers les jugements de pertinence globale. Parmi celles-ci, on trouvait aussi l'échelle liée à *métallique* qui se trouve ici être pertinente pour seulement 9 des 12 timbres, c'est à dire pour tous sauf les timbres SNO, CLA et CFT (pour lesquels l'échelle est tout de même adaptée pour respectivement 44, 50 et 50 % des sujets). Il est intéressant de remarquer que le timbre de CLA est

celui qui est positionné en moyenne comme le *moins métallique*, les timbres CFT et SNO étant par contre en moyenne positionnés au milieu de l'échelle.

A part celle liée à *métallique*, 4 autres échelles sont adaptées pour 9 des 12 sons : celles liées à *sec*, *vibré*, *résonnant* et *clair*. A part l'échelle liée à *vibré*, globalement pertinente pour 65% des sujets, les 3 autres échelles sont considérées comme globalement adaptées par plus de 70% des sujets. On peut remarquer que les échelles liées à *résonnant* et *vibré* sont considérées comme adaptées (par plus de 50% des sujets) pour les mêmes 9 sons, tous sauf ceux imitant des sons soufflés, à l'exception du son TBN pour lequel ces deux échelles sont quand même adaptées.

On trouve ensuite *rond* et *diffus*, adaptés pour quasiment les 8 mêmes sons. Les seules différences apparaissant pour les sons CVC (plutôt adaptés pour *rond*) et SNO (plutôt adapté pour *diffus*). Puis, 2 échelles sont adaptées pour 7 sons, celles liées à *nasillard* et *riche*. On peut remarquer au contraire le peu de sons adaptés communs à ces deux échelles (TPT, TBN et VBN), bien qu'elles relèvent de qualités *apriori* proches. En effet, l'échelle liée à *riche* semble adaptée plutôt pour les sons VBR, PNE, PNO et CVC : tous des imitations de sons pincés ou frappés réels, alors que celle liée à *nasillard* semble adaptée plutôt pour les sons CRA, CFT, OBC et SNO.

Enfin on peut considérer un ensemble de 4 échelles, pertinentes pour 4 à 5 sons, celles liées à *large*, *long*, *brillant* et *pincé*. Les sons adaptés à l'échelle *pincée* seront discutés ci-dessous en comparaison avec ceux adaptés pour *soufflé*. Pour ce qui est des trois autres échelles, on peut remarquer que *brillant* est adapté pour les sons VBR, PNE, CVC, OBC et VBN, soit tous les sons dont le mode de production imité est pincé ou frappé, qu'ils imitent des instruments réels ou qu'ils soient des hybrides. *Large* et *long* sont adaptés pour quasiment les mêmes sons, CLA, CFT, VBR, PNO, et VBN, soit un mélange de sons très différents.

A l'autre extrême, parmi les échelles considérées comme adaptées pour seulement 1 ou 2 sons par plus de 50% des sujets, on retrouve les 5 échelles considérées comme globalement les moins pertinentes, c'est à dire les échelles liées à *haut*, *bas*, *fort*, *aigu* et *grave*. En effet, l'échelle liée à *haut* n'est considérée comme adaptée par plus de 50% des sujets à aucun des 12 sons proposés. Les échelles liées à *bas* et *fort* le sont seulement pour un des 12 sons (respectivement le son VBN pour 53% des sujets qui est aussi le son jugé le plus bas en moyenne et le son PNE pour l'échelle liée à *fort* pour 84% des sujets). Cette échelle représente justement le cas d'une échelle qui ne serait pertinente que pour un son. D'où l'intérêt de poser les deux questions de pertinence globale et de pertinence pour chacun des sons. Ensuite on trouve les échelles liées à *aigu* et *grave* qui ne sont pertinentes (pour plus de 50% des sujets) que pour deux sons, CVC et OBC, c'est à dire respectivement l'imitation de son de clavecin et l'hybride de clavecin et de hautbois appelé *obochord*. Ce sont les sons jugés les plus aigus et les moins graves en moyenne par les sujets. Ceci nous permet de penser qu'une échelle *aigu-grave* pourrait être une vraie échelle bipolaire : l'ensemble des sujets considère que l'échelle *grave* est adaptée à des sons qui sont les moins graves. Il ne s'agit donc pas d'un jugement de type

catégoriel où on considérerait que l'échelle est adaptée à un son si le son possède la qualité associée directement à cette échelle (ce qui semble le cas au contraire pour les échelles liées à *pincé* et *soufflé*, comme on le verra plus loin). Ces deux sons contiennent tous les deux au moins l'enveloppe temporelle d'un son de clavecin et toujours beaucoup d'harmoniques aigus. En effet, ce sont les deux sons qui ont les valeurs de CGS (Centre de Gravité Spectral) les plus élevées de l'ensemble de 12 sons considérés.

Il est intéressant de remarquer qu'autant les sujets semblent considérer les échelles liées à *aigu* et *grave* comme des échelles antonymes, autant on observe de grandes différences de comportement avec *haut* et *bas* qui sont censés être leurs synonymes. En effet, l'échelle liée à *bas* est considérée comme adaptée pour le son jugé le plus bas (VBN) et celle liée à *haut* pour aucun des sons. On pourrait donc en conclure que *haut* n'est pas synonyme de *aigu* (ni *bas* de *grave*) dans ce contexte, et que les échelles liées à *haut* et *bas* ne sont pas antonymes. Ceci nous montre une nouvelle fois la difficulté d'utiliser des échelles sémantiques bipolaires.

Parmi les échelles pertinentes pour relativement peu de sons, on trouve aussi celles liées à *soufflé* et *pincé*. Ces deux échelles ont été construites à partir des deux modes de productions imités des sons de notre ensemble. On s'attendait donc à ce qu'elles se comportent comme des antonymes. De plus, comme la notion de début des sons semble si importante à la fois dans la perception du timbre (première dimension de l'espace perceptif de McAdams, et al. (1995) et pour comparer les timbres (la notion d'attaque et de début est très souvent citée dans les verbalisations), nous nous attendions à deux échelles très pertinentes et pour l'ensemble des sons proposés. Or il n'en est rien. D'une part, l'échelle liée à *soufflé* semble pertinente uniquement pour les sons qui sont positionnés en moyenne comme les *plus soufflés* : TBN pour 81% des sujets, CLA pour 68% des sujets et TPT pour 55% des sujets. Il est intéressant de remarquer que ces trois sons sont des imitations de sons d'instruments à vent. Le seul autre exemple de son imitant un instrument à vent et non cité dans cet ensemble est le son CRA qui est en fait tellement synthétique que la qualité de "souffle" en a totalement disparu. D'autre part, l'échelle liée à *pincé* n'est pertinente que pour 5 sons tous disjoints des précédents et eux aussi parmi ceux qui sont positionnés comme les *plus pincés* : CVC pour 100% des sujets (c'est le seul cas où tous les sujets sont unanimes), OBC pour 81% des sujets, puis les sons VBR, PNE et PNO pour respectivement 56%, 69% et 56% des sujets. Ce sont là aussi tous les sons imitant des instruments dont le mode de production est pincé ou frappé, à part le son OBC qui est un hybride construit avec l'enveloppe temporelle du clavecin donc la même impression d'attaque.

Les sons cités ni pour l'échelle liée à *soufflé*, ni pour celle liée à *pincé* sont le son CRA, particulièrement synthétique, et les sons CFT et SNO, tous les deux construits à partir de l'enveloppe temporelle d'une corde frottée, donc ni soufflés, ni pincés, bien qu'en général les sujets semblent confondre une émission soufflée et frottée. On trouve aussi le son VBN qui est le son hybride dont l'enveloppe temporelle est modifiée, donnant l'impression d'une double attaque au début du son.

Au contraire de ce qu'on a pu observer pour les échelles liées à *aigu* et *grave*, cette répartition laisse penser que ces échelles ont été utilisées de manière catégorielle plutôt que continue. Il est vrai qu'on

peut trouver difficile de juger sur une échelle *peu pincé - très pincé* des sons qui ne sont absolument PAS pincés.

Ces mesures de pertinence des échelles sémantiques stimulus par stimulus pourraient être une manière de tester si la qualité testée est vraiment perçue comme un continuum à travers les stimuli ou si, au contraire, elle n'est que spécifiquement associée à certains. Un moyen de tester le matériel est proposé dans un paradigme de sémantique différentielle : dans quelle mesure les échelles peuvent être considérées comme vraiment bipolaires ou non, dans quelle mesure elles peuvent concerner des qualités dont la variation est continue sur l'ensemble des stimuli testés.

#### Différences musiciens / non-musiciens

Si on teste la différence de répartition en fonction de la compétence des sujets du nombre de sujets ayant considéré qu'une des échelles est adaptée pour chacun des timbres, aucune des 23 différences n'est significatives ( $\chi^2(11) < 3$ ,  $p > .99$ ) (les données sont en annexe tableaux A.7.2 et A.7.3). La compétence des sujets n'influe donc pas sur le choix de considérer qu'une échelle est adaptée pour un son donné. De même le nombre de sons pour lesquels chacune des 23 échelles est adaptée pour plus de 50% des sujets n'est pas significativement différent en fonction de la compétence des sujets ( $\chi^2(22) = 25$ ,  $p = .28$ ).

#### Différences entre les sons

Nous avons calculé, pour chaque échelle et pour chaque son, le nombre de sujets qui avaient trouvé l'échelle appropriée pour juger ce son. Les différences entre les sons, en terme de nombre de sujets ayant trouvé l'échelle adaptée pour ce son, ne sont significatives que pour l'échelle liée à *pincé* ( $\chi^2 = 8.28$ ,  $p = .004$ ) et pour l'échelle liée à *soufflé*, bien que cette différence soit moins importante ( $\chi^2 = 2.80$ ,  $p = .09$ ). Ces calculs ont aussi été effectués sur le groupe des sujets musiciens et sur celui des non-musiciens. Les différences entre les sons ne sont significatives que pour le groupe des musiciens et l'échelle liée à *pincé* ( $\chi^2 = 8.60$ ,  $p = .003$ ). On peut remarquer que ces deux échelles sont les seules qui n'ont pas été construites à partir des mots les plus utilisés dans notre première expérience pour comparer les sons présentés par paire. Ce sont en effet des termes plutôt utilisés pour catégoriser les sons que pour les comparer. Dans notre contexte, et surtout si les sujets faisaient référence au type d'instrument et de mode de production des sons entendus, il était en effet difficile de juger par exemple sur l'échelle *peu pincé - très pincé* un son imitant la trompette qui n'est en fait PAS émis en pinçant quelque chose, donc PAS pincé.

Nous allons donc regarder la répartition des jugements de pertinence des sujets sur les 12 sons pour ces deux échelles (tableau 7.4. ci-dessous).

Tableau 7.4. Nombre de sujets (sur 32) considérant que l'échelle est adaptée à chacun des sons, moyenne (M) et écart-type (ET).

Timbre	CVC	OBC	PNE	VBR	PNO	TPT	SNO	VBN	CLA	TBN	CFT	CRA	M	ET
pincé	<b>32</b>	<b>26</b>	22	18	18	15	15	15	14	13	12	<b>8</b>	17.33	6.57
soufflé	11	10	10	13	11	18	14	10	<b>22</b>	<b>26</b>	15	13	14.62	4.90

En gras, les valeurs s'écartant de la moyenne de plus d'un écart-type.

Il est intéressant de remarquer que pour ces deux échelles, le nombre de sujets considérant que l'échelle est adaptée pour chacun des sons est très fortement corrélé à la position moyenne de ces sons sur chacune des deux échelles ( $p < .01$  avec respectivement  $r = .91$  et  $r = .85$ ,  $df = 10$  pour les échelles liées à *pincé* et *soufflé*). Ces deux échelles sont donc considérées comme d'autant plus adaptées à un son donné que ce son est positionné plus proche de l'extrémité "*très soufflé*" ou "*très pincé*".

En positionnant les sons comme dans le tableau 7.4, en allant du son pour lequel l'échelle liée à *pincé* est pertinente pour le plus grand nombre de sujets (CVC), jusqu'au son pour lequel c'est vrai pour le plus faible nombre de sujets (CRA), on retrouve une bonne répartition des sons en fonction des classes de familles d'instruments imités.

#### Comparaison avec les jugements globaux

Nous avons calculé pour chaque échelle, la moyenne sur les 12 sons du nombre de sujets ayant jugé cette échelle pertinente pour chacun des sons. Les différences entre les échelles sont significatives ( $\chi^2(1) = 197$ ,  $p = 10^{-30}$ ). Il n'y a, par contre, pas de différence significative entre les jugements pour les deux groupes de sujets : musiciens et non-musiciens ( $\chi^2(22) = 5$ ,  $p = .99$ ).

Globalement, les moyennes calculées pour chaque échelle sur les 12 sons, à partir des jugements de pertinence donnés pour chaque son et chaque échelle, sont très fortement corrélées aux jugements de pertinence globale de chaque échelle ( $r(21) = .91$ ,  $p < .01$ ). Les échelles considérées comme plus pertinentes en moyenne pour les 12 sons sont celles liées à *attaque rapide*, *doux* et *net*. On retrouve donc les deux premières échelles, considérées globalement comme les plus pertinentes pour l'ensemble des sons. Par contre, l'échelle liée à *sourd*, globalement pertinente ne sort de la moyenne que sur le groupe des non-musiciens et celle liée à *métallique* que sur le groupe des musiciens. Ces derniers font aussi ressortir l'échelle liée à *rond*.

On retrouve par contre exactement les mêmes échelles globalement les moins pertinentes et les moins pertinentes en moyenne sur les 12 sons : ce sont celles liées à *fort*, *haut*, *bas*, *aigu*, et *grave*.

#### **I.1.3. Lien entre le jugement de pertinence globale et l'utilisation spontanée (annexe A.7.4).**

La moitié des sujets qui ont passé cette expérience avaient aussi participé à notre première expérience de verbalisations libres tout en jugeant la dissemblance de paires de ces mêmes sons. Il est donc intéressant pour chacun des sujets concernés de comparer leurs jugements de pertinence au nombre d'occurrences spontanées qu'ils avaient produit de ces mêmes mots.

Nous avons donc repris les données de chacun des 16 sujets qui ont passé à la fois notre expérience 1 de verbalisation libre en comparant les sons 2 à 2, et notre expérience 2 de positionnement sur les 23 échelles sémantiques. Pour chacun des sujets, nous avons regardé s'ils avaient utilisé (ou non) au



moins une fois spontanément chacun des mots liés à chacune de nos 23 échelles sémantiques. Dans chacun des cas, nous l'avons comparé avec leur jugement de pertinence globale de chacune des échelles. Les 4 colonnes principales du tableau 7.5 rassemblent donc les cas où :

- un sujet a utilisé spontanément un mot et juge l'échelle associée globalement adaptée ;
- un sujet a utilisé spontanément un mot et juge l'échelle associée globalement pas adaptée. On calcule alors pour chacune des échelles le pourcentage de sujets qui jugent une échelle pas adaptée malgré le fait qu'ils aient eux-mêmes utilisé ce mot spontanément.

De même, pour les sujets qui n'ont pas utilisé spontanément un mot associé à une des échelles sémantiques testées, les trois colonnes suivantes rassemblent :

- un sujet n'a pas utilisé spontanément un mot et juge l'échelle associée globalement adaptée ;
- un sujet n'a pas utilisé spontanément un mot et juge l'échelle associée globalement pas adaptée. On calcule alors pour chacune des échelles le pourcentage de sujets qui jugent une échelle adaptée malgré le fait qu'ils ne l'aient pas utilisée spontanément.

Tableau 7.5. Nombre de sujets (sur 16 anciens) ayant utilisé spontanément dans l'expérience 1 les mots associés aux échelles sémantiques et lien avec leur jugement de pertinence globale dans l'expérience 2.

Utilisation (Exp 1) Jugement (Exp 2)	OUI			NON		
	adaptée	pas adaptée	% pas adaptée	adaptée	pas adaptée	% Adaptée
aigu	6	9	60	0	0	
attaque	11	1	8	3	1	75
bas	1	4	80	3	7	30
brillant	3	1	25	4	5	44
clair	11	1	8	2	1	67
diffus	3	0	0	7	5	58
doux	7	0	0	7	1	88
fort	1	8	89	2	3	40
grave	5	7	58	0	2	0
haut	0	6	100	4	5	44
large	2	2	50	5	6	45
long	8	6	43	0	1	0
métallique	9	0	0	5	1	83
nasillard	9	0	0	5	2	71
net	8	1	11	4	3	57
pincé	12	1	8	1	1	50
résonn(ant)(e)	11	2	15	0	0	
riche	6	1	14	1	8	11
rond	9	1	10	2	4	33
sec	10	1	9	2	2	50
soufflé	7	3	30	3	2	60
sourd	9	0	0	5	1	83
vibré	7	3	30	2	3	40

Remarque : pour l'échelle liée à *résonnant*, nous avons regardé aussi les utilisations spontanées du verbe "résonner".

On a donc un certain nombre d'échelles dont les mots associés ont été utilisés spontanément par des sujets qui pourtant jugent l'échelle sémantique associée non pertinente. C'est le cas en particulier pour

plus de 50% des sujets qui ont utilisé spontanément les 6 mots suivants : *haut, fort, bas, aigu, grave* et *large* (le pourcentage de sujets les ayant utilisés et les considérant comme non pertinents étant respectivement de 100%, 89%, 80%, 60%, 58% et 50%). Ceci est d'autant plus fort que cela concerne plus de sujets : 9 sujets dans le cas d'*aigu* ont utilisé ce mot spontanément mais le considèrent comme non pertinent comme échelle sémantique. Certaines échelles sont par contre toujours considérées comme pertinentes quand le mot auquel elles sont associées a été utilisé spontanément. C'est le cas notamment des 5 échelles suivantes, liées à *métallique, nasillard, sourd, doux* et *diffus* (bien que cela ne concerne que 3 sujets dans le cas de *diffus*). Il y a quelques cas intermédiaires où le fait de considérer une échelle comme non pertinente alors que le mot a été utilisé spontanément ne concerne qu'un seul sujet. C'est les cas des 8 échelles suivantes, liées à *pincé, attaque, clair, sec, rond, net, riche* et *brillant*.

Les sujets qui ont participé aux deux expériences ont été aussi parfois confrontés lors de la deuxième expérience à des échelles correspondant à des mots qu'ils n'avaient personnellement pas utilisés spontanément dans la première expérience. Lorsqu'on leur propose un vocabulaire nouveau pour eux, l'acceptent-ils et le reconnaissent-ils comme pertinent ? Là aussi, cela dépend des échelles, qu'on répartit en deux groupes.

Pour un premier groupe d'échelles, la plupart des sujets qui n'avaient pas utilisé spontanément un mot trouvent néanmoins l'échelle adaptée globalement. Il s'agit par exemple de l'échelle liée à *doux* : un seul des nouveaux utilisateurs l'a considéré comme pas adapté (88% de sujets n'ont pas utilisé spontanément ce mot et ont considéré son échelle comme adaptée). C'est aussi le cas des échelles liées à *sourd, métallique* (83%), *attaque rapide* (75%), *clair* (67%). C'est aussi le cas de *long* et *pincé*, mais vu le nombre très faible de sujets qui n'avaient pas utilisé spontanément ces mots, les pourcentages n'ont plus tellement de sens ici. Pour les échelles liées à *nasillard* (71%) et *soufflé* (60%), deux des sujets n'ayant pas utilisé spontanément ces mots ont jugé leur échelle globalement inadaptée.

Dans un deuxième groupe d'échelles on trouve celles qui ont été souvent jugées comme inadaptées par les sujets qui n'avaient pas utilisé spontanément les mots sur lesquelles elles se basent. On retrouve les échelles liées à *grave, bas, haut* et *fort*, qui semblent apparaître à chaque fois qu'il y a des échelles dont la robustesse est mise en doute. Pourtant d'autres échelles apparaissent dans ce groupe, notamment celles liées à *brillant, rond* et *riche*, avec respectivement 44%, 33% et 11% seulement de leurs nouveaux utilisateurs qui les considèrent comme adaptées. Pourtant, ces termes font partie des rares termes que les musiciens utilisent de manière quasiment conventionnelle pour décrire les qualités des sons.

#### **I.1.4. Lien avec l'utilisation resserrée ou dilatée des échelles**

Nous avons repris la classification de l'utilisation des échelles par les sujets de manière dilatée, normale, resserrée ou très resserrée, définie et analysée dans le chapitre 4 §II.6, pour la confronter

avec les jugements de pertinence globale de chaque échelle par chacun des sujets. Notre hypothèse est qu'une échelle utilisée de manière très resserrée par un sujet correspondra à une échelle qu'il considère comme non pertinente. Notre deuxième hypothèse est qu'une échelle utilisée de manière dilatée par un sujet correspondra à une échelle qu'il considère comme pertinente. Les résultats sont rassemblés dans le tableau 7.6.

Tableau 7.6. Nombre d'échelles utilisées de manière très resserrée et dilatée et comparaison avec le jugement de pertinence effectué par les sujets concernés.

Echelles	Très resserrées	Dilatées
Nb total (sur 23 échelles *32sujets)	25	378
Nb pertinentes	1 (4%)	275 (73%)
Nb non-pertinentes	24 (96%)	103 (27%)

Nous trouvons donc, comme le prévoyait notre première hypothèse que dans 96% des cas où un sujet utilise une échelle de manière très resserrée, c'est qu'il la considère comme non pertinente. Il y a pourtant une exception : un sujet qui utilise une échelle de manière très resserrée (qui ne permet donc pas de discriminer les 12 sons écoutés) bien qu'il considère cette échelle comme globalement appropriée. C'est le cas du sujet n12m (musicien) qui, pour l'échelle liée à *soufflé*, positionne quasiment tous les sons à l'extrémité *peu soufflé* (0) sauf deux exceptions : le SNO positionné à 0.16 et le TBN positionné à 0.01. Ce sujet a aussi utilisé l'échelle liée à *résonnant* de manière très resserrée mais cette fois-ci en la jugeant pas pertinente. Ce sujet n'avait pas participé à notre première expérience et nous n'avons pas ses commentaires écrits. Nous ne pouvons donc expliquer son comportement qu'à partir des suppositions liées au caractère particulièrement synthétique des sons imitant des instruments à vent. En effet, la plupart des autres sujets ont choisi de juger le caractère plus ou moins *soufflé* des sons sur cette échelle en fonction de l'appartenance des instruments évoqués par ces sons à la classe des instruments à vent. Ce sujet semble avoir fait abstraction de connaissances évoquées par les sons pour ne juger qu'à partir des perceptions. Or les sons qui imitaient des sons d'instruments percussifs paraissaient plus vraisemblables que les sons imitant des instruments à vent, particulièrement statiques, linéaires et synthétiques. Il est donc tout à fait compréhensible d'avoir positionné tous ces sons à zéro si ce qui était recherché était la perception d'un souffle. Aucun de ces sons n'étant réellement fabriqué au moyen d'un souffle. Il est intéressant de remarquer que le son que ce sujet a fait ressortir de l'ensemble est le SNO, un son hybride de corde frottée et de piano, qui présentait justement le souffle résultat du frottement superposé au son de piano. Ceci nous permet de poser la question de ce qui est réellement jugé par un sujet : le son réellement entendu ou un mélange de cette perception et de tout ce le son évoque pour lui, en particulier l'instrument imité.

Par contre notre deuxième hypothèse, qu'une échelle utilisée de manière dilatée par un sujet est forcément une échelle qu'il considère comme pertinente, n'est validée que dans 2 tiers des cas. De plus on trouve que pour toutes les échelles, au moins un sujet l'utilise de manière dilatée bien qu'il la juge globalement non pertinente. On trouve globalement beaucoup d'échelles utilisées de manières

dilatées (378 sur 736 contre 279 normales 37 resserrées et 25 très resserrées). Le fait que c'est le comportement le plus courant et le comportement attendu dans ce type de paradigme (donc induit implicitement chez les sujets), il peut donc être plus facilement décorrélé de leur jugement de pertinence. Alors qu'elles sont considérées comme non pertinentes par les sujets, on retrouve là encore, parmi les échelles pour lesquelles on a le plus fort pourcentage d'utilisation dilatée, celles liées à *bas, haut, long, fort* et *aigu* (avec comme pourcentage d'utilisations dilatées non pertinentes respectivement 61, 61, 58, 54 et 50% des échelles utilisées de manières dilatées). On retrouve aussi parmi les échelles dilatées dont le pourcentage d'échelles non pertinentes est le plus faible celles liées à *doux, attaque, sourd* et *net* (respectivement 5, 9, 10 et 14%). Ce sont toutes des échelles globalement pertinentes pour plus de 80% des sujets.

La manière dont un sujet va répartir les sons sur les échelles sémantiques proposées est donc liée à la fois à la pertinence que chacune de ses échelles a pour lui et à un comportement plus ou moins conforme à l'attente qu'on a de lui : répartir les sons sur l'ensemble de l'échelle proposée.

## I.2. Commentaires et préférences des sujets

Nous allons analyser ici les commentaires sur les échelles données par certains sujets après avoir participé à l'expérience ainsi que les échelles qu'ils ont déclaré avoir préféré utiliser. Ceci concerne principalement les 24 premiers sujets qui ont passé la première version de l'expérience 2. En effet, avant que le tirage au sort de l'ordre des échelles ne soit automatisé par une nouvelle version de PsiExp, il se faisait à l'aide d'un chapeau [sic !] et les sujets devaient remplir la grille reproduite en annexe A.7.5.

### **I.2.1. Échelles préférées par les sujets**

Vingt-deux des 24 sujets ayant rempli la fiche de suivi de l'expérience ont donné à l'issue de l'expérience les 4 ou 5 échelles qu'ils avaient préférées parmi les 23 qui leur avaient été proposées au cours de cette expérience.

Tableau 7.7. Nombre de sujets ayant choisi une échelle comme une de ses quatre préférées.

<u>Échelles préférées</u>	
entre 0 et 3	bas (0), haut (0), fort (1), large (1), riche (1), diffus (2), grave (2), long (2), net(2), aigu (3), vibré (3)
entre 4 et 6	brillant (4), pincé (4), sec (4), soufflé (4), doux (5), rond (5), résonnant (6)
entre 7 et 10	clair (7), nasillard (8), sourd (8), attaque rapide/lente (9), métallique (10).

Si on considère pour chacune des 23 échelles le nombre de sujets qui l'ont choisi parmi leurs préférées, on obtient une différence de répartition du nombre de sujets sur les 23 échelles qui est significative ( $\chi^2(1)=9.97, p=.001$ ). Ici aussi, la différence de répartition entre les échelles par les sujets musiciens et non-musiciens n'est pas significative ( $\chi^2(20)=16.12, p=.71$ ).

Parmi les 22 sujets qui se sont exprimés sur leurs 4 échelles préférées, celle qui a été préférée par le plus grand nombre de sujets (*métallique*) ne l'a été que par 10 sujets donc moins de la moitié des sujets (soit seulement 45% des sujets). On peut donc considérer que globalement les préférences des sujets se répartissent sur plus de 4 échelles différentes. On trouve en effet parmi les 5 échelles qui ont eu la préférence du plus grand nombre de sujets (entre 7 et 10 sujets) celles liées à *métallique*, *attaque rapide*, *nasillard*, *sourd* et *clair*. Ces échelles avaient été considérées comme adaptées par plus de 50% des sujets au moins pour 9 des 12 sons présentés (à part pour *nasillard*, adaptée pour 7 sons seulement), et font partie des échelles considérées comme globalement les plus pertinentes (pour entre 71 et 91% des sujets). Il est par contre étonnant de ne pas trouver l'échelle liée à *doux* dans ce groupe, vu qu'elle a été considérée comme globalement pertinente par 90% des sujets.

Par contre deux échelles n'ont été citées par aucun des 22 sujets concernés, ce sont celles liées à *haut* et *bas*. Bien que synonymes de *aigu* et *grave*, elles n'avaient pas été utilisées spontanément par beaucoup de sujets. On trouve aussi 3 échelles choisies comme une des 4 préférées par un seul sujet : celles liées à *fort*, *large* et *riche*. A part les échelles liées à *large* et *riche*, globalement pertinentes pour la moitié des sujets, les 3 autres échelles font partie de celles ayant été considérées comme globalement pertinentes par le moins de sujets (moins de 30%).

Afin de comprendre l'utilisation des échelles considérées souvent parmi les moins pertinentes par l'ensemble des sujets et pourtant parmi les 4 préférées de quelques sujets, nous avons essayé de considérer les commentaires effectués par les sujets qui considéraient ces échelles comme une de leurs préférées, ainsi que leur éventuelle utilisation de ces termes de manière spontanée (voir tableau A7.I.5.) En effet, certains des sujets qui déclarent, contre l'avis général, qu'une échelle est parmi leurs préférées avaient été les principaux utilisateurs spontanés de ces termes lors de notre première expérience (ce qui relève d'une certaine logique). Par contre, certains de ces sujets avaient très peu, voir jamais, utilisé ces termes spontanément.

Ainsi, l'échelle liée à *aigu* fait partie des préférées de 3 des sujets (a3, a7, et n7) et est la première citée par 2 d'entre eux (a3 et n7). On peut remarquer que ces 3 sujets sont de "vrais" non-musiciens, n'ayant eu aucune formation et aucune pratique musicale (voir A.4.2). Deux de ces 3 sujets avaient participé à notre première expérience, et faisaient partie des 4 sujets qui avaient le plus utilisé spontanément ce mot. Parmi les 2 autres sujets qui avaient beaucoup utilisé ce mot spontanément (une trentaine de fois, lors de la comparaison des 66 paires de sons), un seul autre a participé aussi à la deuxième expérience, le sujet a12, qui n'est pas un vrai non-musicien puisqu'elle a pratiqué le violoncelle dans son enfance. Ce sujet a considéré l'échelle liée à *aigu* comme globalement adaptée sans faire de commentaire à la fin de cette expérience. Elle avait commenté son utilisation spontanée de *aigu* lors de l'expérience 1 comme ceci : "comme s'il était un petit peu plus aigu (même si c'est pas vrai du point de vue musical vu que c'est la même note)" et "ils ont l'air d'être plus aigus, enfin, c'est pas vraiment aigu, mais, on a l'impression qu'on les entend, dans une partie de l'oreille qui est différente, et qui est la partie de l'oreille qui sert à écouter les sons aigus d'habitude, je sais pas si

c'est très logique du point de vue euh de la médecine, mais bon, on va faire comme si ! / c'est comme si on les entendait un peu plus haut dans la tête". Cette utilisation, même si elle était pour elle spontanée, la dérangeait et l'obligeait à faire beaucoup de commentaires.

De même, pour *grave* deux sujets l'ont choisi parmi leurs échelles préférées : a11, qui remarque : "les sons paraissent plus graves : une note pincée paraît plus aiguë" et a3, qui ne fait pas de remarque. Tous les deux étaient des anciens sujets, 2 des 3 qui avaient utilisé ce mot le plus spontanément (respectivement 33 et 26 fois). De même encore, alors que seuls 23% des sujets considèrent l'échelle liée à *fort* comme pertinente globalement, elle fait partie des 4 préférées du sujet a12 qui commente : "très facile de décider si un son est fort ou non". Il est intéressant de remarquer que le sujet a12 est de loin celui qui a le plus utilisé ce mot spontanément dans l'expérience 1.

Par contre, pour l'échelle liée à *long*, deux sujets l'ont choisie parmi leurs préférées : le sujet a14m ("approprié à beaucoup de sons") et le sujet a3 ("long = note maintenue non (forcée ??)"). Ces deux sujets avaient pourtant peu utilisé ce mot spontanément (respectivement 3 et 2 occurrences). Le sujet a9, qui avait le plus utilisé ce mot spontanément (22 occurrences spontanées), considère l'échelle associée comme adaptée globalement mais pas parmi ses préférées. De même, l'échelle *diffus* est une des préférées du sujet a12 qui n'écrit pas de commentaires et n'avait pas utilisé spontanément ce mot. C'est aussi une des préférées du sujet n5, qui n'avait pas participé à la première expérience, et qui commente : "plus adapté que large", "les sons ont la même durée : large, est-ce l'attaque ? -> diffus". Il en va de même pour l'échelle *vibré*, parmi les préférées (1) du sujet a11 qui commente : "le plus pertinent en opposé à sourd" et en fait une utilisation spontanée faible (6 occurrences contre 58 fois pour certains) et (2) du sujet a16m qui ne donne pas de commentaires et n'avait jamais utilisé ce mot spontanément.

L'échelle liée à *large* (souvent peu pertinente et mauvaises corrélations inter-sujets, une des 4 préférées du sujet n8 qui commente : "généreux (riche) ou qui résonne". Ce sujet n'avait pas participé à la première expérience.

### **I.2.2. Commentaires sur les échelles**

Nous avons rassemblé en annexe (A.7.6 à A.7.29) l'ensemble des commentaires écrits par les sujets à la fin de l'expérience à propos de leur utilisation des différentes échelles sémantiques proposées. Nous avons classé ces commentaires en positifs ou négatifs (tableau 7.8).

Ont été classés "positifs" les commentaires qui essayaient d'expliquer comment cette échelle pouvait être utilisée dans la tâche donnée ou qui exprimaient la satisfaction des sujets vis-à-vis d'elle ; "négatif", les commentaires qui visaient à exprimer des difficultés ou à la déprécier (tableau 7.8).

Il est intéressant de remarquer que plus une échelle est considérée comme non pertinente, plus le nombre de commentaires positifs abonde. En effet, le nombre de commentaires positifs lorsqu'une

échelle est jugée pas adaptée est fortement anticorrélée avec le pourcentage de sujets ayant trouvé cette échelle adaptée ( $r(21)=-.72$ ,  $p<.01$ ). Ceci pourrait s'expliquer en considérant que ces mots ont été tous utilisés spontanément un grand nombre de fois. Ils ont donc forcément quelque chose d'adapté aux sons considérés. Les sujets les jugeant non adaptés se sentent obligés d'essayer de définir le contexte dans lequel cette échelle serait adaptée ou comment on pourrait quand même l'utiliser. C'est un moyen de dire que ces échelles sont à la fois pertinentes (puisqu'on leur trouve un sens) et non-pertinentes (jugées comme telles dans le contexte de notre expérience).

Tableau 7.8. Nombre de commentaires positifs ou négatifs écrits à la fin de l'expérience en fonction du jugement de pertinence de l'échelle sémantique considérée.

Échelle jugée Commentaires	adaptée		Pas adaptée		Pertinence globale % sujets
	positif	négatif	positif	négatif	
aigu	3	1	7	4	39
attaque	4	3	1	1	91
bas	6	0	4	3	29
brillant	5	4	1	6	62
clair	8	2	0	1	71
diffus	4	3	3	3	57
doux	9	2	0	0	90
fort	5	0	5	5	23
grave	3	4	5	3	40
haut	4	0	5	4	26
large	7	1	3	2	52
long	5	2	6	1	52
métallique	7	1	0	2	84
nasillard	4	3	5	0	78
net	11	2	0	3	81
pincé	7	5	0	2	68
résonnant	4	6	1	1	79
riche	7	4	3	4	55
rond	7	1	2	2	69
sec	9	0	1	2	74
soufflé	5	2	4	6	48
sourd	7	1	3	0	87
vibré	6	3	2	2	65

### I.2.3. Quand les sujets expliquent ce qu'ils ont l'impression de faire.

Plusieurs remarques écrites par les participants de cette expérience paraissent rejoindre nos propres questionnements et méritaient d'être soulignées.

- Par exemple, le sujet n4 écrit, vis à vis de l'échelle liée à *bas* : "à ma grande surprise, je pensais que mes jugements seraient plus identiques que ceux relatifs au terme *haut*. Or il n'y a pas d'inversion réelle. Il semble que le terme influence notre perception." Ceci rejoint des questions sur la possibilité de mettre en évidence un processus descendant (*top-down*) qui expliquerait qu'en jugeant les sons par rapport à une échelle donnée, on les perçoit différemment que lorsqu'on a à les juger par rapport à une autre échelle. Ceci rejoint aussi la discussion sur l'existence d'une perception "pure" et indépendante du langage et des concepts formés.

- D'autre part, plusieurs sujets remarquent que les qualificatifs du timbre sont liés à un mode de jeu, un style de musique, faisant ainsi se rejoindre les notions de timbre instrumental et de timbre orchestral. On peut ainsi considérer les deux remarques d'un même sujet musicien concernant les échelles liées à *doux* : "les sons sont difficiles à dissocier du type de musique qu'on fait avec" (a14m) ; et à *résonnant* : "difficile à dissocier de la façon dont il est joué" (a14m).

- Enfin certaines remarques des sujets renvoient à la difficulté de dissocier des jugements effectués à partir de nos perceptions et des jugements influencés par nos connaissances des concepts qu'on associe à nos perceptions. Par exemple, quand on doit juger le son TPT (imitation de son de trompette) par rapport à l'échelle *métallique*, le fait-on directement à partir de notre perception ou est-ce influencé par l'idée que l'on se fait de l'instrument qu'on y a associé : un instrument en métal de la famille des cuivres, utilisé dans les fanfares, ayant plutôt des interventions fortes, musclées. On peut ainsi considérer les remarques concernant les cinq échelles suivantes :

- *doux* : "ici *doux* est associé à "velouté". On tient également compte de l'attaque et de la catégorie de l'instrument (ex : percussion n'apparaît pas comme "douce" sauf si le bruit est quelque peu étouffé)" (n4) ;

- *pincé* : "absolument pas adapté. N'est pas évocateur comme qualificatif. On tient compte des connaissances (ex : clavecin)" (n4) ;

- *résonnant* : "la résonance n'est pas réelle ; elle est provoquée par l'imaginaire (instrumental ?)" (a15m) ;

- *riche* : "= type du son (identification) = synthétique" (n5) ;

- *sec* : "tendance à considérer les percussions comme sèches car brèves. Cependant, des sons non percussifs peuvent également être considérés comme tels, si considérés comme "nets, clairs"" (n4).

Cet effet est particulièrement présent vis à vis de l'échelle liée à *soufflé*, qui totalise 5 remarques allant dans ce sens : "= instrument à vent" (a5) ; "problème d'identification" (a8m) ; "pour moi dépend de la NATURE de ce qui produit le son" (n2m) ; "problème car facile à définir mais perception possible d'un souffle même si pas l'instrument" (a10m) ; "pas du tout adapté. Cette fois ci même si on a conscience de l'origine du son ou de la catégorie de l'instrument (aérophone), ça n'est pas du tout évocateur" (n4).

### I.3. Conclusion : comparaison entre ces différentes mesures

Nous avons rassemblé dans le tableau 7.9 l'ensemble des analyses effectuées sur les échelles sémantiques dans ce chapitre et dans le chapitre 4. Une analyse des corrélations entre ces différentes mesures (voir tableau A.7.30 en annexe) fait apparaître peu de corrélations significatives. Pourtant, le pourcentage de corrélations inter-sujets significatives pour une échelle donnée est significativement corrélé avec le nombre d'occurrences spontanées de ce terme ( $r(22)=.69$ ,  $p<.01$ ), ainsi qu'avec la



pertinence globale et le nombre de sujets ayant préféré cet échelle, bien plus faiblement (respectivement,  $r(22)=.43$  et  $.47$ ,  $p<.05$ ). De plus, le nombre de sujets ayant préféré une échelle est significativement corrélé à la pertinence (globale et par son) de cette échelle (respectivement,  $r(22)=.79$  et  $.66$ ,  $p<.01$ ). Sinon, on retrouve les corrélations significatives entre le nombre d'occurrences spontanées d'un terme et le nombre de sujets différents qui l'ont utilisé spontanément, entre la pertinence par son et la pertinence globale, et entre l'utilisation resserrée des échelles et la pertinence. Si on effectue une analyse de cluster sur les résultats rassemblés dans le tableau 7.9. (voir annexe A.7.31), on peut dégager les groupes d'échelles suivants. Ceci peut nous amener à une sorte de portrait-robot de ces 23 échelles testées dans cette expérience.

Tableau 7.9. Résumé des analyses sur les échelles sémantiques

Echelles	Spontané		pertinence globale % S (sur 32)	pertinence par son nb de T	corrélation IS % de cor avec $p<.1$	Echelles préférées nb S (sur 22)	Echelles resserrées	
	nb d'occ.	nb S (sur 20)					nb E<1	nb E>1
aigu	<b>252</b>	<b>19</b>	<b>39</b>	<b>2</b>	12	3	<b>7</b>	<b>25</b>
attaque	<b>499</b>	13	<b>91</b>	<b>12</b>	<b>43</b>	<b>9</b>	0	32
bas	52	7	<b>29</b>	<b>1</b>	10	<b>0</b>	2	30
brillant	23	<b>4</b>	62	5	8	4	0	32
clair	110	<b>16</b>	71	9	12	<b>7</b>	0	32
diffus	36	<b>3</b>	57	8	5	2	0	32
doux	77	10	<b>90</b>	<b>11</b>	12	5	0	32
fort	103	14	<b>23</b>	<b>1</b>	4	<b>1</b>	<b>8</b>	<b>24</b>
grave	188	<b>16</b>	<b>40</b>	<b>2</b>	7	2	<b>7</b>	<b>25</b>
haut	34	8	<b>26</b>	<b>0</b>	11	<b>0</b>	<b>8</b>	<b>24</b>
large	25	<b>4</b>	52	4	<b>3</b>	<b>1</b>	2	30
long	108	<b>17</b>	52	5	11	2	5	27
métallique	140	11	<b>84</b>	9	13	<b>10</b>	0	32
nasillard	121	11	78	7	8	<b>8</b>	0	32
net	166	12	81	<b>11</b>	14	2	0	32
pincé	<b>242</b>	14	68	5	<b>33</b>	4	0	32
resonn(ant)	101	14	79	9	19	6	1	31
riche	61	9	55	7	<b>2</b>	<b>1</b>	5	27
rond	118	11	69	8	7	5	3	29
sec	83	14	74	9	19	4	1	31
soufflé	94	14	48	3	20	4	<b>7</b>	<b>25</b>
sourd	122	10	<b>87</b>	<b>12</b>	14	<b>8</b>	2	30
vibré	174	13	65	<b>9</b>	5	3	4	28
Moyenne	122.04	11.13	61.74	6.48	12.70	3.96	2.70	29.30
Ecart-type	103.24	4.48	20.61	3.73	9.52	2.90	2.99	2.99

En gras, valeurs s'écartant de la moyenne de plus d'un écart-type

1- Tout d'abord, le comportement de l'échelle liée à *attaque rapide / lente* semble s'écarter beaucoup de toutes les autres échelles. D'une part, c'est la seule échelle bipolaire, ce terme n'ayant pas été extrait de comparaisons graduelles entre les sons, bien qu'il totalise le plus grand nombre d'occurrences spontanées. Mais c'est surtout une échelle qui se distingue par sa forte pertinence (pour 91% des sujets, tous les sons) et la plus forte corrélation inter-sujets. C'est aussi une des échelles préférées par le plus grand nombre de sujets et utilisée toujours de manière dilatée. Unanimité donc pour cette échelle qui permet de bien juger les 12 sons de notre ensemble, "première des différences

entre sons" pour le sujet a11, les seuls commentaires plus nuancés critiquant plutôt les extrémités de l'échelle : plusieurs sujets auraient préféré : attaque précise / floue, plutôt que rapide / lente. On peut remarquer aussi que cette échelle est corrélée principalement à la première dimension perceptive obtenue à partir de cet ensemble de sons et liée au temps d'attaque de ces sons (LTM\*).

2- Les 22 autres échelles se scindent en 2 grands groupes. Le premier est constitué des échelles liées à *pincé* et *aigu* d'une part, à *grave*, *vibré* et *net* d'autre part.

Le cas de l'échelle liée à *aigu* a déjà été beaucoup discuté. Ce terme est *apriori* inadapté dans cette expérience. En musique, il a un sens précis qui est lié à la notion de hauteur tonale, or cette dimension a été égalisée dans nos expériences, vu que tous les sons sont joués à la même hauteur (Mib4). Pourtant, ce mot est un des plus souvent utilisés spontanément (252 occurrences et 19 des 20 sujets), même par des gens qui savent que le sens conventionnel dans la musique pour ce mot n'est pas adapté (cf. remarques du sujets a12). La plupart des sujets à qui on le propose sous forme d'échelle sémantique ne le trouvent pas adapté. Et pourtant, il capte quelque chose de la réalité de ces sons là... On a vu chapitre 6 que son utilisation est essentiellement corrélée à la dimension 2 de notre espace perceptif, soit le centre de gravité spectral (CGS). Ceci est repris dans les commentaires des sujets : "idée de perception sonore sans base" (a10m), "il ne s'agit pas de fréquence mais de *l'angulosité* du son qui serait qualifiable pour moi de *pointu*" (a15m), "je l'ai pris au sens *nasillard*, *brillant*" (n1m). On peut aussi noter que c'est une des échelles qui a été le plus souvent utilisée de manière resserrée et qui obtient les jugements de pertinence (globale et par son) parmi les plus bas.

L'échelle liée à *pincé* se caractérise elle aussi par une des plus fortes utilisations spontanées, notamment en terme de nombre d'occurrences, mais au contraire de *aigu* elle présente une des plus fortes corrélations inter-sujets. Elle est toujours utilisée de manière dilatée et présente une bonne pertinence globale (68% des sujets) bien qu'elle ne soit en réalité pertinente que pour un petit sous ensemble de 5 ans. C'est en effet un terme qui se réfère au mode d'attaque des sons et permettrait plutôt de les séparer en 2 catégories : ceux dont l'émission imite un pincement (ou frappement) et ceux dont l'émission imite un souffle. On peut tout de même noter une utilisation différente : "pas dans le sens attaque mais plutôt "nez pincé"" (a8m).

Le comportement lié à l'utilisation de *grave* se rapproche assez de celui de son antonyme : *aigu* (bonne production spontanée, jugé peu pertinent, et beaucoup d'utilisations sous forme resserrée. Les commentaires recueillis vont aussi dans une compréhension de son sens liée à la présence d'harmoniques graves dans les sons entendus : "grave parce que plus sourd, plus d'harmoniques graves" (a10m), même s'ils soulignent la difficulté de juger les sons avec cette échelle : "difficile de l'entendre d'une oreille non-musicienne : toujours la même note !" (a14m) ou "la fréquence est identique" (a15m).

L'échelle liée à *vibré*, présente une meilleure pertinence globale (65%) et pour beaucoup de sons (9 sur 12), bien qu'elle ait été peu préférée par les sujets (3 fois) et qu'elle présente une des plus mauvaises corrélations inter-sujets. On observe aussi une grande diversité dans les commentaires : "trémolo" (a6), "très vibrés dans le sens "pincé"" (a9), "exprime la non constance d'un son et son "hésitation" entre plusieurs" (a14m), "proche de résonnant" (n3m), "? identique à nasillard ?" (a8m), "je ne suis pas fixée sur la notion de vibré" (n8). Il semble que la qualité jugée ne soit facile, ni à définir, ni à partager.

Quant à l'échelle liée à *net*, elle est jugée plutôt pertinente (pour 81% des sujets et 11 des 12 sons) et est toujours utilisée de manière dilatée, bien qu'elle apparaisse rarement parmi les échelles préférées des sujets (2 fois seulement). C'est l'échelle qui présente la plus forte corrélation avec la première dimension perceptive, de caractère temporel (cf. chapitre 5), ce qui est souligné dans les commentaires des sujets : "juge l'attaque plus autre chose..." (a8m), "plus l'attaque est rapide, plus il est net" (a9), "parallèle à attaque + clair. Sons soufflés moins nets ?"(a10m), "= pincé, attaque, + ou - sec" (a11). On peut tout de même noter quelques hésitations : "pas assez précis : net dans quelle partie du son : attaque ou ensuite ?" (n2m) et "j'oscille entre écouter la netteté du timbre et celle de l'attaque" (a15m).

Les 17 échelles restantes peuvent se scinder en 3 sous-groupes.

3- Le premier est composé des échelles liées à *métallique*, *sourd*, *nasillard*, *rond*, *clair* et *résonnant*, d'une part, *sec* et *doux* d'autre part. Il rassemble des échelles appréciées par les sujets pour décrire nos sons.

L'échelle liée à *métallique* apparaît comme une des plus pertinentes (pour 84% des sujets et 9 des 12 sons) et la plus souvent préférée par les sujets (10 fois). Elle est en outre toujours utilisée de manière dilatée. Pourtant, même s'il permet de bien discriminer les sons ("bien discriminant, mot utile" (n7)), son sens peut être discuté : "en fait, j'ai beaucoup de difficultés avec ce mot car je ne sais pas comment l'employer : sens métal ? ou sens brillant ? pas réussi à me décider." (n1m). On trouve en effet des définitions assez variées : "=aigu, froid, corde" (a6), "lié à pincé, sec et opposé à sourd" (a11), "proche de la couleur des sons" (a14m), envie d'assimiler "métallique" et "froid"... Ceci est peut-être à lier avec le fait que l'utilisation de cette échelle est franchement composite des dimensions perceptives 1 et 2 (cf. chapitre 5), un son apparaissant comme plus métallique si son attaque est rapide et son centre de gravité spectral élevé (présence d'harmoniques aigus).

L'échelle liée à *sourd*, qui lui est souvent opposée, est elle aussi globalement très pertinente (pour 87% des sujets et tous les sons). Elle est souvent parmi les préférées des sujets. Elle semble représenter une manière plus "convenable" de dire d'un son qu'il est *grave* : "=grave mais long" (a3), "=grave, gras" (a6), même si certains sujets le prennent dans le sens "qui n'entend pas" : "j'associe à l'intensité sonore." (n2m et a10m) ou "j'ai considéré "sourd" comme "étouffé" et opposé à "clair"" (n4), "tous sourds (= étouffés)" (a9). Certains sujets semblent avoir été surpris au premier abord qu'on puisse qualifier un son de *sourd* si cela peut vouloir dire : qui n'entend pas.

L'échelle liée à **nasillard** est elle aussi parmi les plus pertinentes (pour 78% des sujets et 7 sons) et les plus préférées, bien que dans une moindre mesure. Ce terme a été pourtant utilisé spontanément moins souvent. Souvent considéré comme un synonyme de métallique pour décrire le timbre des sons, nous avons déjà vu chapitre 5 qu'il s'en distingue pourtant beaucoup, qualifiant plutôt des sons ayant une attaque lente. Il est souvent associé à "aigu et désagréable" (n8 et n5).

L'échelle liée à **rond** apparaît dans la moyenne pour tous nos indicateurs. Il qualifie plutôt des sons perçus comme agréables : "chaud, voluptueux" (a4), "un son rond = un son doux, grave, pas incisif" (n3m), "serein, doux, pas strident, pas nasillard, habituel, simple" (n8).

L'échelle liée à **clair** se démarque surtout par le grand nombre de sujets différents qui l'on utilisée spontanément (16 sur 20), mais aussi par le fait qu'elle apparaît parmi les échelles préférées de 7 sujets. Là aussi il s'agit d'une qualité agréable dont la définition oscille entre un caractère plutôt *aigu* ou plus proche de la notion de luminosité : "lumineux, léger, pur, froid" (a6), "brillant = clair" (n3m), "=simple" (n5), "agréable, résonnant" (n8), "entre la couleur et la netteté ?" (a14m).

Quant à l'échelle liée à **résonnant**, elle apparaît dans la moyenne pour tous nos indicateurs. On peut noter tout de même que spontanément, les sujets produisent plus souvent des occurrences du verbe résonner que le mot *résonnant* lui-même. De plus, cette qualité semble se référer au mode de production des sons : "la résonance n'est pas réelle, elle est provoquée par l'imaginaire (instrumental ?)" (a15m), "difficile à dissocier de la façon dont il est joué" (a14m), l'impression de présence d'un "écho" (a5), d'un effet de longueur étant ici contrarié à la fois par le caractère synthétique des sons et le fait qu'ils étaient égalisés en longueur.

L'échelle liée à **sec** apparaît elle aussi dans la moyenne pour tous nos descripteurs. Elle est souvent associée à l'échelle liée à *net* : "= bref" (a3) ; "sec = net" (n3m) et à un caractère nettement temporel : "= fin du son brutale ou non" (a5), "cf. net ? notion d'attaque + résonance" (a8m), "très sec = attaque rapide" (a9 et a11).

Enfin, l'échelle liée à **doux** est caractérisée par une des plus fortes pertinences (pour 90% des sujets et 11 des 12 sons). Elle est toujours utilisée de manière dilatée. C'est toujours associé à un caractère agréable : "agréable" (n8), "ici "doux" est associé à "velouté"" (n4), même si les sujets discutent de son lien avec la brillance : "cela me semble être un qualificatif composite : attaque lente et peu brillant" (n3m), "Corrélié avec brillance" (n1m), "juge la brillance ?" (a8m).

4- Le deuxième sous-groupe est composé des échelles liées à *soufflé, long et fort*.

L'échelle liée à **soufflé** se caractérise surtout par un grand nombre d'utilisations de manière resserrée. Elle est peu pertinente (48% des sujets) mais surtout caractérise très peu de sons de notre

ensemble (3 seulement). Pourtant, elle décrit *a priori* une des qualités les plus discriminantes pour notre ensemble de sons : son mode d'attaque. On retrouve d'ailleurs ceci dans les remarques des sujets : "pour moi dépend de la NATURE de ce qui produit le son" (n2m), "instrument à vent" (a5), "= à vent" (a6). La réserve liée à son utilisation est peut-être liée au caractère particulièrement synthétique des sons imitant des sons soufflés : "pas du tout adapté. Cette fois-ci même si on a conscience de l'origine du son ou de la catégorie de l'instrument (aérophone) ça n'est pas du tout évocateur" (n4)

L'échelle liée à **long** se caractérise surtout par une utilisation spontanée par un grand nombre de sujets différents. Comme la longueur des sons présentés avait été égalisée, les sujets ont cherché à définir ce qu'ils pouvaient percevoir comme *long* toujours associé à la fin du son : "long = note maintenue (forcée?)" (a3), "= écho" (a5), "juge la résonance" (a8m), "long = pas net, la fin du son se prolonge" (n3m), "sensation de même longueur mais différences de résonances après l'attaque" (n2m). Une autre difficulté réside dans les 360 ms de durée moyenne des sons entendus : "mais ici les sons semblent trop courts pour être appelés longs ! (souvent utilisé dans le sens négatif : "pas long")" (a12)

L'échelle liée à **fort** se caractérise surtout par le faible nombre de sujets qui l'ont jugée pertinente (23% pour 1 seul son) et le grand nombre d'utilisations de façon resserrée. Il s'agit là aussi d'une qualité *a priori* égalisée sur l'ensemble des sons (sonie). Les sujets compensant par un grand nombre de commentaires : "dans sa présence revendication, si rayonne ou pas" (a4), "évoque la puissance et la persistance du son" (a14m), "= aigu, clair, stress" (a6), "j'ai l'impression d'associer fort et strident" (a7), "trouve fort dans la présence sonore, l'attaque" (a10m)... bien que le fait que le "volume sonore" de tous les sons soit le même, est souvent noté (a5, a15m).

5- Enfin le dernier sous-groupe d'échelles est constitué des échelles liées à *haut, bas, riche, diffus, large et brillant*. Il est principalement caractérisé par un nombre d'occurrences spontanées parmi les plus bas (inférieur à 52) et une très faible préférence des sujets.

Les échelles liées à **haut** et **bas**, sont parmi les moins pertinentes globalement (moins de 29% des sujets et pour au maximum 1 son). Elles ne sont jamais préférées par les sujets, présentent des corrélations inter-sujets plutôt faibles et l'échelle liée à haut est une de celles qui est utilisée le plus souvent de manière resserrée. Elles sont utilisées souvent comme synonymes de *aigu* et *grave*, comme le soulignent les commentaires des sujets : "haut = aigu" (a3), "où est la différence avec aigu ?" (a14m) ou "bas = grave" (a3 et a6), "parallèle à grave" (a10m).

Les sujets ont souligné là aussi que la hauteur (en terme de fréquence) était pourtant la même pour tous les sons : "ils sont tous à la même hauteur" (a4 et n8), "tous ces sons sont de même hauteur (fréquence) pour moi un son peut être haut par rapport à un diapason, une fréquence recherchée" (a15m), "la note ne change pas" (a16m), "la hauteur est la même pour moi dans tous ces sons" (n2m).

L'échelle liée à **riche** se caractérise surtout par la très faible corrélation inter-sujets dans son utilisation, et le fait qu'elle ne soit préférée que par un seul sujet. Pourtant, ce terme est souvent utilisé pour parler

du timbre des sons de manière usuelle. Ceci est peut-être dû au caractère synthétique des sons présentés : "tes *sons* me semblent plutôt pauvres !" (a12), "gène des sons synthétiques qui cherchent à imiter des sons naturels" (n2m), "ces sons synthétiques peuvent difficilement être qualifiés de riches à mon avis, ils sont assez simples" (n3m).

L'échelle liée à *diffus* a été celle utilisée spontanément par le moins de sujets différents (3). Son sens a été difficile à saisir par l'ensemble des sujets entre la notion de netteté, et celle de précision : "précis / imprécis ?" (a10m), "j'associe la diffusion à la longueur du son" (n2m), "un son qualifié de "diffus" serait un son dont l'extinction n'est pas nette." (n3m), "proche d'*étouffé*" (n4), "diffus différent de bref et sec" (a3).

L'échelle liée à *large* a été elle aussi utilisée spontanément par très peu de sujets différents (4). Elle présente de plus une des plus faibles corrélations inter-sujets. Son sens semble en effet varier suivant les sujets entre quelque chose d'ordre temporel : "large différent de bref" (a3), "cf. résonnant ?" (a8m), "sens de long" (a9) et quelque chose plus lié à la richesse : "large = grave, le contraire serait pincé" (n3m), "généreux (riche) ou qui résonne" (n8)...

Enfin, l'échelle liée à *brillant* est elle aussi surtout caractérisée par un faible nombre de sujets différents qui l'ont utilisée spontanément (4). C'est pourtant un des termes les plus couramment utilisé pour décrire le timbre des sons. Notre remarque rejoint celle concernant riche : "aucun ne brille ou presque car tous "synthétiques"" (a9), "la luminosité est plutôt dans le sombre (terne)" (a4), voir plus catégorique : "n'est pas représentatif d'une qualité du timbre" (n4).

## II. Expérience 3 : sens des portraits

Dans ce paragraphe vont être analysés les jugements sur le sens des portraits proposés aux sujets lors de l'expérience 3. En effet, les sujets, après avoir fait chacune des conditions, revoyaient défiler les portraits auxquels ils avaient associé les sons et devaient donner leur avis pour chacun d'eux : est-ce que pour vous (dans ce contexte) ce portrait a un sens "précis", "vague" ou "pas de sens" du tout. Nous allons tout d'abord analyser ces jugements de sens, puis les comparer aux autres données recueillies, notamment aux jugements de pertinence analysés dans le paragraphe précédent.

### II.1. Jugement de sens des portraits

Nous allons donc analyser ici les jugements recueillis à la fin de chaque passation d'un sous-groupe de portraits sur le sens de chacun d'entre eux ("précis", "vague" ou "pas de sens"). Ces résultats seront analysés d'abord en moyenne sur l'ensemble des sujets, puis en considérant des groupes de sujets différents : musiciens et non-musiciens.

#### **II.1.1. Jugement de sens en fonction du portrait proposé**

On peut regarder la répartition des réponses des sujets dans les trois catégories proposées (sens précis / vague / pas de sens), en fonction du type de portrait (selon qu'il décrivent le son ou la source sonore), puis en fonction des sous-groupes de portraits et de chacun des portraits qu'ils contiennent. Dans ces conditions, on remarque tout d'abord que sur l'ensemble de tous les portraits testés, les sujets considèrent en moyenne que 54% des portraits proposés ont "un sens précis", 36% ont "un sens vague" et 10% "pas de sens" à leur yeux dans le contexte.

#### Dépendance du type de portrait

Nous avons fabriqué les portraits à tester dans l'expérience 3 dans 2 catégories, selon qu'ils décrivaient une source sonore possible des sons ou qu'ils décrivaient des qualités de ces sons. La répartition des jugements sur le sens des portraits en fonction du type de portraits proposés est présentée dans le tableau 7.10.

Tableau 7.10. Répartition (en pourcentage) des réponses de sens des portraits en fonction du groupe ou sous-groupe auquel il appartient.

Portraits décrivant :	TOUS	le son	la source	famille d'instruments	Autres sources
% sens Précis	54	57	52	70	52
% sens Vague	36	36	36	27	36
% Pas de sens	10	7	12	3	12

Il y a un lien entre l'appartenance des portraits aux deux grandes catégories : "décrit le son" / "décrit la source" et le choix de la réponse sur le sens de ceux-ci ( $\chi^2(2)=21.8, p= 2.10^{-5}$ ). En effet, on observe une sous représentation des réponses "pas de sens" pour l'ensemble des portraits appartenant au

groupe "décrit le son" (7% seulement des réponses) et une sur représentation de ces réponses ("pas de sens") pour les portraits appartenant au groupe "décrit la source" (représente 12% des réponses).

Dépendance du sous-groupe de portrait

De même, si on regarde parmi les portraits décrivant la source, les 6 sous-groupes de portraits (source, excitateur, matériau, famille d'instruments, instruments et noms de fichiers), il y a un lien significatif entre le type de jugement de sens des portraits et leur appartenance à chacun des 6 sous-groupes ( $\chi^2(10)=33.9$ ,  $p= 2.10^{-4}$ ). En effet, on observe une sur représentation des portraits dont le sens est considéré comme "précis" dans le sous-groupe de noms de famille d'instruments, et une sous-représentation de ceux considérés comme n'ayant "pas de sens" dans ce même sous-groupe. Par contre, parmi les portraits décrivant le son, les 4 sous-groupes de portraits (qualités simples utilisées dans l'expérience 2, autres qualités simples, qualités doubles et qualités triples), il n'y a pas de lien entre le jugement de sens des portraits et le sous-groupe auquel le portrait appartient ( $\chi^2(6)=9.8$ ,  $p= 0.13$ ).

Dépendance du portrait

Si on regarde parmi chacun des sous-groupes de portraits d'éventuelles différences de jugement de sens de ces portraits en fonction du portrait considéré, on observe les résultats suivants. Pour ce qui concerne les portraits décrivant une source sonore possible pour les sons, les résultats obtenus en moyenne sur l'ensemble des sujets sont rassemblés dans le tableau 7.11.

Tableau 7.11. Pourcentage des réponses de sens des portraits en fonction du sous-groupe de source auquel il appartient.

Sens	Source	Excitation	Matériau	Famille	Instrument	Fichiers
précis	48	46	56	71	53	56
vague	38	41	32	27	34	36
pas de sens	14	13	12	3	12	8

Il n'y a pas de lien significatif entre le jugement de sens et un des 26 portraits du sous-groupe des sources (combinaison d'un mode d'excitation et d'un matériau) ( $\chi^2(50)=48.04$ ,  $p= 0.55$ ). Il en est de même pour les 17 portraits du sous-groupe des modes d'excitation ( $\chi^2(32)=26.77$   $p= 0.73$ ) et pour les 12 portraits du sous-groupe des matériaux ( $\chi^2(22)=14.72$   $p= 0.87$ ).

De même, il n'y a pas de lien significatif entre le jugement de sens et un des 12 portraits du sous-groupe des noms de fichiers ( $\chi^2(22)=22.11$   $p= 0.45$ ) et ce, bien que le portrait "piano échantillonné" recueille 20% de réponses "pas de sens", contre 8% seulement en moyenne pour l'ensemble des portraits de ce sous-groupe. Il en est de même pour les 37 portraits de noms d'instruments ( $\chi^2(70)=52.43$   $p= 0.94$ ).

Par contre, pour ce qui est des 6 portraits noms de familles d'instruments, qui se démarquent déjà en moyenne des autres sous-groupes de portraits décrivant une source sonore possible, il y a un lien significatif entre le choix de la réponse et un des 6 portraits de ce sous-groupe ( $\chi^2(10)=19.46$   $p= 0.03$ ). En effet, il y a une sur représentation des réponses "pas de sens" liées au portrait : *un*



*instrument de la famille des percussions* (10% des réponses) et une sur représentation des réponses "un sens vague" liées au portrait *un instrument de la famille des bois* (50% des réponses).

Pour ce qui concerne les portraits décrivant des qualités des sons, les résultats obtenus en moyenne sur l'ensemble des sujets sont rassemblés dans le tableau 7.12.

Tableau 7.12. Pourcentage des réponses de sens des portraits en fonction du sous-groupe décrivant le son auquel il appartient.

Adjectifs	simples (expé.2)	autres simples	doubles	triples
Précis	60	58	54	
Vague	31	35	41	
Pas de sens	9	7	6	

Il n'y a pas de lien significatif entre le jugement de sens et un des 21 portraits du sous-groupe des adjectifs simples utilisés dans l'expérience 2 ( $\chi^2(40)=33.22$ ,  $p= 0.77$ ). Il en est de même pour les 16 portraits du sous-groupe des autres adjectifs simples ( $\chi^2(30)=29.9$ ,  $p= 0.47$ ), pour les 21 portraits du sous-groupe des adjectifs doubles ( $\chi^2(40)=41.52$ ,  $p= 0.42$ ) et pour les 12 portraits du sous-groupe des adjectifs triples ( $\chi^2(22)=16.02$ ,  $p= 0.81$ ).

### II.1.2. Jugement de sens en fonction de la compétence musicale des auditeurs

Nous avons voulu tester d'éventuelles différences de comportements entre les sujets, notamment en fonction de leur compétence musicale. Si on considère la totalité des portraits testés, on n'observe pas de lien entre l'expertise des sujets et les réponses qu'ils ont données à la question du sens des portraits ( $\chi^2(2)=1.57$   $p= 0.46$ ).

#### Types de portraits

Pourtant, si on considère seulement les portraits qui décrivent le son, la différence entre les réponses des musiciens et celles des non-musiciens devient significative ( $\chi^2(2)=17.6$   $p= 10^{-4}$ ). En effet, on observe une sur-représentation des réponses "pas de sens" pour les musiciens et une sous-représentation pour les non-musiciens (5% de leurs réponses alors qu'elles concernent 9% de celles des musiciens). Réciproquement, les non-musiciens ont tendance à considérer plus souvent que ces portraits ont un sens "précis" que les non-musiciens (62% contre 52% des portraits considérés).

On observe le phénomène inverse quand il s'agit de l'ensemble des portraits décrivant une source sonore. La différence de comportement est significative ( $\chi^2(2)=8.35$   $p= .01$ ), bien que la force du lien entre l'expertise et la réponse choisie soit plus faible ( $\phi = 0.06$  contre 0.1 pour les portraits décrivant le son). On observe ici en effet une sur-représentation des musiciens (et une sous-représentation des non-musiciens) qui ont considéré l'ensemble de ces portraits comme ayant un sens précis (56% pour les musiciens contre 49% pour les non-musiciens). Par contre, on observe une légère sur-représentation des non-musiciens (et une sous-représentation des musiciens) qui ont considéré ces portraits comme de sens "vague" (respectivement 38% pour les non-musiciens contre 33% pour les musiciens).

Tableau 7.13. Répartition des jugements sur le sens des portraits des différents groupes suivant l'expertise des sujets.

Adjectifs %	le son		la source	
	musiciens	non-musiciens	musiciens	non-musiciens
Précis	52	61	56	49
Vague	39	34	33	39
Pas de sens	9	5	11	12

Sous-groupes de portraits

En fait, cette différence de réponses entre les deux groupes de sujets ne se retrouve pas systématiquement dans tous les sous-groupes de portraits considérés.

En effet, si on considère les 4 sous-groupes de portraits décrivant le son (adjectifs simples utilisés dans l'expérience 2, autres adjectifs simples, adjectifs doubles et triples), les différences entre les comportements des deux groupes de sujets ne sont significatives que dans 2 cas (voir tableau 7.14)

- Pour les portraits formés à partir des adjectifs simples utilisés dans l'expérience 2 ( $\chi^2(2)=11.89$   $p=0.003$ ). Cela est dû à une sur-représentation des musiciens qui considèrent ces portraits comme n'ayant pas de sens (13% pour les musiciens contre 5% pour les non-musiciens).

- Pour les portraits formés à partir d'adjectifs doubles ( $\chi^2(2)=11.02$   $p=0.004$ ). Ce qui est dû un peu au même phénomène (8% de ces portraits sont considérés comme n'ayant pas de sens par les musiciens contre 3% pour les non-musiciens) mais aussi au fait que les non-musiciens considèrent le sens de ces portraits plus souvent comme précis (pour 61% d'entre eux contre 46% pour les musiciens).

Tableau 7.14. Répartition des jugements sur le sens des portraits des différents sous-groupes décrivant le son en fonction de l'expertise des sujets.

Adjectifs %	simples (expé.2)		doubles	
	musiciens	non-musiciens	musiciens	non-musiciens
Précis	53	67	46	61
Vague	34	29	46	36
Pas de sens	13	5	8	3

Par contre, on n'observe pas de lien significatif entre l'expertise et le choix de réponse de sens des sujets dans le cas des deux autres sous-groupes décrivant les qualités des sons : adjectifs simples utilisés spontanément pour catégoriser les sons ( $\chi^2=3.13$   $p=0.21$ ,  $df=2$ ) ou adjectifs triples ( $\chi^2(2)=.94$ ,  $p=0.62$ ).

D'autre part, pour ce qui concerne les 6 sous-groupes de portraits qui décrivent la source sonore, les différences de réponses à la question du sens des portraits ne dépendent de l'expertise musicale que pour 2 de ces sous-groupes (voir tableau 7.15) :

- Celui concernant les portraits de noms de fichiers (instruments imités ou mélanges) ( $\chi^2=11.37$ ,  $p=0.003$ ,  $df=2$ ). Cela est dû à une sur-représentation des portraits jugés précis pour les musiciens et une sous représentation de ces réponses chez les non-musiciens (elles représentent respectivement 66% et 45% de leurs réponses).

- Celui concernant les portraits de noms de famille d'instruments ( $\chi^2=10.98$   $p=0.004$ ,  $df=2$ ). Ce qui est dû surtout à la sur-représentation des réponses "sens vague" chez les non-musiciens (contre une sous-représentation chez les musiciens), représentant respectivement 40 et 13% de leurs réponses, mais aussi à la sur-représentation des réponses "sens précis" chez les musiciens (contre une sous-représentation de ces réponses chez les non-musiciens), représentant 83% des réponses des musiciens contre seulement 58% de celles des non-musiciens).

Tableau 7.15. Répartition des jugements sur le sens des portraits des différents sous-groupes décrivant une source sonore possible en fonction de l'expertise des sujets.

Adjectifs %	noms de Fichier		Famille d'instruments	
	Musiciens	Non-Musiciens	Musiciens	Non-Musiciens
Précis	66	45	83	58
Vague	29	43	13	40
Pas de sens	5	12	3	2

Par contre, on n'observe pas de lien significatif entre l'expertise des sujets et leur jugement de sens dans le cas des 4 autres sous-groupes : noms de source (association d'un mode d'excitation et d'une matière) ( $\chi^2(2)=0.08$   $p=0.96$ ), noms de mode d'excitation ( $\chi^2(2)=5.16$   $p=0.08$ ), noms de matériaux ( $\chi^2(2)=0.2$   $p=0.9$ ) ou même noms d'instruments ( $\chi^2(2)=2.76$   $p=0.25$ ).

## II.2. Comparaisons avec les autres données

Les portraits construits pour l'expérience 3 l'ont été à partir des verbalisations libres récoltées lors de notre première expérience. Il peut donc être intéressant de comparer les jugements de sens des portraits donnés par les sujets avec les conditions de leur production spontanée (par exemple leur nombre d'occurrences spontanées et le nombre de sujets différents les ayant produites). De plus pour certains de ces portraits, nous pouvons comparer les jugements de sens des portraits avec les jugements de pertinence des échelles correspondantes analysés dans le paragraphe précédent.

### **II.2.1. Lien avec les jugements de pertinence de l'expérience 2**

Comme une partie du matériel verbal (les adjectifs produits spontanément dans l'expérience 1 pour comparer les sons 2 à 2) a été testé dans les expériences 2 et 3, nous disposons alors à la fois des jugements de pertinences des échelles construites à partir de ces adjectifs (par exemple l'échelle "peu doux - très doux" est adaptée/pas adaptée à l'ensemble des sons ou à ce son) et des jugements sur le

sens des portraits correspondants (par exemple "un son doux" a pour vous un sens précis / vague / pas de sens" dans ce contexte). Nous pouvons donc comparer ces jugements (voir tableau 7.16).

Tableau 7.16. Comparaison du nombre de sujets (sur 20) ayant jugé le sens des portraits dans l'expérience 3 (E3) et du nombre de sujets (sur 32) ayant jugé de la pertinence globale des échelles sémantiques associées dans l'expérience 2 (E2).

NB sujets	E3: Sens (20 sujets)			E2: pertinence (32 sujets)	différence P/Ad	différence V/Ad	différence PDS/Ad	différence P+V/Ad
	P	V	PDS	Adaptée	<sup>2</sup> (p)	<sup>2</sup> (p)	<sup>2</sup> (p)	<sup>2</sup> (p)
aigu	9	11	0	12	0.10 ns	0.77 ns	7.74 *	7.42 *
attaque	9	7	3	29	3.11 ns	5.09 †	11.15 *	0.06 ns
brillant	11	6	3	20	0.09 ns	2.96 ns	5.69 †	0.79 ns
clair	14	3	2	23	0.00 ns	7.25 *	9.20 *	0.51 ns
diffus	16	2	2	18	0.94 ns	6.73 *	6.73 *	1.91 ns
doux	11	8	1	29	1.93 ns	4.62 †	15.32 *	0.02 ns
fort	11	7	2	7	3.58 ns	0.45 ns	0.99 ns	10.90 *
grave	12	4	3	13	1.14 ns	1.17 ns	2.16 ns	4.08 †
large	11	7	2	17	0.04 ns	0.92 ns	5.60 †	2.59 ns
long	12	7	1	17	0.20 ns	0.92 ns	7.70 *	3.32 ns
métallique	9	7	3	27	2.35 ns	4.14 †	9.89 *	0.00 ns
nasillard	13	7	0	25	0.22 ns	4.05 †	15.63 *	0.68 ns
net	13	4	2	26	0.35 ns	7.32 *	11.16 *	0.09 ns
pincé	14	4	1	22	0.04 ns	5.08 †	10.86 *	1.12 ns
résonnant	14	6	0	25	0.08 ns	5.49 †	15.86 *	0.60 ns
riche	13	6	1	18	0.29 ns	2.03 ns	8.33 *	2.76 ns
rond	12	6	2	22	0.11 ns	3.90 †	8.78 *	0.65 ns
sec	13	7	0	24	0.21 ns	2.87 ns	14.83 *	0.97 ns
soufflé	8	9	3	15	0.25 ns	0.04 ns	3.42 ns	2.43 ns
sourd	12	6	2	28	1.08 ns	6.74 *	12.28 *	0.00 ns
vibré	13	7	0	21	0.01 ns	2.28 ns	12.90 *	1.98 ns

avec \*: p<.01, †: p<.05, ns: p>.05, P: sens précis, V: sens vague, PDS: pas de sens, Ad : échelle globalement adaptée.

Si on compare, pour chacun des adjectifs, la proportion de sujets ayant jugé le portrait comme ayant "un sens précis" et la proportion de sujets qui ont considéré l'échelle "globalement adaptée" pour l'ensemble des sons, aucune des différences n'est significative. Les sujets ont donc en général jugé du sens précis des portraits du type "un son brillant" de la même façon que de la pertinence des échelles associées "peu brillant - très brillant". Par contre, si on compare les jugements de "sens vague" des portraits aux jugements de pertinence des échelles (globalement adaptées), des différences significatives apparaissent dans 52% des cas. Cette catégorie (sens vague) étant intermédiaire, il est difficile d'interpréter la comparaison avec le jugement d'échelle adaptée.

Mais si on compare la somme des jugements de "sens précis" et "vague" des portraits aux jugements de pertinence des échelles (globalement adaptées), des différences significatives n'apparaissent plus

que pour 3 adjectifs (14% des cas) : *fort, aigu, et grave*. En effet, en général, peu de sujets déclarent qu'un des portraits construits à partir d'un de ces adjectifs n'a pas de sens du tout (au maximum 3 sujets différents sur 20 au total). Par contre, les échelles liées aux trois adjectifs *fort, aigu, et grave*, ont été jugées "pas adaptées" par plus de 19 sujets sur 32. Il semble donc que les sujets jugent plus facilement qu'une échelle n'est pas adaptée plutôt que le portrait associé n'a pas de sens.

Par contre, si on compare la somme des jugements des portraits n'ayant "pas de sens" aux jugements de pertinence des échelles (globalement adaptées), des différences significatives apparaissent dans 86% des cas. En fait, 62% des différences entre le nombre de sujets ayant jugé chacun des portraits comme n'ayant pas de sens et le nombre de sujets ayant jugé l'échelle associée comme "pas adaptée" sont significatives. On retrouve donc ici le fait que globalement, les sujets ont plus facilement jugé qu'une échelle n'était pas adaptée globalement plutôt que le portrait associé n'avait pas de sens.

On peut trouver deux explications à cela. Pour qu'une échelle soit globalement adaptée, il faut qu'elle convienne pour juger de la plupart des sons proposés (c'est d'ailleurs ce qu'on retrouve dans la forte corrélation entre le jugement de pertinence son par son et le jugement de pertinence globale des échelles). Par contre, pour que le portrait associé ait un sens, même vague, il suffit qu'il convienne à un des sons. L'autre explication possible c'est que dans le cas des jugements de sens des portraits, le fait de proposer 3 catégories différentes dans le jugement de sens (précis / vague / pas de sens) permet aux sujets de nuancer leur propos, ils ont donc peut être moins facilement choisi le jugement plus "catégorique" de "pas de sens".

### II.2.2. Lien avec le nombre de productions spontanées

Nous avons comparé le nombre de sujets ayant jugé que chacun des portraits avait un sens précis dans l'expérience 3 au nombre de sujets différents qui avaient produit spontanément ces verbalisations dans l'expérience 1 d'où ces portraits sont extraits. Nous donnerons les résultats de ces comparaisons par sous-groupes de portraits, les données étant consignées en annexe (tableaux A7.II.1 à A7.II.9). Ici aussi, nous commencerons par les portraits décrivant une source sonore possible puis ceux décrivant des qualités des sons.

Pour ce qui est des portraits construits à partir de noms de **sources** sonores (association d'un mode d'excitation et d'un matériau), il s'agit tout d'abord de préciser qu'il y a de grandes différences entre les 26 portraits qui ont été testés dans l'expérience 3. En effet, certains dérivent des locutions très fréquentes chez un grand nombre de sujets différents (par exemple, le portrait : "on pince une corde" est construit à partir des 184 occurrences observées de locutions du type : "une corde pincée", "on pince une corde", "une corde qu'on pincerait"... produites par 14 sujets différents dans l'expérience 1). Par contre, d'autres locutions n'ont été produites que par un sujet pour décrire un son précis (par exemple : "une pièce de monnaie sur des cordes").

Nous avons donc comparé ici d'une part le nombre de sujets (de différentes compétences musicales) qui ont considéré que chacun des portraits de ce sous-groupe avait un sens précis et d'autre part, le nombre de sujets parmi lesquels on pouvait trouver un ensemble d'éléments constituant de ces portraits parmi leurs verbalisations spontanées. Ainsi, par exemple, pour comparer au portrait "un morceau de métal mis en résonance", on a compté le nombre de sujets qui avaient spontanément produit dans l'expérience 1 des locutions contenant le mot "métal" et le mot "résonance". On s'attend donc à ce que la comparaison entre le nombre de sujets qui ont spontanément produit un certain nombre de composantes des portraits proposés, et le nombre de sujets qui les ont considérés comme précis produise un assez grand nombre de différences significatives.

Les résultats sont rassemblés en annexe, dans le tableau A7.II.1. Si on considère l'ensemble des sujets, 11 des 26 portraits testés présentent des différences significatives entre le nombre de sujets ayant produit spontanément un terme dans l'expérience 1 et le nombre de sujets ayant considéré que le portrait associé avait un sens précis dans l'expérience 3. Il s'agit ici des locutions : "du vent dans un tuyau", "on bouge une plaque métallique", "on casse du verre", "on frotte du verre", "on percute du métal", "on percute une peau", "on tape sur une tôle", "un son produit par une touche d'orgue électronique", "une corde enrobée avec un tissu", "une double percussion sur une peau" et "une pièce de monnaie sur des cordes". On peut remarquer que les différences observées sont toutes dues à des locutions produites spontanément par moins de 3 personnes différentes et pour lesquelles le portrait associé est pourtant considéré comme ayant un sens précis pour plus de 7 personnes différentes.

Deux autres locutions n'ont été produites spontanément que par 2 personnes différentes : il s'agit de "on frappe du verre" et "on effleure une corde avec du métal", les différences avec respectivement 8 et 7 sujets considérant le portrait associé comme ayant un sens précis, ne sont pas significatives. Toutes les locutions produites spontanément par 3 sujets ou plus ont abouti à des différences non significatives. Le nombre de différences significatives diminue si on les considère au sein de chacun des deux groupes de sujets de compétences musicales différentes (3 différences significatives chez les musiciens et 3 autres chez les non-musiciens). En général, toutes ces différences reprennent des portraits pour lesquels elles étaient significatives sur l'ensemble des sujets, sauf pour deux portraits : "une anche vibre" n'avait été produit spontanément que par des sujets musiciens. Ici 5 non-musiciens considèrent ce portrait comme ayant un sens précis. Au contraire "une corde vibre dans du verre" n'avait été produit que par des sujets non musiciens. Pourtant, 6 musiciens considèrent ce portrait comme ayant un sens précis. On remarque que toutes les différences observées en fonction de la compétence musicale sont dues au fait que la production spontanée est effectuée par peu de sujets d'un seul des deux groupes (musiciens ou non musiciens) alors que le jugement "sens précis" est réparti de façon homogène dans les deux groupes de sujets. En effet, parmi les productions spontanées, le pourcentage moyen de sujets musiciens qui ont produit chacun des portraits testés est de  $49 \pm 33$  %, alors que parmi les sujets qui ont considéré qu'un portrait avait un sens précis, il est de  $50 \pm 13$  %.

Pour ce qui est des portraits construits à partir de noms de **modes d'excitation**, si on considère l'ensemble des sujets, 5 des 12 portraits testés présentent des différences significatives entre le nombre de sujets ayant produit spontanément un terme dans l'expérience 1 et le nombre de sujets ayant considéré que le portrait associé avait un sens précis dans l'expérience 3 (voir tableau A7.II.2 en annexe). Il s'agit ici des locutions : "froissé", "glissé", "gratté", "raclé" et "effleuré". On peut remarquer que les différences observées sont toutes dues à des locutions produites spontanément par moins de 3 personnes différentes et pour lesquelles le portrait associé est pourtant considéré comme ayant un sens précis pour plus de 8 personnes différentes. On peut remarquer aussi que toutes les locutions spontanées sont au participe passé, donc produites sous forme d'adjectifs alors que les portraits proposés ont été construits avec les verbes associés (par exemple : "on froisse"). Par contre, si on considère les sujets selon leur compétence musicale, on observe beaucoup moins de différences significatives : 2 pour les sujets non musiciens et une seule pour les sujets musiciens. On note donc une plus grande homogénéité dans ces deux sous-groupes de sujets, entre le nombre de personnes d'une certaine compétence musicale qui produisent un terme, et le nombre de personnes de cette même compétence musicale qui considèrent le portrait associé comme ayant un sens précis. Ici aussi, à part pour "on effleure", ces différences sont dues au fait que ces portraits ont été produits spontanément par des sujets d'un seul des deux groupes (musiciens ou non-musiciens) alors que les sujets qui considèrent les portraits associés comme précis se répartissent de façon homogène dans ces deux groupes. Le pourcentage moyen de sujets musiciens qui ont produit chacun des portraits testés est de  $53 \pm 22 \%$ , alors que parmi les sujets qui ont considéré qu'un portrait avait un sens précis, il est de  $53 \pm 9 \%$ .

Par contre, pour ce qui est des portraits construits à partir de noms de **matériaux**, si on considère l'ensemble des sujets, seulement 3 des 17 portraits testés présentent des différences significatives entre le nombre de sujets ayant produit spontanément un terme dans l'expérience 1 et le nombre de sujets ayant considéré que le portrait associé avait un sens précis dans l'expérience 3 (voir tableau A7.II.3. en annexe). Il s'agit ici des locutions : "de la tôle", "une corde en boyau" et "de la peau". Ces locutions ont toutes été produites spontanément par moins de 3 personnes différentes et le portrait associé est pourtant considéré comme ayant un sens précis pour plus de 10 personnes différentes. Toutes les locutions produites spontanément par plus de 4 sujets mènent à des différences non-significatives. Seul le portrait : "de la peau et du métal" n'a été produit que par 2 sujets différents et mène à une différence non significative du fait que seulement 6 sujets le considèrent comme ayant un sens précis. Ici aussi, si on considère les sujets selon leur compétence musicale, on observe beaucoup moins de différences significatives : aucune pour les sujets musiciens et 1 pour les sujets non musiciens, concernant le portrait "une corde en boyau" qui n'avait été produit que par un sujet musicien. Le pourcentage moyen de sujets musiciens qui ont produit chacun des portraits testés est de  $53 \pm 20 \%$ , alors que parmi les sujets qui ont considéré qu'un portrait avait un sens précis, il est de  $49 \pm 13 \%$ . Ainsi, les différences dues à la compétence musicale sont beaucoup plus faibles dans l'expérience 3.

Pour ce qui est des portraits construits à partir de noms de **familles d'instruments**, si on considère l'ensemble des sujets, 1 seul des 6 portraits testés présente des différences significatives entre le nombre de sujets ayant produit spontanément un terme dans l'expérience 1 et le nombre de sujets ayant considéré que le portrait associé avait un sens précis dans l'expérience 3 (voir tableau A7.II.4. en annexe). Il s'agit ici des locutions : "famille des claviers". C'est la locution produite par le moins de sujets différents (2 contre au moins 6 pour les autres portraits de ce sous-groupe). Cette différence se maintient si on considère les sujets musiciens, ce qui est dû cette fois-ci à une différence dans le jugement de sens (plus de musiciens considèrent que ce portrait a un sens précis que de non-musiciens) alors que les 2 sujets qui ont produit spontanément cette locution étaient un musicien et un non musicien. On observe d'ailleurs le comportement inverse de celui observé ci-dessus en terme de différences de pourcentage moyen de sujets musiciens dans les deux conditions. En effet, le pourcentage moyen de musiciens qui ont produit chacun des portraits testés est de  $56 \pm 7 \%$ , alors que parmi les sujets qui ont considéré qu'un portrait avait un sens précis, il est de  $61 \pm 10 \%$ . Plus de musiciens considèrent en moyenne que les portraits ont un sens précis, alors qu'ils ont été produits par des sujets des deux groupes de compétences musicales. Par contre les différences entre musiciens et non musiciens sont plus grandes dans l'expérience 3.

Pour ce qui est des portraits construits à partir de noms d'**instruments**, si on considère l'ensemble des sujets, 22 des 36 portraits testés présentent des différences significatives entre le nombre de sujets ayant produit spontanément un terme dans l'expérience 1 et le nombre de sujets ayant considéré que le portrait associé avait un sens précis dans l'expérience 3 (voir tableau A7.II.4. en annexe). C'est la plus grande différence observée pour les portraits décrivant la source sonore. En fait, ces différences significatives concernent 22 des 24 noms d'instruments produits par 1 à 3 personnes différentes dans l'expérience 1. En effet, ces portraits ont été considérés comme ayant un sens précis par toujours plus de 8 sujets différents. Les deux cas différents sont "un harmonium", produit par deux sujets et ayant un sens précis pour 8 sujets, et "une cloche" produite par 3 sujets et ayant un sens précis pour 10 sujets. On conserve ici un grand nombre de différences significatives dans chacun des deux groupes de sujets (respectivement 11 et 12 chez les musiciens et les non musiciens) même si cela représente un pourcentage des portraits plus faible (respectivement 30 et 33%). La plupart de ces portraits ont été produits spontanément par un faible nombre de sujets différents (en moyenne 3.4). Le pourcentage moyen de sujets musiciens qui ont produit chacun des portraits testés est de  $52 \pm 39 \%$ , alors que parmi les sujets qui ont considéré qu'un portrait avait un sens précis, il est de  $52 \pm 11 \%$ . Ainsi, les différences dues à la compétence musicale sont beaucoup plus faibles dans l'expérience 3.

Nous n'avons pas testé les différences concernant les portraits construits à partir d'**instruments visés par le concepteur**, vu que ces noms ont été déduits de ceux donnés à ces sons par leur concepteur (Wessel et al., 1987), et ce, bien qu'un certain nombre d'entre eux ont été produits spontanément. Dans ce dernier cas, ils font partis du sous-groupe précédent.

Nous allons donc aborder les portraits construits à partir d'adjectifs décrivant les qualités des sons.



Pour ce qui est des portraits construits à partir d'adjectifs de **qualités simples** utilisés spontanément pour **comparer** les sons 2 à 2, les différences entre le nombre de personnes les ayant produits spontanément et le nombre de personnes ayant considéré l'échelle sémantique associée comme globalement adaptée ont déjà été testées dans le paragraphe précédent (§ I.1.3). Ici, si on considère l'ensemble des sujets, seulement 3 des 21 portraits testés présentent des différences significatives entre le nombre de sujets ayant produit spontanément un terme dans l'expérience 1 et le nombre de sujets ayant considéré que le portrait associé avait un sens précis dans l'expérience 3 (voir tableau A7.II.6. en annexe). Cela concerne les deux adjectifs produits spontanément par moins de 4 sujets différents ("diffus" et "résonnant"), mais aussi l'adjectif "aigu", produit spontanément par 19 sujets et considéré comme ayant un sens précis par seulement 8 sujets différents. On retrouve donc ici cette différence entre la production spontanée de cet adjectif et le jugement de pertinence ou de sens émis par les sujets dans un autre contexte. Deux différences significatives sont observées dans chacun des deux groupes de sujets : on retrouve "diffus" et on ajoute "soufflé" chez les musiciens, on retrouve "résonnant" et on ajoute "large" chez les non musiciens. Cela concerne deux types de comportements différents : pour "diffus" et "résonnant", dans l'expérience 3, les sujets ayant jugé du sens précis de ce portrait se répartissent de manière homogène entre musiciens et non musiciens alors que ces adjectifs ont été plus souvent produits par des sujets d'une compétence musicale donnée, ce qu'on a déjà observé précédemment. Par contre, on observe l'effet inverse pour "large" et "soufflé" : ces adjectifs ont été spontanément produits de façon homogène ou plus souvent par des sujets musiciens, par contre, les non-musiciens sont beaucoup plus nombreux à considérer que le sens des portraits associés est précis. D'ailleurs, si on compare les pourcentages moyens de sujets musiciens dans les deux contextes, on remarque  $52 \pm 11$  % de musiciens qui ont produits chacun des portraits testés contre  $39 \pm 13$  % parmi les sujets qui ont considéré qu'un portrait avait un sens précis. Les différences sont équivalentes dans les deux contextes. Par contre, il apparaît que plus de non musiciens que de musiciens considèrent les portraits de ce groupe comme précis.

Pour ce qui est des portraits construits à partir d'adjectifs de **qualités simples** utilisés spontanément pour **catégoriser** les sons, si on considère l'ensemble des sujets, seulement 3 des 16 portraits testés (soit 19%) présentent des différences significatives entre le nombre de sujets ayant produit spontanément un terme dans l'expérience 1 et le nombre de sujets ayant considéré que le portrait associé avait un sens précis dans l'expérience 3 (voir tableau A7.II.7. en annexe). Cela concerne ici aussi les adjectifs produits spontanément par le moins de sujets différents : "franc" et "terminé" (respectivement produits par 5 et 3 sujets différents) mais aussi "chaleureux" produit pourtant par 7 sujets différents. Les différences vont donc toujours dans le même sens : globalement plus de sujets considèrent les portraits comme ayant un sens précis que ne pouvait en laisser prévoir le nombre de personnes qui produisent spontanément ces adjectifs. On retrouve le fait que "chaleureux" a été produit par plus de non musiciens alors que les sujets des deux groupes le considèrent comme ayant un sens précis. Chez les non-musiciens, on retrouve "terminé" (produit uniquement par des musiciens), ainsi que "double" produit essentiellement par des musiciens. On retrouve donc ici l'idée

que dans l'expérience 3, les différences de comportement liées à la compétence musicale sont moins fortes que dans l'expérience 1. D'ailleurs, on retrouve  $54 \pm 20$  % de musiciens qui ont produit chacun des portraits testés contre  $47 \pm 9$  % parmi les sujets qui ont considéré qu'un portrait avait un sens précis. Soit à la fois l'idée que plus de non musiciens considèrent les portraits comme ayant un sens précis dans l'expérience 3 et que les différences y sont moins marquées.

Pour ce qui est des portraits construits à partir de **deux** et a fortiori de **trois adjectifs**, on trouve très peu de locutions spontanées mettant côte à côte ces 2 adjectifs (au moins une fois mais en général seulement une fois) et jamais de locutions spontanées mettant côte à côte ces 3 adjectifs, puisque ces portraits ont été construits à partir des échelles sémantiques de l'expérience 2 (voir chapitre 4 § III.1.). Les comparaisons ont donc été effectuées, un peu comme dans le cas des sources sonores, en recueillant parmi les productions spontanées le nombre de sujets différents qui ont produit au moins une fois chacun des adjectifs contenus dans les portraits doubles ou triples. (Par exemple pour le portrait : "nasillard, peu pincé et peu doux", on a comparé le nombre de sujets qui ont considéré ce portrait comme ayant un sens précis au nombre de sujets qui avaient produit spontanément au cours de leurs verbalisations les adjectifs "nasillard", "pincé" et "doux" au moins une fois chacun.).

Ceci étant précisé, pour ce qui est des portraits construits à partir de **deux adjectifs**, si on considère l'ensemble des sujets, 5 des 21 portraits testés présentent des différences significatives entre le nombre de sujets ayant produit spontanément un terme dans l'expérience 1 et le nombre de sujets ayant considéré que le portrait associé avait un sens précis dans l'expérience 3 (voir tableau A7.II.8. en annexe). Cela concerne ici aussi les adjectifs produits spontanément par moins de 3 sujets différents : "court et ramassé", "mou et agressif", "pincé et résonnant" et "sourd et flasque" mais aussi "froid et agréable" produits pourtant par 4 sujets différents, mais jugé précis par 14 sujets différents. On retrouve toutes les différences significatives dues à un faible nombre de sujets parmi les musiciens, toutes dues à une plus grande homogénéité entre musiciens et non musiciens dans le jugement de l'expérience 3 que lors de la production spontanée. Seules deux différences sont significatives parmi les non musiciens et explicables de la même façon. D'ailleurs, on retrouve  $40 \pm 24$  % des musiciens qui ont produit chacun des portraits testés contre  $43 \pm 10$  % parmi les sujets qui ont considéré qu'un portrait avait un sens précis. Soit l'idée que les différences sont moins marquées dans l'expérience 3.

Par contre, pour ce qui est des portraits construits à partir de **trois adjectifs**, si on considère l'ensemble des sujets, 5 des 12 portraits testés présentent des différences significatives entre le nombre de sujets ayant produit spontanément un terme dans l'expérience 1 et le nombre de sujets ayant considéré que le portrait associé avait un sens précis dans l'expérience 3 (voir tableau A7.II.9. en annexe). C'est la plus grande différence observée pour les portraits décrivant des qualités des sons. On trouve en général peu de sujets qui ont spontanément produits les 3 adjectifs de chacun des portraits considérés (en moyenne 5.2). Pourtant, cela ne concerne pas de manière systématique les adjectifs ayant été produits spontanément par le moins de sujets différents. On retrouve 4 différences

significatives chez les musiciens contre seulement 2 chez les non-musiciens. La majorité de ces différences sont dues à une plus grande homogénéité entre musiciens et non-musiciens dans l'expérience 3 que dans l'expérience 1, à l'exception du portrait "long, diffus et attaque lente" dont les 3 composants ont été produits par 2 sujets, un musicien et un non-musicien, mais qui a été considéré comme ayant un sens précis par plus de musiciens que de non-musiciens. On retrouve, ici aussi  $52 \pm 35$  % de musiciens qui ont produit chacun des portraits testés contre  $53 \pm 9$  % parmi les sujets qui ont considéré qu'un portrait avait un sens précis ; d'où l'idée que les différences sont beaucoup moins marquées dans l'expérience 3.

Il est difficile de comparer le nombre d'occurrences de chacun des termes lors des verbalisations libres et les jugements de sens des portraits. En effet, les jugements de sens sont beaucoup plus homogènes entre eux que les nombres d'occurrences lors des verbalisations spontanées. Pour donner un exemple, on enregistre 266 occurrences spontanées de "pincé/pincement/pincer" contre seulement 1 de "froissé" et 2 de "effleuré". Pourtant, le nombre de sujets ayant considéré les portraits associés (respectivement "on pince", "on froisse" et "on effleure") comme ayant un sens précis est respectivement de 9, 8 et 6. Nous n'avons donc finalement pas comparé ces données.

### II.3. Résumé

Les jugements sur le sens des portraits présentent des différences en fonction d'une part, du type de portrait proposé, et d'autre part de la compétence musicale des auditeurs. Les sujets considèrent que 54% de l'ensemble des portraits proposés ont un sens précis, 36% un sens vague et seulement 10% n'ont pas de sens pour eux. Plus de portraits sont jugés n'ayant "pas de sens" s'ils décrivent la source sonore plutôt que le son (respectivement 12% contre 7%). En général, la répartition des réponses sur le sens des portraits ne dépend pas du sous-groupe de portraits considéré ni des portraits rassemblés dans chacun des sous-groupes, sauf dans le cas des 6 portraits de noms de familles d'instruments (un plus grand nombre de portraits de ce sous-groupe est considéré comme ayant "un sens précis" et un moins grand nombre comme n'ayant "pas de sens" que pour l'ensemble des autres portraits décrivant une source sonore possible). De plus, il y a un lien significatif entre le jugement de sens et un des 6 portraits de ce sous-groupe. En effet, on observe plus de réponses "pas de sens" liées au portrait : *un instrument de la famille des percussions* (10% des réponses) et plus de réponses "un sens vague" liées au portrait : *un instrument de la famille des bois* (50% des réponses).

Il semble que ces différences reflètent des comportements différents entre les musiciens et les non-musiciens : les musiciens ont tendance à considérer que plus de portraits décrivant la source ont "un sens précis" (respectivement 56% contre 49% ), et plus de portraits décrivant le son n'ont "pas de sens" (respectivement 9% contre 5%) que pour les non-musiciens . Ceci peut s'expliquer par le fait que les adjectifs décrivant les sons sont les moins spécifiques du domaine sonore que les noms qui constituent les portraits de sources sonores. En effet, ces différences de jugement sont d'autant plus importantes lorsqu'il s'agit pour les portraits décrivant le son des adjectifs simples utilisés dans

l'expérience 2 (par exemple : *aigu*) et des adjectifs associés deux à deux (par exemple : *doux et froid*), et pour les portraits décrivant la source des noms d'"instruments" électroniques et des noms de familles d'instruments (dont 83% sont jugés précis par les musiciens contre 58% pour les non-musiciens).

Pour ce qui est de la comparaison entre les jugements de sens dans l'expérience 3 et les jugements de pertinences des échelles sémantiques associées dans l'expérience 2, les sujets ont en général jugé du sens précis des portraits du type "un son brillant" de la même façon que de la pertinence des échelles associées "peu brillant - très brillant". Les jugements "sens vagues" sont difficiles à comparer, quand aux jugements "pas de sens", les sujets ont plus facilement jugé qu'une échelle était "pas adaptée" plutôt que le portrait associé n'avait pas de sens.

Pour ce qui est de la comparaison avec les productions spontanées, les différences observées (entre 14 et 61% des portraits) vont très souvent dans le même sens : globalement on trouve plus de sujets qui considèrent les portraits comme ayant un sens précis que ne pouvait le laisser prévoir le nombre de personnes qui produisent spontanément ces adjectifs, ce qui est d'autant plus marqué quand le nombre de locuteurs spontanés des expressions testées est inférieur à 3. De plus, on observe en général une plus grande homogénéité entre les compétences musicales des sujets lors du jugement de sens que lors de la production spontanée.

### **III. Conclusion**

Un résultat important découvert dans ce chapitre est qu'on utilise, pour décrire spontanément les sons qu'on entend, des mots qu'on peut pourtant considérer comme "pas adaptés". En effet, il est extrêmement inattendu d'observer qu'en moyenne, chacune des échelles proposées, pourtant extraites des verbalisations libres les plus fréquentes (produites avec le nombre d'occurrences le plus élevé), n'est jugée pertinente que par 62% des sujets (dont 64% des anciens sujets qui ont pourtant produit spontanément ces termes).

De plus, si on considère uniquement les sujets ayant participé aux deux expériences, on peut remarquer que leurs jugements de pertinence des échelles proposées ne coïncident pas avec leur propre utilisation spontanée de ce même vocabulaire pour décrire ces mêmes sons. En effet, si on trouve des mots à la fois utilisés spontanément et considérés comme pertinents par un grand nombre de sujets, on observe aussi des échelles considérées comme pertinentes par moins de 50% des sujets bien que plus de 70% d'entre eux les ont utilisées spontanément. A l'inverse, pour beaucoup de sujets, certains termes peu employés spontanément peuvent paraître au contraire globalement adaptés.

Le vocabulaire utilisé spontanément pour décrire des sons n'est donc pas forcément considéré comme pertinent, même par leur locuteur. Pourtant, ayant recueilli d'éventuels commentaires écrits par les sujets sur leur utilisation des échelles au terme de notre deuxième expérience, on peut remarquer que plus une échelle est globalement considérée comme non-pertinente par un grand nombre de sujets,

plus ceux-ci ajoutent des commentaires pour nuancer leur propos, essayant de définir une acception ou un contexte qui permettrait d'utiliser quand même ce mot, une sorte de mode d'emploi possible dans ce contexte. Ainsi par exemple, l'échelle liée à *fort*, considérée globalement comme pertinente par le moins de sujets (23%), suscite de nombreux commentaires du genre : "fort dans la présence sonore, l'attaque", ou encore "évoque la puissance et la persistance du son".

Ces conclusions forment déjà une première étape pour repousser une conception référentielle du langage, qu'on peut schématiser par une bijection "mots  $\Leftrightarrow$  chose", pour lui préférer une conception plus pragmatique, en proposant au contraire l'idée que le sens des mots est adapté au contexte : les adjectifs "fort" ou "aigu", pourtant spécifiques du domaine de l'audition, n'ont pas été utilisés ici dans leur sens "prototypique" liés respectivement aux notions d'intensité et de hauteur tonale des sons. En effet, les sons présentés aux sujets étaient égalisés selon ces deux dimensions. Ils ont pourtant été utilisés, d'une manière que les sujets ont essayé de préciser dans leurs commentaires.

Il reste à savoir si ce vocabulaire permet de reconnaître un son, sinon la description des sons resterait totalement inutile pour permettre leur évocation. Ce point est analysé dans le chapitre suivant .

## **Chapitre 8. Données d'association de sons aux portraits**

Nous allons analyser, dans ce dernier chapitre de la partie expérimentale, la façon dont les participants ont associé un ou plusieurs des 12 sons de cette étude à chacun des portraits verbaux, construits à partir des verbalisations libres récoltées dans la première expérience et proposés dans l'expérience 3. La méthodologie de cette expérience est décrite dans le chapitre 4, §III.

Les participants devaient associer au moins un son (mais pouvaient en associer plusieurs) à chacun des portraits verbaux proposés. La définition d'une mesure de la cohérence de choix entre les participants, habituellement effectuée avec des corrélations inter-sujets, nous a demandé ici un peu plus de réflexion (paragraphe I). Ensuite, nous nous sommes intéressés plus particulièrement aux sons choisis par le plus grand nombre de participants. Nous voulions voir si les différents types de portraits menaient à des utilisations différentes, en fonction des compétences des participants, mais aussi du nombre de timbres différents qui y étaient associés, certains portraits apparaissant comme spécifiques d'un ou deux sons ou au contraire associés à un grand nombre de sons différents. Enfin, nous avons comparé les sons choisis lors de l'association aux portraits verbaux aux sons qui avaient été à l'origine de ces descriptions lors des verbalisations libres, celles récoltées lors de notre première expérience mais aussi celles récoltées au début de l'expérience 3 (voir chapitre 6 pour une description de ces verbalisations). Nous pouvons alors analyser dans quelle mesure on peut reconnaître un son à partir d'une description verbale et quel type de description est le plus efficace dans la communication.

De ces résultats dépend l'intérêt de continuer à étudier la façon dont on parle des sons. En effet, s'il est de toute façon impossible de comprendre de quel son quelqu'un a parlé, il paraît vain de tenter de décrire les sons.

## **I. Cohérence des associations entre les participants**

Nous avons développé une mesure de la cohérence des choix entre les différents participants, afin de savoir s'ils associent les mêmes sons aux mêmes portraits. Or, comme chaque sujet pouvait choisir un ou plusieurs sons, nous avons dû mettre en place deux mesures de cohérence, que nous avons appelées pourcentage de Réponses Communes (%RC) et pourcentage de Choix Communs (%CC), qui sont définis comme des moyennes, sur les 190 paires de participants (20 dans chaque condition). Le pourcentage de réponses communes de deux sujets, pour un portrait considéré, est le nombre de sons communs que ces 2 personnes ont choisis d'associer au portrait plus celui de sons communs qu'ils ont choisis de NE PAS associer au portrait. Cette somme est ensuite divisée par le nombre total de sons possibles.

Cette mesure considère donc de manière analogue les choix d'associer ET de ne pas associer un son donné. Ainsi, elle ne dépend pas du nombre de sons choisis. Par contre, elle a tendance à donner de forts pourcentages. En effet, si deux sujets A et B choisissent respectivement les sons  $T_i$  et  $T_j$  pour un portrait, ils présentent 83%RC (ils ont tous les deux rejeté les 10 autres timbres). Le pourcentage de Choix Communs de deux sujets, pour un portrait considéré, est le nombre de sons choisis en commun par les deux personnes divisé par le nombre de choix total des deux personnes.

Cette mesure ne considère que les sons associés à chaque portrait et non pas ceux rejetés. (Les sujets A et B dans notre exemple précédent auraient 0% de CC). Elle dépend du nombre de sons choisis pour chaque portrait et a donc tendance à avantager les portraits qui induisent plus de sons choisis par rapport aux portraits plus spécifiques d'un seul son. Chacune de ces deux mesures a ses avantages et ses inconvénients, nous garderons donc les deux types de mesure de cohérence inter-sujets.

Sur l'ensemble des 201 portraits proposés, on obtient en moyenne 78.4 % de Réponses Communes et 39.5% de Choix Communs. Nous analyserons d'éventuelles différences inter-groupes de sujets pour chacun de ces résultats.

### I.1. Cohérence d'association en fonction des types de portraits (%RC et %CC)

Globalement, si on rassemble d'une part, les 91 portraits qui décrivent le son (ceux formés à partir d'adjectifs de qualité) et d'autre part les 110 portraits qui décrivent une source sonore possible de ces sons (sous forme de noms d'instruments ou de modes d'excitation et matériaux), on obtient des différences de comportement entre les deux groupes. En effet, le groupe de portraits décrivant la source sonore obtient un pourcentage moyen de Réponses Communes (%RC) plus important que celui décrivant le son (respectivement 83.8 et 71.8%RC,  $t(199)=11$ ,  $p<.0001$ ). Par contre la différence en terme de pourcentage de choix communs (%CC) n'est pas significative ( $t(199)=1.1$ ;  $p=.27$ ), la moyenne obtenue sur le groupe de portraits décrivant le son étant de 38.3 contre 40.5 pour

les portraits décrivant la source. Le tableau 8.1 donne les %RC et %CC de chaque sous-groupe de portraits.

Tableau 8.1. Répartition des %RC et %CC moyen de chaque sous-catégorie de portraits

Type de Portrait	NB de portraits	%RC moyenne	Ecart-type	%CC moyenne	Ecart-type
Total SON	91	71.81		38.26	
Qualité simple positive	21	69.81	4.62	41.05	8.89
Qualité simple négative	21	61.81	4.57	43.90	8.73
Autres Qualités simples	16	72.38	5.61	37.44	10.24
Qualités doubles	21	78.05	4.57	31.10	11.29
Qualités triples	12	81.08	3.96	37.17	10.35
Total SOURCE	110	83.82		40.49	
Actions	12	81.08	3.58	42.50	18.12
Matériaux	17	77.47	8.87	36.59	12.52
Sources	26	82.73	6.63	39.19	15.78
Familles d'Instruments	6	80.33	4.97	56.17	8.01
Noms d'instruments	37	87.00	6.10	39.49	18.64
Noms de fichiers	12	89.75	3.19	42.08	17.83
TOUS	201	78.37		39.48	

Si on compare les moyennes obtenues par chaque sous-groupe à l'intérieur de son groupe de portraits, dans le groupe décrivant le son, les différences de moyenne de %CC ne sont significatives par rapport à la moyenne du groupe que pour les sous-groupes des adjectifs doubles, qui présentent un plus faible %CC, et des adjectifs simples négatifs, qui présentent un plus fort %CC (respectivement  $t(20)=2.91$  et  $2.96$ ,  $p<.01$ ). Par contre, pour le %RC, 3 sous-groupes présentent des différences de moyenne significatives par rapport à l'ensemble des portraits décrivant le son. Les portraits construits à partir d'adjectifs triples, de même que ceux construits à partir d'adjectifs doubles, ont un plus fort %RC que la moyenne, et ceux construits à partir d'adjectifs négatifs ont un plus faible %RC que la moyenne (respectivement  $t(11)=8.12$ ,  $t(20)=6.28$  et  $t(20)=10.01$ ,  $p<.0001$ ). Dans le groupe de portraits décrivant une source sonore possible, seul le %CC moyen du sous-groupe de portraits de noms de famille d'instruments est significativement plus fort que l'ensemble du groupe ( $t(5)=4.79$ ,  $p<.01$ ). Par contre, pour le %RC, les noms de fichiers et les noms d'instruments présentent un plus fort %RC alors que les noms de matériaux et de modes d'excitation présentent un plus faible %RC (respectivement  $t=6.44$ ,  $3.18$ ,  $2.94$  et  $2.63$ ;  $p<.0001$ ,  $p=.003$ ,  $.009$  et  $.02$ ;  $df=11$ ,  $36$ ,  $16$  et  $11$ ).

Nous avons rassemblé en annexe (tableaux A.8.1 à A.8.11) toutes les valeurs de %RC et %CC calculées pour chaque portrait de chaque sous-groupe, en fonction de la compétence musicale des participants. On peut analyser les résultats de la manière suivante.

Dans le groupe des portraits décrivant des qualités sonores des sons à l'aide d'adjectifs, pour le sous-groupe des **adjectifs simples positifs** (voir tableau A.8.1), la moyenne des %RC=70 environ. 14 des 21 portraits de ce sous-groupe présentent des différences de moyennes de %RC significatives. En particulier, les portraits construits à partir des adjectifs *aigu*, *fort*, *attaque rapide* et *soufflé* ont des



%RC supérieurs à la moyenne, et ceux construits à partir de *riche* et *vibré* ont des %RC inférieurs à la moyenne. Ce qui concerne les adjectifs *aigu* et *fort* est assez étonnant compte tenu que ces deux adjectifs ont été considérés dans l'expérience 2 comme n'étant globalement pas adaptés pour décrire ces sons. On retrouve le fait que des portraits peuvent à la fois être jugés non pertinents et être utilisés de la même façon par les participants. En ce qui concerne les %CC, en moyenne égal à 41%, on obtient des différences significatives avec cette moyenne pour 14 des 21 portraits de ce sous-groupe. En particulier, on retrouve les portraits construits avec *attaquerapide* et *soufflé* comme ayant un plus fort %CC que la moyenne, ceux construits à partir de *large*, *vibré* et *doux* ayant un plus faible %CC.

Pour le sous-groupe des **adjectifs simples négatifs** (voir tableau A.8.2), 13 portraits sur 21 (soit 62%) présentent des différences significatives de %RC par rapport à la moyenne (de 62%RC environ). Il s'agit principalement de *attaque lente*, *pas soufflé*, *pas net* et *pas résonnant*, qui ont des %RC plus forts que la moyenne et de *pas large* et *pas brillant*, qui ont des %RC inférieurs à la moyenne. On trouve par contre 16 portraits sur 21 qui présentent des différences significatives en terme de %CC par rapport à la moyenne du sous-groupe (environ 44%). Il s'agit principalement de *pas soufflé*, *pas pincé* et *pas doux*, qui ont des %CC plus forts que la moyenne et de *pas aigu*, *pas riche*, *pas grave*, *pas fort* et *pas large*, qui ont des %CC plus faibles que la moyenne.

Pour l'**autre** sous-groupe d'**adjectifs simples** (extraits des verbalisations libres les plus fréquentes dans l'expérience 1 pour caractériser les sons) (voir tableau A.8.3), la majorité des portraits (11 sur 16) présentent des %RC significativement différents de la moyenne du sous-groupe, égale à 72%. Il s'agit principalement de *cuivré*, *pur*, *agressif*, *agréable* et *chaleureux*, qui présentent un %RC plus fort que la moyenne, et *terminé* et *intérieur* qui présentent un %RC inférieur à la moyenne. Ces résultats semblent en rapport avec ceux sur le sens de ces portraits (cf. chapitre 7, §II). Le nombre de portraits présentant des différences de %CC par rapport à la moyenne du sous-groupe est encore plus important (14 sur 16), ce qui est dû à un fort écart type (10.2). On retrouve principalement *constant*, *pur* et *cuivré* parmi les %CC les plus forts et *double*, *intérieur*, *fin*, *terminé* et *chaleureux* parmi les %CC les plus faibles.

Pour le sous-groupe des **adjectifs doubles** (voir tableau A.8.4), on trouve 13 portraits sur 21 dont les %RC sont significativement différents de la moyenne du sous-groupe. Il s'agit principalement de *rond et soufflé*, *froid et agréable*, *doux et froid* et *clair et rond*, qui présentent de plus forts %RC que la moyenne, et de *ni mou ni agressif*, *rapide et précis*, *court et ramassé*, qui présentent des %RC plus faibles que la moyenne. On peut remarquer que le portrait présentant le plus faible %RC est le seul construit avec la double négation *ni... ni*, et que parmi ceux qui ont les plus fort %RC, on retrouve deux portraits construits à partir de l'adjectif *froid* qui "visait" le son VBR. En ce qui concerne les %CC, on obtient de forts écarts par rapport à la moyenne (ET=11.2, 17 différences significatives sur 21 portraits). Il s'agit principalement de *aigu et métallique* et *pur et net*, qui présentent de plus forts %CC que la moyenne et de *court et ramassé* (12%CC), *lourd et continu*, *sourd et métallique*, qui présentent les plus faibles %CC.

Enfin, pour le sous-groupe des portraits construits à partir de **trois adjectifs** de l'expérience 2 et visant chacun un des 12 timbres de notre ensemble (voir tableau A.8.5), on obtient, par rapport au %RC moyen du sous-groupe, égal à 81%, 9 différences significatives sur les 12 portraits, dont 6 différences aboutissant à des %RC plus forts que la moyenne (il s'agit principalement des portraits visant les sons CLA, CVC, TPT et VBR, dont parmi les plus vraisemblables). Trois différences aboutissent à des %RC plus faibles que la moyenne. Il s'agit des portraits visant les sons de SNO, TBN et VBN. On peut noter que deux de ces sons sont des hybrides, que le portrait visant le SNO est le seul construit à partir de *assez... assez*, et que les deux derniers portraits (visant le TBN et le VBN) contiennent les mêmes 2 premières composantes : *sourd et bas*. En terme de %CC, 8 des différences sont significatives (sur 12). On retrouve parmi les portraits présentant les plus fort %CC ceux visant les sons VBR, CVC et PNE, et parmi ceux ayant le plus faible %CC ceux visant le TBN, mais aussi le VBN et le SNO.

Dans le groupe des portraits décrivant des sources sonores possibles des sons, pour le sous-groupe des **modes d'excitation** (voir tableau A.8.6), 7 différences par rapport à la moyenne du %RC du sous-groupe sont significatives (sur 12 portraits). Il s'agit principalement du portrait *on frotte*, mais aussi dans une moindre mesure *on racle* et *on effleure*, qui présentent les plus forts %RC, et d'autre part des portraits *on frappe*, *on glisse* et *on tire*, qui présentent les plus faibles %RC. Le %CC moyen de ce sous-groupe présentant un énorme écart type (18.1), la plupart des différences par rapport à la moyenne sont significatives (pour 11 des 12 portraits). On retrouve en particulier les portraits *on frotte*, *on percute*, *on pince*, parmi ceux qui ont les plus forts %CC et *on froisse* et *on tire*, parmi ceux qui ont les plus faibles %CC.

Dans le sous-groupe des noms de **matériaux** (voir tableau A.8.7), pour 15 des 17 portraits les %RC sont significativement différents de celui du groupe. En particulier, on obtient de très forts %RC pour le portrait *du verre* (93%), mais aussi les portraits *de la peau*, *du cristal*, *de l'eau*... On peut remarquer que la plupart de ceux-ci visaient le son VBR. Les portraits menant aux plus faibles %RC sont les plus généraux, construits comme *un matériau naturel*, *synthétique*, *électronique*, mais aussi *du métal*. On retrouve ces résultats en terme de %CC. En effet, 10 des 17 portraits présentent des différences significatives, et on retrouve parmi ceux ayant les plus forts %CC *du verre* (71%), mais aussi *du cristal* et *de la peau*, et parmi les plus faibles %CC *un matériau synthétique*, *naturel*, mais aussi *de la tôle* et *une corde en boyau*.

Dans le sous-groupe des noms de **source**, construits à partir de modes d'excitation et de matériaux (voir tableau A8.I.8), on obtient des différences significatives en terme de %RC pour 18 des 26 portraits. En particulier, *on frappe du verre*, *on souffle dans du bois* et *on percute une peau* présentent les plus fort %RC, *on étouffe le son* et *un son produit par une touche d'orgue électronique* ayant les plus faibles %RC. En terme de %CC, on retrouve aussi 18 différences significatives, avec *on frappe*

*du verre et on souffle dans du cuivre* parmi les plus forts %CC, *une anche vibre et une pièce de monnaie sur des cordes* parmi les plus faibles %CC.

Dans le sous-groupe des noms de **familles d'instruments** (voir tableau A.8.9), 4 des 6 portraits mènent à des différences de %RC significatives par rapport à la moyenne du sous-groupe. Il s'agit des *familles des cuivres, des bois et des percussions*, qui mènent à de plus forts %RC, et de la *famille des cordes* qui mène à un plus faible %RC. On retrouve ces différences en terme de %CC : 4 différences sont significatives, il s'agit cette fois-ci des *familles des cuivres et des vents* qui présentent de plus fort %CC et des *familles des percussions et des claviers* qui présentent de plus faibles %CC.

Dans le sous-groupe des noms **d'instruments de musique** (voir tableau A.8.10), qui possède un %RC moyen très élevé (87%), 30 des 37 portraits présentent des %RC significativement différents de cette moyenne. Il s'agit en particulier pour ceux qui présentent les plus forts %RC du portrait *xylophone* (99%), mais aussi de ceux de *violoncelle, flûte, violon, cloche* (pour lesquels %RC>93) et pour ceux qui présentent les plus faibles %RC des portraits *synthétiseur* (59%) et *orgue*, qui sont aussi les moins spécifiques d'un des 12 sons du groupe. On retrouve ces résultats en terme de %CC : 27 des 37 portraits présentent des différences significatives avec la moyenne, en particulier, les portraits de *xylophone* (97%), *violon* et *violoncelle* présentent les plus forts %CC, ceux de *synthétiseur* et d'*harmonica* présentent les plus faibles %CC (< 22%).

Enfin, pour le sous-groupe des **instruments visés par le concepteur** (voir tableau A.8.11), on retrouve de très forts %RC (en moyenne 98.8) avec 9 des 12 portraits qui présentent des différences significatives par rapport à cette moyenne. Il s'agit en particulier des portraits *trompette, clarinette, vibraphone* et *corde frottée*, qui présentent les plus forts %RC et d'autre part des portraits de *mélange hautbois et clavecin, piano électrique, mélange de corde frottée et de piano, mélange de vibraphone et de trombone* et *cor anglais*, qui présentent les plus faibles %RC. On peut remarquer qu'on trouve dans ce groupe les trois portraits de mélange, qui correspondent aux trois sons hybrides, ainsi que les deux portraits utilisant des noms plus spécifiques, comme cor anglais et piano électrique. On retrouve ces résultats en terme de %CC : beaucoup de portraits présentant une différence significative de %CC par rapport à la moyenne du sous-groupe (10 sur 12), notamment *clarinette, trompette* et *trombone* parmi les plus forts %CC, on retrouve les mêmes portraits présentant les plus faibles %CC que pour les %RC.

## I.2. Cohérence d'association en fonction de la compétence des sujets (%RC et %CC)

Globalement, sur l'ensemble des 201 portraits de l'expérience, les différences de %RC en fonction de la compétence des sujets ne sont pas significatives ( $t(400)=1.23$ ,  $p=.22$ ). Par contre, les différences de %CC sont significatives, les musiciens ayant globalement un plus fort pourcentage de Choix Commun entre eux que les non-musiciens (respectivement 43 et 37%,  $t(400)=3.67$ ,  $p<.001$ ). Les différences entre les %RC des musiciens et non-musiciens ne sont pas non plus significatives si on

considère les deux groupes de portraits décrivant respectivement la source et le son, de même que celles entre les %CC du groupe décrivant le son ( $t < 1.6$ ,  $p > .09$ , dl entre 180 et 218). Par contre, le %CC concernant l'ensemble des portraits décrivant une source sonore possible est plus important dans le groupe des musiciens que chez les non musiciens (respectivement %CC=45.7 et 36.7,  $t(218)=3.50$ ,  $p < .001$ ).

En fait, sur l'ensemble des sous-groupes de portraits décrivant le son, seul le sous-groupe construit à partir d'adjectifs triples présente des différences significatives de %RC en fonction de la compétence des participants. En effet, les musiciens présentent un plus fort %RC que les non musiciens (%RC=83.00 et 78.91,  $t(22)=2.11$ ,  $p < .05$ ). Par contre, c'est le %CC concernant le sous-groupe des adjectifs négatifs qui donne la seule différence significative en fonction de la compétence des sujets, les musiciens ayant ici aussi un plus fort %CC que les non musiciens (%CC=48.8 et 40.0,  $t(40)=2.69$ ,  $p < .01$ ).

Sur l'ensemble des sous-groupes de portraits décrivant une source sonore possible, seul le sous-groupe des instruments visés par le concepteur présente une différence significative, les musiciens ayant une fois encore un %RC plus fort que les non musiciens (%RC=92.67 et 78.83,  $t(22)=3.16$ ,  $p < .01$ ). Par contre, on observe plus de différences significatives de %CC entre musiciens et non musiciens. En effet, si ces différences sont déjà significatives au niveau de l'ensemble des portraits décrivant la source sonore, on retrouve une différence particulièrement importante de pourcentage de choix communs pour les portraits d'instruments visés par le concepteur, où les musiciens présentent un %CC presque deux fois plus grand que les non musiciens (respectivement %CC=60.08 et 32.33,  $t(22)=3.34$ ,  $p = .003$ ). Mais cette différence est aussi significative pour les portraits de noms de familles d'instruments et d'instruments (respectivement  $t=3.16$  et 2.14,  $p < .01$  et  $< .05$ , dl=10 et 72), les musiciens ayant une fois encore des %CC plus forts que les non musiciens (respectivement %CC=65.50 et 47.66 pour les portraits de familles d'instruments, %CC= 45.81 et 35.24 pour les portraits de noms d'instruments). Pour ce qui est d'éventuelles différences de comportement entre les musiciens et les non musiciens pour un des portraits de chaque sous-groupe, nous avons comparé les moyennes de %RC et de %CC obtenues à l'intérieur de chacun des deux sous-groupes de sujets (voir annexes A.8.1 à A.8.11).

Dans le groupe des portraits décrivant les qualités des sons, pour le sous-groupe des **adjectifs simples positifs** (voir tableau A8.I.1), on n'observe globalement pas de différence significative de comportement entre musiciens et non musiciens en ce qui concerne le %RC. Pourtant, 7 des 21 portraits présentent des différences significatives. Quatre de ces différences mènent à un plus fort %RC chez les musiciens que chez les non musiciens. Il s'agit des portraits *attaque rapide*, *vibré*, mais aussi *sourd* et *long* dans une moindre mesure. Trois différences mènent à un plus fort %RC pour les non musiciens que pour les musiciens. Il s'agit des portraits *riche*, *pincé* et *rond*. Par contre, en terme de %CC, les différences de comportement entre musiciens et non musiciens sont significatives au niveau de la moyenne correspondant à l'ensemble des portraits. On observe en effet de plus forts

%CC chez les musiciens que chez les non musiciens. On retrouve des différences significatives pour 6 des 21 portraits. Quatre portraits mènent à un plus fort %CC chez les musiciens, ceux liés à *attaque rapide, sec, long et brillant*, 2 mènent à un plus fort %CC chez les non musiciens, ceux liés à *pincé et grave*.

Dans le sous-groupe des **adjectifs simples négatifs** (voir tableau A8.I.2), on observe en moyenne un plus fort %RC chez les non musiciens que chez les non musiciens. Cette différence est à souligner car c'est rare qu'elle se présente dans ce sens. On retrouve des différences significatives chez 7 des 21 portraits de ce sous-groupe, 4 pour lesquels on observe aussi un plus fort %RC chez les non musiciens (pour *pas grave, pas pincé, pas clair et pas vibré*), et 3 pour lesquels on observe un plus fort %RC chez les musiciens (pour *attaque lente, pas rond et pas soufflé*). On retrouve des différences entre musiciens et non musiciens sur l'ensemble de ce sous-groupe en terme de %CC mais dans l'autre sens : les %CC moyens étant plus fort au sein des musiciens que chez les non musiciens. Par contre, on trouve plus de différences significatives (pour 13 des 21 portraits). Seules 2 de ces différences mènent à un plus fort %CC chez les non musiciens. Il s'agit des portraits *pas riche et pas grave*. Les autres différences menent toutes à un plus fort %CC chez les musiciens, en particulier pour les portraits *pas soufflé, pas rond, pas métallique et pas diffus*.

Pour l'autre sous-groupe d'**adjectifs simples** (voir tableau A8.I.3), on observe globalement de plus forts %RC chez les musiciens que chez les non musiciens, des différences significatives entre les deux groupes de sujets apparaissant pour 7 des 16 portraits (soit 44%). Deux de ces différences mènent à de plus forts %RC chez les non musiciens, il s'agit des portraits *constant et terminé*, les autres menant tous à un plus fort %RC chez les musiciens, notamment pour les portraits *franc, agréable, fin, double et intérieurs*. En ce qui concerne les %CC, les différences entre musiciens et non musiciens ne sont globalement pas significatives et ne sont significatives que pour 3 des 16 portraits. Il s'agit pour deux d'entre eux de portraits où les non musiciens obtiennent de plus fort %CC (*constant et double*) et pour un d'entre eux (*agréable*) un portrait pour lequel les musiciens ont un plus fort %CC.

Pour le sous-groupe des **adjectifs doubles** (voir tableau A8.I.4), les différences en terme de %RC ne sont globalement pas significatives, malgré le fait qu'elles le soient pour 5 des 21 portraits de ce groupe. On observe en effet pour 4 d'entre eux de plus forts %RC chez les musiciens (pour *rapide et précis, sourd et flasque, métallique et cuivré et rond et soufflé*), et pour *pur et net*, un plus fort %RC chez les non musiciens. Par contre, les différences en terme de %CC sont globalement significatives, dans le sens de plus fort %CC chez les non musiciens, ce qui est rare. En effet, pour 8 des 21 portraits, on observe des différences significatives entre les deux groupes de sujets. Pour 6 d'entre elles, les non musiciens obtiennent de plus forts %CC. Il s'agit principalement des portraits *ni mou, ni agressif et froid et agréable*, 2 d'entre eux menant à de plus forts %RC chez les musiciens, notamment *pincé et résonnant*.

Enfin, pour le sous-groupe des portraits construits à partir de **trois adjectifs** (voir tableau A8.I.5), les différences en terme de %RC sont globalement très significatives et mènent à de plus forts %RC chez les musiciens. On retrouve des différences significatives pour 8 des 12 portraits, une seule allant dans le sens de plus forts %RC chez les non musiciens (il s'agit du portrait visant le son TPT), les 7 autres menant à de plus forts %RC chez les musiciens. Il s'agit principalement des portraits visant les sons CFT, TBN et VBR. En terme de %CC, les différences sont elles aussi globalement très significatives et dans le sens d'un plus fort %CC chez les musiciens. Elles ne se retrouvent pourtant que pour 4 des 12 portraits, qui mènent toutes à de plus forts %CC chez les musiciens. Il s'agit des portraits visant les sons VBR, CFT, puis PNO et CLA.

Dans le groupe des portraits décrivant des sources sonores possibles des sons, pour le sous-groupe des **modes d'excitation** (voir tableau A8.I.6), on observe globalement de plus forts %RC chez les musiciens que chez les non musiciens, ce qui se retrouve pour 4 des 12 portraits de ce sous-groupe. En effet, sur ces 4 différences significatives, une seule aboutit à de plus forts %RC chez les non musiciens (pour le portrait *on percute*), les 3 autres menant à des plus forts %RC chez les musiciens (pour *on tire*, *on effleure* et *on froisse*, qui visaient tous les trois le mode de production de la CFT de façon moins directe que *on frotte*). Les différences de %CC ne sont pas significatives globalement, même si elles le sont pour 2 des 12 portraits. Il s'agit de *on effleure*, qui présente un plus fort %CC chez les musiciens et *on tape*, qui présente un plus fort %CC chez les non musiciens.

Dans le sous-groupe des noms de **matériaux** (voir tableau A8.I.7), les différences entre musiciens et non musiciens ne sont globalement pas significatives, même si on en observe pour 4 des 17 portraits. Parmi celles-ci, une seule différence mène à un plus fort %RC pour les non musiciens (*du bois*), les 3 autres ayant de plus forts %RC chez les musiciens. Il s'agit des portraits *du cristal*, *de la peau* et *du métal* et *une corde en boyau*. On peut remarquer que les deux derniers portraits sont parmi les plus précis et les plus élaborés du sous-groupe. En terme de %CC, les différences globales ne sont pas non plus significatives, même si on les retrouve pour 3 des 17 portraits. On retrouve *du bois*, qui présente un plus fort %CC chez les non musiciens, *un matériau synthétique* et *une corde métallique* présentant de plus fort %CC chez les musiciens.

Dans le sous-groupe des noms de **sources** construits à partir de modes d'excitation et de matériaux (voir tableau A8.I.8), on trouve globalement de plus forts %RC chez les musiciens que chez les non musiciens, ce qui se retrouve chez 11 des 26 portraits de ce sous-groupe. Deux de ces différences sont en faveur de plus forts %RC chez les non musiciens (pour *on étouffe le son* et *une corde vibre dans du verre*), les 9 autres menant à de plus forts %RC chez les musiciens, en particulier pour *on frotte une corde*, *on frappe du verre* et *un morceau de métal mis en résonance*. En terme de %CC les différences sont globalement non significatives, même si on en observe pour 10 des 26 portraits. Sept de ces différences mènent à de plus forts %CC chez les musiciens, en particulier pour *on frappe du verre* et *on frotte une corde*, 3 menant à de plus forts %CC chez les non musiciens, en particulier *on étouffe le son* et *une corde vibre dans du verre*.

Dans le sous-groupe des noms de **familles d'instruments** (voir tableau A8.I.9), on trouve globalement de plus fort %RC chez les musiciens que chez les non musiciens, ce qui se retrouve pour 4 des 6 portraits (en particulier, les *familles des vents, bois, cuivres* et *cordes*) pour lesquels les différences mènent toutes à de plus forts %RC chez les musiciens. On retrouve le même phénomène en terme de %CC : globalement et pour 5 des 6 portraits on obtient de plus forts %CC au sein du groupe des musiciens, en particulier pour les portraits des *familles des vents, cordes* et *cuivres*.

Dans le sous-groupe des noms **d'instruments de musique** (voir tableau A8.I.10), les différences entre musiciens et non musiciens sont globalement non significatives pour les %RC, même si on en retrouve pour 17 des 37 portraits. Parmi celles ci, seules 6 d'entre elles vont dans le sens de plus forts %RC chez les non musiciens (en particulier pour les portraits : *synthétiseur, trompe*, mais aussi *accordéon* et *cor Anglais*, ce qui est plus surprenant), les 11 autres différences significatives menant à de plus forts %RC chez les musiciens (en particulier pour les portraits *clavecin, cornet à piston, clarinette, trombone...*). Par contre, les différences en terme de %CC sont globalement significatives et mènent à de plus forts %CC chez les musiciens. On les retrouve pour 15 des 37 portraits (soit 40%). Seules 2 de ces différences mènent à de plus forts %CC chez les non musiciens (pour *violon*, ce qui est étonnant, et pour *tambour*), les autres différences étant en faveur de plus forts %CC entre musiciens. Il s'agit en particulier des portraits *clavecin, trombone, cor, synthétiseur, cornet à piston* et *trompette*. On peut noter que chez les musiciens, le portrait xylophone totalise 100% de réponses communes ainsi que 100% de Choix Communs, ce portrait ayant été associé par les 10 participants musiciens uniquement au son VBR.

Enfin, pour le sous-groupe des **instruments visés par le concepteur** (voir tableau A8.I.11), les musiciens obtiennent en moyenne de plus forts %RC que les non musiciens. Ceci se retrouve pour 5 des 12 portraits, en particulier ceux de *vibraphone, piano électrique, clavecin, clarinette* et *trombone*. Les différences en terme de %CC sont encore plus significatives : les musiciens obtenant des %CC près de 2 fois plus grands que les non musiciens, ce qui se retrouve dans 7 des 12 portraits de ce sous-groupe. On retrouve en particulier des portraits de *vibraphone, piano électrique, clavecin* et *trombone*. Il faut remarquer ici aussi que le portrait de *vibraphone* mène à 100% de RC et de CC chez les musiciens.

### I.3. Conclusion

L'accord entre les participants est en moyenne plus important sur l'ensemble des portraits décrivant la source sonore que sur ceux décrivant le son, qu'on considère le %RC ou le %CC, même si seule la différence correspondant au %RC est significative. De plus, parmi les portraits référant à la source sonore, ce sont ceux des instruments visés par le concepteur et ceux des noms d'instruments (tous deux construits à partir de noms d'instruments de musique) qui mènent au plus fort pourcentage de Réponses Communes (plus de 87%), le sous-groupe de noms de familles d'instruments présentant lui

le plus fort pourcentage de Choix Communs entre les participants (56%). On peut noter que la plupart des noms de sources sonores visant le son VBR, que ce soit sous forme de matériau (*du verre, du cristal...*) ou d'instrument de musique (*xylophone, vibraphone...*) obtiennent les plus forts %RC et %CC, faisant même parfois l'unanimité des musiciens.

On peut aussi noter globalement une plus forte cohérence du choix des associations des sons aux portraits entre les sujets quand ils sont musiciens que quand ils sont non-musiciens. Ces différences sont significatives en terme de %CC pour l'ensemble des portraits, mais aussi dans le groupe des portraits décrivant une source sonore, en particulier ceux des instruments visés par le concepteur, de noms de familles d'instruments et de noms d'instruments de musique. Pour les portraits décrivant le son, seul le sous-groupe construit à partir d'adjectifs négatifs présente un plus fort %CC significatif pour les musiciens. En terme de %RC, on n'observe que deux différences significatives, toujours dans le sens d'un plus fort %RC pour les musiciens que pour les non musiciens. Il s'agit des sous-groupes de portraits construits à partir de trois adjectifs et celui des instruments visés par le concepteur.



## II. Sons choisis pour chaque portrait

Nous allons analyser ici les résultats liés aux choix par les sujets, pour chacun des portraits proposés, "du ou des sons qui correspondent le mieux" parmi les 12 sons proposés. Pour cela, nous allons d'une part décrire les résultats les plus fréquents, notamment à l'aide du pourcentage de sujets ayant fait chacun des choix et d'autre part la répartition des réponses des sujets sur l'ensemble des 12 sons proposés à l'aide d'une mesure de  $\chi^2$ .

Ces analyses ont été réparties en fonction des différentes parties de l'expérience que nous avons menée et qui correspondent aux différentes façons pour les sujets de décrire les sons, soit en décrivant des qualités de ces sons (à partir d'un ou plusieurs adjectifs associés entre eux), soit en décrivant une source sonore (sous forme d'action, de matériaux, d'un assemblage des deux ou sous forme de nom d'instruments de musique ou de nom de familles d'instruments).

### II.1. Dispersion des réponses sur l'ensemble des timbres proposés

Nous avons calculé pour chaque portrait un  $\chi^2$  correspondant à la répartition des réponses sur l'ensemble des 12 timbres proposés, comparée à une distribution égale sur l'ensemble des sons. Cette mesure est en effet significative si la répartition des réponses des sujets sur les 12 sons ne se fait pas au hasard mais au contraire est focalisée sur un nombre réduit de sons.

Tableau 8.2. Répartition des  $\chi^2$  moyens de chaque sous-catégorie de portraits et comparaison par rapport aux moyennes des 2 groupes son et source.

Type de Portrait	NB de portraits	$\chi^2$ moyen	dl	t	p
Total SON	91	45.4			
Qualités simples positives	21	47.0	20	0.47	
Qualités simples négatives	21	31.9	20	5.14	*
Autres Qualités simples	16	46.2	15	0.17	
Qualités doubles	21	46.8	20	0.31	
Qualités triples	12	62.3	11	2.64	†
Total SOURCE	110	73.1			
Actions	12	68.5	11	0.49	
Matériaux	17	56.8	16	2.16	†
Sources	26	68.5	25	0.70	
Familles d'Instruments	6	85.2	5	1.82	
Noms d'instruments	37	78.6	36	0.80	
Instruments visés par le concepteur	12	88.1	11	1.32	
TOUS	201	60.6			

Avec \* :  $p < .01$ , † :  $p < .05$ .

Si on considère le  $\chi^2$  calculé pour chaque portrait pour tester la répartition des réponses des sujets sur les 12 timbres proposés, la différence entre les moyennes correspondant aux deux groupes de portraits : ceux qui décrivent le son et ceux qui décrivent la source, est significative ( $t(199) = -6.46$ ,

$p < .0001$ ). Les choix des participants sont en effet moins dispersés sur l'ensemble des 12 sons quand les portraits décrivent une source sonore possible que lorsqu'ils décrivent des qualités des sons.

De plus, parmi les portraits décrivant le son, ceux construits à partir d'adjectifs simples négatifs présentent des choix significativement plus dispersés sur l'ensemble des 12 sons que la moyenne de ce groupe ( $t(21)=5.14$ ,  $p < .01$ ). Ceux construits à partir de trois adjectifs extraits de l'expérience 2 et visant chacun un des 12 sons en particulier présentent au contraire des choix moins dispersés sur l'ensemble des 12 sons ( $t(12)=2.64$ ,  $p < .05$ ). Par contre, parmi les portraits décrivant une source sonore possible, seuls ceux construits à partir de noms de matériaux présentent une répartition sur l'ensemble des 12 sons significativement plus dispersée que pour l'ensemble des portraits de ce groupe.

## II.2. Portraits décrivant le son

Nous allons retrouver ici les différents sous-groupes de portraits construits à partir d'adjectifs de qualité, notamment l'ensemble des adjectifs testés sous forme d'échelle sémantique dans notre deuxième expérience, et choisis comme les adjectifs les plus fréquemment employés par les sujets pour comparer les sons deux à deux lors des verbalisations libres de notre première expérience. Nous y avons ajouté un ensemble de 16 autres adjectifs utilisés les plus fréquemment pour décrire ces mêmes sons mais pas forcément dans un processus de comparaison de deux sons d'une paire. Nous avons ensuite testé 21 associations de deux de ces adjectifs, produites spontanément par les sujets, afin de voir si les portraits devenaient plus spécifiques de certains sons. Enfin, nous avons testé 12 portraits construits en associant trois adjectifs de qualité choisis comme portrait de chacun des sons à partir des échelles sémantiques de l'expérience 2 que les sujets ont utilisées de façon extrême pour ces sons.

L'ensemble des réponses donnant les nombres de sujets ayant associé chacun des sons aux portraits proposés ainsi que les pourcentages de réponses dues aux musiciens, se trouvent dans l'Annexe du chapitre 8, tableaux A.8.12. à A.8.31.

### **II.2.1. Qualités simples testées dans l'expérience 2**

Nous allons donc analyser les choix des sujets dans la partie de l'expérience : qualités simples issues de l'expérience 2. Dans ce paragraphe, nous allons appeler portraits "positifs" ceux de la forme "un son aigu" et "négatifs" ceux de la forme "un son pas aigu" (voir tableaux 8.2 et 8.3). Pour chacun des portraits proposés, nous avons testé la répartition des réponses des sujets afin de voir si elle se faisait au hasard. Sur ces 21 portraits de qualité simple, les réponses concernant un seul d'entre eux "*un son vibré*" semblent se répartir au hasard ( $\chi^2(11)=18$ ,  $p=1$ ). Ensuite, deux autres portraits ont un  $\chi^2$  correspondant à  $.01 > p > .001$ , ce sont les portraits "*un son large*" et "*un son riche*". Pour tous les autres portraits, on a  $\chi^2 > 34$  et  $p < .001$ . Globalement donc, pour la plupart de ces portraits, les sujets ne répondent pas au hasard.

De même que pour les qualités "positives", pour un grand nombre de portraits "négatifs" (19 sur 21, voir tableau 8.3) la répartition des réponses des sujets se fait de manière significativement différente sur les différents sons proposés. Seules les réponses concernant "un son pas vibré" et "un son pas large" semblent se répartir au hasard (respectivement  $\chi^2=14$  et  $18$ ,  $p>.05$ ). Par contre, on a ici 10 portraits qui ont un  $\chi^2$  correspondant à  $.01>p>.001$ . On peut donc dire que la répartition des réponses des sujets est moins spécifique pour les portraits "négatifs" que pour les portraits "positifs", même si ici aussi les sujets ne répondent pas au hasard.

**Tableau 8.3.** Timbres associés à chacun des portraits de qualités simples testés dans l'expérience 2 : 2 des répartitions ( $dl=11$ ) et proportion des sujets ayant choisi chaque son pour un portrait donné.

Portrait	$\chi^2$ p	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
aigu	70 °	*	.	*	.	.	.	.	.	**	***	.	.
attaque rapide	80 °	.	.	.	.	.	***	***	*	**	***	**	.
brillant	36 °	*	.	.	.	.	**	*	.	**	*	.	.
clair	61 °	.	.	.	.	.	***	**	.	**	*	.	.
diffus	40 °	.	*	.	*	***	.	.	*	.	.	**	**
doux	47 °	.	.	.	**	*	*	.	*	.	.	.	.
fort	62 °	*	*	.	.	.	*	***	.	.	.	.	.
grave	40 °	.	**	.	**	*	*	.	*	.	.	*	.
large	25 †	.	*	.	*	*	*	.	*	.	.	*	.
long	40 °	.	*	.	.	**	*	.	***	.	.	**	*
métallique	56 °	.	.	.	.	.	**	**	.	***	**	.	.
nasillard	55 °	**	*	**	.	**	.	.	.	*	**	.	.
net	41 °	*	.	.	*	.	**	**	.	**	*	.	.
pincé	54 °	.	.	.	.	.	*	*	.	***	**	.	.
riche	27 *	.	*	.	.	*	**	.	*	*	*	**	.
rond	51 °	.	**	.	**	*	**	.	*	.	.	*	.
résonnant	41 °	.	.	.	.	*	**	.	**	.	*	**	*
sec	36 °	*	.	*	.	.	*	**	.	**	**	*	.
soufflé	73 °	**	***	*	***	*	.	.	.	.	.	*	.
sourd	34 °	.	*	.	**	.	.	.	**	.	.	**	*
vibré	18 ns	.	*	.	.	*	*	*	*	*	*	**	*

avec ° :  $p<.001$ ; \* :  $p<.01$ , † :  $p<.05$ , ns : non significatif, et avec \*\*\* si plus de 3/4, \*\*: plus de 1/2 et \* : plus de 1/4 des sujets ("." de 1 à 4 sujets) ont choisi ce son comme correspondant au portrait considéré.

Sur les 21 portraits de qualités positives, seuls 9 d'entre eux sont associés à au moins un même son par plus de 3/4 des participants (on retrouve ceux qui ont un %RC supérieur à la moyenne dans le §I). Ainsi, le portrait *aigu* est associé au son OBC, *clair* au VBR, *diffus* à la CFT, *fort* au PNE, *long* au PNO, *métallique* et *pincé* au CVC. De plus *soufflé* est associé par plus de 3/4 des participants aux sons TBN et CNT et le portrait *attaque rapide* aux 4 sons : VBR, PNE, CVC et OBC.

De même, sur les 21 portraits de qualités négatives, 14 d'entre eux sont associés à au moins un même son par plus de 3/4 des participants. Ainsi, les portraits *attaque lente* et *pas fort* sont associés au son CFT, *pas brillant* et *pas métallique* à la CNT, *pas nasillard* au VBR, *pas net* au VBN, *pas rond* au OBC et *pas vibré* à la TPT. De plus, 3 portraits sont associés par plus de 3/4 des sujets à 2 sons différents. Il s'agit de *pas résonnant*, associé à la TPT et au CRA, *pas sec* associé au TBN et à la CFT et *pas sourd*, associé au VBR et au CVC. Enfin, *pas doux* est associé par plus de 3/4 des sujets aux

trois sons TPT, PNE et OBC, *pas soufflé* aux 4 sons VBR, PNE, CVC et OBC et *pas pincé* aux 5 sons TPT, TBN, CRA, CNT et CFT. On peut remarquer que les réponses des sujets correspondant au portrait *attaque lente* ne concernent que des sons, soit imitant des sons soufflés, soit des sons imitant des productions par frottement. Ce sont tous les sons qui n'ont pas été cités dans le portrait *attaquerapide*. Ces deux portraits permettent donc bien de définir deux groupes de sons dans notre ensemble. Pourtant seul le son CFT est cité ici par plus de 3/4 des sujets.

Tableau 8.4. Timbres associés à chacun des portraits de qualités simples "négatives" : <sup>2</sup> des répartitions et proportion de sujets ayant donné chaque réponse.

Portrait	2 p	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
pas aigu	21 †	*	**	*	**	*	*	.	**	.	.	*	*
attaque lente	61 °	*	**	*	**	***	.	.	.	.	.	.	*
pas brillant	25 *	*	**	**	***	*	.	.	**	.	*	**	*
pas clair	28 *	*	*	*	*	**	.	.	*	.	*	**	*
pas diffus	24 †	**	*	*	*	.	*	**	.	**	**	.	.
pas doux	34 °	***	*	**	.	.	.	***	*	**	***	*	*
pas fort	35 °	.	.	.	*	***	*	.	*	*	.	**	*
pas grave	26 *	*	.	*	.	*	**	*	.	**	**	.	*
pas large	18 ns	**	.	**	*	*	*	*	.	**	**	*	*
pas long	25 *	**	*	**	**	.	*	**	.	**	**	.	.
pas métallique	33 °	**	**	**	***	**	.	.	*	.	.	*	*
pas nasillard	30 *	*	*	.	**	*	***	*	**	*	.	*	*
pas net	40 °	.	*	*	.	**	.	.	*	.	*	***	*
pas pincé	27 *	***	***	***	***	***	*	*	*	.	.	**	*
pas riche	20 †	**	*	**	**	.	.	*	*	.	.	*	*
pas rond	26 *	**	.	**	.	*	.	**	.	**	***	*	*
pas résonnant	46 °	***	*	***	**	.	.	.	.	*	*	.	.
pas sec	41 °	*	***	*	**	***	*	.	**	.	.	**	*
pas soufflé	55 °	.	.	.	.	*	***	***	**	***	***	**	*
pas sourd	41 °	*	.	*	.	*	***	**	*	***	**	.	.
pas vibré	14 ns	***	*	**	**	*	*	*	*	*	*	*	*

avec °: p<.001; \*: p<.01, † : p<.05, ns : non significatif, et avec \*\*\* si plus de 3/4, \*\*: plus de 1/2 et \*: plus de 1/4 des sujets ("." de 1 à 4 sujets) ont choisi ce son comme correspondant au portrait considéré.

Le portrait *pas soufflé* semble correspondre à l'ensemble des 6 sons imitant une production pincée ou frappée et de manière systématique (plus de 3/4 des sujets) pour 4 d'entre eux. Les deux sons imitant une production par frottement (CFT et SNO) ne sont cités ici que par 1/4 des sujets. Ce sont les seuls sons qui permettent de différencier le couple de portraits *soufflé / pas soufflé* du couple *attaquerapide / lente*. Ils sont cités uniquement dans le deuxième couple de portraits. On peut enfin remarquer qu'une seule réponse a été donnée à l'unanimité par les 20 sujets : il s'agit du caractère *pas soufflé* du son VBR.

#### Différence entre portraits "positifs" et "négatifs"

Dans la comparaison entre les portraits "positifs" et "négatifs", il est intéressant de remarquer aussi que sur l'ensemble des 21 portraits, 13 sons ont été associés à l'un des portraits "positifs" par plus de 3/4 des sujets alors que cela concerne 26 sons pour les portraits "négatifs". Les productions spontanées des sujets étant très souvent de type positif, sont sans doute plus spécifiques de quelques sons. Les sujets associent moins de sons différents à une version positive qu'à une version négative

du même portrait. Par exemple, alors que seul le son CVC est considéré comme correspondant au portrait *un son pincé* par plus de 3/4 des sujets, les sons TPT, TBN, CRA, CNT et CFT sont tous les 5 choisis par plus de 3/4 des sujets comme correspondant au portrait *un son pas pincé*. De plus, pour une même qualité, si on compare les résultats "positifs" et "négatifs", ils ne sont pas fortement anti-corrélés. En moyenne le coefficient de corrélation entre l'ensemble des données correspondant aux portraits "positifs" et "négatifs" est de  $r(440)=-.78$  ( $p<.01$ ).

#### Différence entre musiciens et non-musiciens

Sur l'ensemble des choix communs aux 3/4 des sujets, la moyenne du pourcentage du aux musiciens est de 52% pour les portraits "positifs" et de 47% pour les portraits "négatifs" (cf. annexes A.8.18 et A.8.19). Pour les portraits "positifs", les associations sons/portrait choisies par plus de 3/4 des sujets sont dues plus souvent aux musiciens dans le cas de l'association OBC/*attaquerapide* (67% dû aux musiciens) et aussi PNE et CVC/*attaque rapide*. Par contre, l'association CVC/*pincé* est due aux non musiciens pour 56%. On trouve aussi des réponses données par plus de 3/4 des musiciens mais qu'on ne retrouve pas dans les réponses choisies par plus de 3/4 de l'ensemble des sujets. Il s'agit des 6 associations suivantes : CVC/*net*, TBN et CNT/*rond*, VBR/*résonnant*, CVC/*sec* et VBN/*sourd*. De la même façon, certaines réponses sont choisies par plus de 3/4 des sujets non musiciens bien que cette proportion ne se retrouve pas au niveau de l'ensemble des sujets. C'est le cas des associations TBN/*grave*, PNE/*net*, PNE et OBC/*pincé* et VBN/*rond*. Il est étonnant par exemple que l'association OBC/*pincé* ne soit pas réalisée par les musiciens, c'est en effet une des associations attendues.

Pour les portraits "négatifs", parmi les associations communes à plus de 3/4 des sujets, les associations suivantes : TPT/*pas doux*, VBN/*pas net*, OBC/*pas rond*, PNE et OBC/*pas soufflé* et VBR/*pas sourd* sont plus souvent dues aux musiciens, alors qu'aucune de ces associations n'est plus souvent due aux non-musiciens. Parmi les Réponses Communes à 3/4 des musiciens seulement, on trouve 10 autres associations : OBC, TPT et CVC/*pas diffus*, TBN/*pas métallique*, CNT/*pas nasillard*, CVC/*pas rond*, CNT/*pas résonnant*, CNT/*pas sec*, PNE/*pas sourd* et CNT/*pas vibré* alors que seulement une association est faite par 3/4 des non musiciens, seulement CRA/*pas doux*. Il semble donc que les réponses des deux groupes de sujets soient assez différentes pour les portraits de qualité simple "négatifs". De plus, si on calcule un  $\chi^2$  sur l'ensemble des résultats des musiciens et des non musiciens (42\*12 cases) on obtient une répartition qui ne dépend pas du niveau d'expertise musicale des sujets ( $\chi^2(503)=361$ ,  $p=1$ ).

#### **II.2.2. Autres qualités simples**

Nous avons repris l'ensemble de ces analyses pour les portraits de qualité simple non testés dans l'expérience 2, mais choisis comme faisant partie des adjectifs les plus utilisés pour décrire les sons dans l'expérience 1 pour caractériser les sons et non plus les comparer 2 à 2. Les résultats sont rassemblés dans le tableau 8.5. En moyenne sur l'ensemble de ces 16 portraits, la valeur de  $\chi^2$  est de  $M=46\pm 20$ . On observe donc de grandes variations entre ces différents portraits. Ainsi, le portrait *un son terminé* voit la répartition de ses réponses considérée comme aléatoire sur l'ensemble des sons. Le

portrait *un son double* a aussi une valeur de  $\chi^2$  plus faible que les autres qui correspondent à un seuil  $p < .05$ . Pourtant, 11 de ces 16 portraits ont des  $\chi^2$  qui correspondent à des valeurs de  $p < .001$ . Ils sont donc relativement spécifiques de certains sons. Sur l'ensemble des 12\*16 associations possibles entre un son et un portrait, seules 6 d'entre elles (soit 3%) ont été faites par plus de 3/4 des sujets. Elles concernent toutes un seul son associé à un seul portrait. En effet, on trouve les associations VBR/*agréable*, CNT/*constant*, CRA/*continu*, TBN/*cuivré*, VBN/*étouffé* et VBR/*pur*.

Tableau 8.5. Timbres associés à chacun des portraits d'autres qualités simples :  $\chi^2$  des répartitions et proportion de sujets ayant donné chaque réponse.

Portrait	$\chi^2$ p	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
agressif	57 °	*	.	*	.	.	.	*	.	.	**	.	.
agréable	63 °	.	.	.	*	.	***	.	*	*	.	.	.
chaleureux	37 °	.	*	.	*	.	*	.	*	*	.	.	.
constant	66 °	**	.	**	***	.	.	.	*	*	.	.	.
continu	46 °	**	*	***	**	.	.	.	*	.	.	.	.
court	27 *	*	.	*	*	.	*	**	.	.	*	*	.
cuivré	81 °	**	***	.	.	.	.	.	.	.	.	*	.
double	25 †	.	.	.	.	*	.	.	.	*	*	**	.
éloigné	50 °	.	.	.	.	**	.	.	*	.	.	**	*
étouffé	51 °	.	*	*	**	*	.	.	*	.	.	***	*
fin	27 *	.	.	.	.	.	*	.	.	**	*	*	.
franc	48 °	**	.	*	.	.	**	**	.	*	.	.	.
intérieur	26 *	.	*	.	**	*	.	.	*	.	.	*	.
précis	50 °	.	.	*	*	.	**	**	.	**	*	.	.
pur	76 °	.	.	.	*	.	***	.	.	*	.	.	.
terminé	9 ns	.	.	*	*	.	*	*	*	*	*	*	.

avec °:  $p < .001$ ; \*:  $p < .01$ , †:  $p < .05$ , ns: non significatif, et avec \*\*\* si plus de 3/4, \*\*: plus de 1/2 et \*: plus de 1/4 des sujets ("." de 1 à 4 sujets) ont choisi ce son comme correspondant au portrait considéré.

#### Différence entre musiciens et non-musiciens

Pour l'ensemble des 6 réponses communes à plus de 3/4 des sujets, le % moyen dû aux musiciens est de 48% (voir tableau A.8.17). Aucune de ces réponses ne correspond donc à un choix principalement dû aux musiciens. Par contre, les associations CNT/*constant* et CRA/*continu* sont plus souvent dues aux non musiciens (pour 56%). On observe aussi des associations qui sont choisies par plus de 3/4 des musiciens bien que ce ne soit pas le cas sur l'ensemble des sujets. Il s'agit ici des associations CNT/*continu* et PNE/*franc*. De même, les associations CRA/*constant*, VBR/*précis* et VBN/*double* sont produites par plus de 3/4 des non musiciens bien que ce ne soit pas le cas sur l'ensemble des sujets. Il est par exemple étonnant que le caractère *double* du son VBN ne soit ici noté de façon systématique que par les non musiciens. On peut aussi noter le fait que tous les non musiciens font l'association CNT/*constant*.

#### II.2.3. Association de deux qualités

Nous avons repris ces analyses sur les portraits construits en associant deux qualités simples. On rappelle que ces portraits correspondent aux associations produites spontanément par les sujets dans l'expérience 1 et contenant au moins un adjectif déjà testé dans la partie qualités simples. Les résultats sont rassemblés dans le tableau suivant. Les valeurs de  $\chi^2$ , correspondant à la répartition des réponses

des sujets sur l'ensemble des 12 sons proposés pour chacun de ces 21 portraits, dépend beaucoup du portrait . En effet, en moyenne sur ces 21 portraits on a  $\chi^2=47\pm 21$ . Les valeurs correspondant aux portraits *un son lourd et continu* et *un son court et ramassé* ne sont pas significatives. La répartition des sons choisis pour ces deux portraits se fait donc de manière non différente du hasard. Pourtant, globalement, 15 des 21 portraits de qualités doubles bénéficient de  $\chi^2$  associés à des  $p < .001$ . On peut donc considérer que ici aussi, le choix des sujets ne se fait globalement pas au hasard.

Tableau 8.6. Timbres associés à chacun des portraits d'association de 2 qualités :  $\chi^2$  des répartitions, et proportion de sujets ayant donné chaque réponse.

Portrait	$\chi^2$	p	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
aigu et métallique	78	°	.	.	.	.	.	.	*	.	**	***	.	.
clair et doux	60	°	.	.	.	*	.	***	.	.	*	.	.	.
clair et rond	78	°	.	.	.	.	.	***	.	.	.	.	.	.
court et ramassé	13	ns	.	*	.	.	.	.	*	.	.	.	.	.
doux et froid	31	*	.	.	.	*	*	*	.	.	.	.	.	.
froid et agréable	43	°	.	.	.	.	.	**	.	.	*	.	.	.
long et diffus	45	°	.	.	.	.	**	.	.	*	.	.	*	.
lourd et continu	18	ns	*	.	.	.	*	.	.	*	.	.	.	.
métallique et cuivré	27	*	*	*	.	.	.	.	.	.	.	*	*	.
métallique et froid	50	°	.	.	.	.	.	.	.	.	*	**	.	.
net et franc	67	°	*	.	.	.	.	**	***	.	**	.	.	.
ni mou, ni agressif	25	*	.	**	.	*	.	**	.	*	*	.	.	.
pincé et résonnant	63	°	.	.	.	.	.	.	*	*	**	**	.	.
pincé et étouffé	32	°	.	.	.	.	.	.	.	*	.	*	*	.
pur et net	61	°	.	.	.	.	.	***	*	.	**	.	.	.
rapide et précis	38	°	*	.	*	*	.	*	**	.	*	*	.	.
rond et doux	38	°	.	*	.	*	.	*	.	*	.	.	.	.
rond et soufflé	84	°	.	**	.	**	.	.	.	.	.	.	.	.
sourd et flasque	41	°	.	*	.	.	*	.	.	.	.	.	**	.
sourd et grave	71	°	.	**	.	**	.	.	.	*	.	.	*	.
sourd et métallique	20	†	.	*	*	.	.	.	.	*	.	.	*	.

avec °:  $p < .001$ ; \*:  $p < .01$ , † :  $p < .05$ , ns : non significatif, et avec \*\*\* si plus de 3/4, \*\*: plus de 1/2 et \*: plus de 1/4 des sujets ("." de 1 à 4 sujets) ont choisi ce son comme correspondant au portrait considéré.

Sur l'ensemble des 12\*21 choix possibles, 5 d'entre eux (soit 2%) ont été effectués par plus de 3/4 des sujets. Ce taux est très faible. Une des raisons possibles est que ces portraits sont issus de combinaisons spontanées qui pour la plupart n'ont été produites que par rapport à un seul son et très souvent avec très peu d'occurrences. Il s'agit des associations : OBC/*aigu et métallique*, VBR/*clair et doux*, VBR/*clair et rond*, PNE/*net et franc* et VBR/*pur et net*. On peut remarquer que sur ces 5 portraits 3 sont associés au son VBR.

#### Différence entre musiciens et non-musiciens

Sur les 5 réponses choisies par plus de 3/4 des sujets on a 47% de ces réponses dues aux musiciens en moyenne. La participation des musiciens est un peu plus forte pour l'association PNE/*net et franc* et celle des non musiciens plus forte pour l'association OBC/*aigu et métallique*. Cette dernière association était pourtant attendue lors de la construction des portraits. Parmi les réponses dues à 3/4 des sujets musiciens mais pas à l'ensemble des sujets, on trouve deux associations : CVC/*aigu et*

métallique et TBN/*rond et soufflé*. De même les associations VBR/*ni mou ni agressif* et OBC/*pincé et résonnant* se retrouvent chez 3/4 des non musiciens mais pas chez l'ensemble des participants. On peut noter aussi que l'association OBC/*aigu et métallique* est faite par tous les non musiciens.

#### II.2.4. Association de 3 qualités

Nous allons maintenant nous pencher sur les résultats obtenus à partir des portraits construits comme association de trois qualités simples choisies parmi les échelles sémantiques de l'expérience 2 qui caractérisent chacun de nos 12 sons de la manière la plus extrême. Chacun de ces portraits étant construit pour décrire un des 12 sons de notre ensemble, nous allons par commodité les nommer par leur code plutôt que par la liste des 3 adjectifs qui les composent. Ainsi le portrait *nasillard, peu pincé et peu doux*, construit pour caractériser le son TPT sera noté P3-TPT (voir tableau 8.7).

Tableau 8.7. Timbres associés à chacun des portraits d'association de 3 qualités : proportion de sujets ayant donné chaque réponse et  $\chi^2$  de répartition des réponses sur les 12 sons.

Portrait	$\chi^2$	p	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
P3-TPT: nasillard, peu pincé et peu doux	76	°	**		**		.		.			.	.	
P3-TBN: sourd, bas et peu pincé	28	*	.	*	.	*	.	.	.	*			*	.
P3-CRA: nasillard, assez aigu, peu résonnant	72	°	**	.	**				.		.	**	.	.
P3-CNT: doux, sourd et soufflé	67	°		**		**	.						.	.
P3-CFT: long, diffus et attaque lente	71	°		*	.	.	***	.		.			.	*
P3-VBR: attaque rapide, clair et doux	89	°		.	.	.		***	.	*	*	.	.	
P3-PNE: attaque rapide, net et fort	75	°	*	.	.	.		*	***	.	*	.	.	
P3-PNO: peu soufflé, résonnant et doux	45	°		.	.	.	*	*		*	.		*	.
P3-CVC: pincé, aigu et métallique	92	°	.	.	.	.	.	.	.		**	***	.	.
P3-OBC: pincé, nasillard et métallique	68	°	.	.	.	.	.	.	.		**	**	.	.
P3-VBN: sourd, bas et peu net	29	*	.	*	.	.	.			*	.		*	*
P3-SNO: assez long, assez doux et peu sec	36	°	.	*	.	*	*	*		*	.		.	.

avec °:  $p < .001$ ; \*:  $p < .01$ , † :  $p < .05$ , ns : non significatif, et avec \*\*\* si plus de 3/4, \*\*: plus de 1/2 et \*: plus de 1/4 des sujets ("." de 1 à 4 sujets) ont choisi ce son comme correspondant au portrait considéré.

Les réponses des sujets pour chacun de ces 12 portraits se répartissent de manière non aléatoire sur les 12 sons proposés ( $\chi^2 > 28$ ,  $p < .01$  et pour 10 d'entre eux  $\chi^2 > 36$ ,  $p < .001$ ). Nous avons vu §II.1. qu'en moyenne, la répartition sur les 12 sons était ici la moins dispersée du groupe des portraits décrivant le son. De même, nous avons vu dans le §I. que les pourcentages de réponses communes sont plus forts que dans les autres cas de qualités simples ou doubles (%RC=81±4). Ceci est sans doute dû au fait qu'on n'a ici que 12 portraits visant chacun un seul des 12 sons. On peut en effet constater une tendance des choix des participants à s'aligner sur la diagonale du tableau. Pourtant, sur les 12\*12 choix possibles, seulement 4 d'entre eux ont été effectués par plus de 3/4 des sujets. Il s'agit des associations : VBR/P3-VBR, PNE/P3-PNE, CFT/P3-CFT et OBC/P3-CVC. On peut remarquer que seule la dernière de ces 4 associations correspond à un son qui n'est pas attendu, le son OBC comme hybride de clavecin et de hautbois étant le plus proche du son CVC, imitant un son de clavecin. Ces associations montrent donc la pertinence des portraits correspondants pour reconnaître les sons en question.



#### Différence entre musiciens et non-musiciens

Sur 4 choix effectués par plus de 3/4 des sujets, en moyenne, les réponses sont dues pour 53% aux musiciens. La participation des musiciens est particulièrement forte pour l'association VBR/P3-VBR (*attaque rapide, clair et doux*), mais aussi pour celle CFT/P3-CFT (*long, diffus et attaque lente*). Par contre la contribution des musiciens et non musiciens est égale pour l'association PNE/P3-PNE (*attaque rapide, net et fort*) et celle des non musiciens est plus forte pour l'association OBC/P3-CVC (*pincé, aigu et métallique*). Aucune réponse n'est due à 3/4 des sujets musiciens aux non musiciens mais pas à 3/4 de l'ensemble des sujets.

#### Analyse de la diagonale

Compte tenu du caractère particulier de nos attentes en matière d'association des sons à chacun de ces portraits, nous avons ici analysé la diagonale de la matrice de choix de chacun des sujets. Pour l'ensemble des sujets, on obtient en général 56% de sons attendus sur la première diagonale (60% chez les musiciens contre 48% chez les non musiciens). Le pourcentage de sujets ayant fait l'association attendue varie de 90% pour le son PNE à 15% pour le SNO, mais est supérieur à 50% pour 8 des 12 portraits. On retrouve les associations visant les sons PNE, VBR et CFT parmi celles réalisées par le plus grand nombre de sujets (respectivement 90, 85 et 80% des sujets). Les portraits visant les sons CVC, TPT, OBC, CRA et CLA étant associés aux sons attendus par respectivement 65, 65, 60, 55 et 50% des sujets. Seuls les portraits visant le PNO, le VBN, le TBN et le SNO ne mènent à l'association attendue que pour respectivement 45, 35, 30 et 15% des sujets. Il semble donc que ces portraits construits à partir de 3 adjectifs peuvent toucher leur cible, même si c'est effectivement le cas de très peu d'entre eux. On observe en effet le cas d'un portrait qui touche une autre cible, même si c'est le son le plus proche de la cible, et beaucoup ne touchent pas leur but...

### II.3. Portraits décrivant une source sonore possible

Nous allons retrouver ici les différents sous-groupes de portraits décrivant une source sonore possible des sons. L'ensemble des réponses donnant les nombres de sujets ayant associé chacun des sons aux portraits proposés ainsi que les pourcentages de réponses dues aux musiciens, se trouve dans les tableaux des annexes A.8.32 à A.8.55.

#### **II.3.1. Noms d'action**

Pour ce qui est des noms de modes d'excitation, on constate qu'un seul 2 n'est pas significatif. Globalement donc, et à part pour le portrait *on tire*, la répartition des sons sur chacun des portraits de noms de modes d'excitation ne se fait pas au hasard. Pourtant, on peut noter que les portraits *on gratte* et *on froisse* semblent moins spécifiques que les autres et que *on souffle* et *on frotte*, qui aboutissent à des 2 > 100, semblent par contre très spécifiques.

Sur les 12\*12 choix possibles, 11 d'entre eux (soit 8%) ont été effectués par plus de 3/4 des sujets. En effet, 5 portraits sont associés à un son par plus de 3/4 des sujets : *on frotte*, *on glisse* et *on effleure* sont associés à la CFT, *on frappe* et *on tape* au VBR. Trois autres portraits sont associés à 2 sons différents par plus de 3/4 des sujets : *on souffle* est associé au TBN et à la CNT, *on percute* au VBR et au PNE, *on pince* au CVC et au OBC. Ces associations sont dues en moyenne seulement pour 51% aux musiciens, ceux-ci étant plus nombreux (56%) à associer *on percute* au PNE. Tous les non musiciens ont par contre associé *on frotte* au son CFT. De plus les non musiciens sont plus de 3/4 à associer *on tape* avec le PNO, les musiciens associant pour plus de 3/4 d'entre eux *on frappe* avec les sons PNE et VBN et à l'unanimité *on effleure* avec la CFT.

Tableau 8.8. Timbres correspondant à chacun des portraits de noms d'action, pourcentage de sujets et 2 des répartitions sur chaque portrait.

Portrait	2	p	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
on souffle	101	°	**	***	*	***	.	.	.	.	.	.	.	.
on frotte	113	°	.	.	.	***	.	.	.	.	.	.	*	**
on glisse	54	°	.	*	.	***	.	.	.	.	.	.	.	**
on racle	42	°	.	.	.	**	.	.	.	.	.	.	.	.
on gratte	28	*	.	.	.	*	.	.	.	.	.	*	.	*
on tire	17	ns	.	.	.	*	.	.	.	.	.	*	.	.
on effleure	86	°	.	.	.	***	.	.	.	.	.	.	*	*
on froisse	30	*	.	*	.	*	.	.	.	.	.	.	*	.
on frappe	76	°	.	.	.	.	***	**	**	.	.	*	**	.
on percute	91	°	.	.	.	.	***	***	**	**	.	*	*	.
on tape	88	°	.	.	.	.	***	**	**	.	.	.	**	.
on pince	96	°	.	.	.	.	.	*	*	.	***	***	.	.

avec °: p<.001; \*: p<.01, † : p<.05, ns : non significatif, et avec \*\*\* si plus de 3/4, \*\*: plus de 1/2 et \*: plus de 1/4 des sujets ("." de 1 à 4 sujets) ont choisi ce son comme correspondant au portrait considéré.

Ceci nous fait donc apparaître clairement 4 catégories de sons : les sons TPT, TBN, CRA et CNT, associés le plus souvent avec *on souffle* et qui sont tous 4 des imitations de sons d'instruments à vent ; le son CFT, associé principalement aux portraits *on frotte*, *on glisse*, *on effleure* et dans une moindre mesure *on racle*, *on gratte* et *on tire* ; les sons VBR, PNE, PNO (et dans un moindre mesure VBN), associés aux portraits *on frappe*, *on percute* et *on tape*, qui sont pour les 3 premiers des imitations de sons d'instruments frappés ; enfin les sons CVC et OBC, associés aux portraits *on pince* et qui imitent des sons de corde pincée (le OBC étant voulu comme mélange de clavecin et de hautbois présente l'enveloppe temporelle du CVC). Reste le son SNO, jamais associé de manière massive à un de ces portraits de modes de production, mais pour lequel on peut tout de même observer une tendance à être considéré comme associé à *on frotte* et *on glisse*.

On observe donc que globalement, les modes de production imités lors de la synthèse de chacun de ces sons ont bien été retrouvés par les participants, qui font notamment la différence entre cordes pincées et frappées ainsi qu'entre les soufflés et les frottés. Décrire les sons sous forme de modes de production semble donc très efficace.

### II.3.2. Noms de matériaux

La plupart des portraits de noms de matériaux se répartissent sur les 12 sons de manière non aléatoire. C'est particulièrement le cas du portrait *du verre* qui présente la plus forte valeur de  $\chi^2$ . En fait, la répartition ne semble se faire au hasard que pour le portrait *un matériau naturel*, et semble un peu moins spécifique pour le portrait *un matériau synthétique*.

Tableau 8.9. Timbres correspondant à chacun des portraits de noms de matériaux : pourcentage de sujets et  $\chi^2$  des répartitions sur chaque portrait.

Portrait	$\chi^2$	p	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
du cuivre	73	°	*	***	.	.	.	.	.	.	.	.	*	.
du métal	40	°	.	.	.	.	.	*	*	*	**	*	*	.
du bois	55	°	.	.	.	**	.	.	.	.	.	.	*	.
du vent	35	°	*	*	*	**	*	.	.	.	.	.	.	*
de l'air	57	°	*	**	.	**	.	.	.	.	.	.	.	.
de l'eau	63	°	.	.	.	.	.	**	.	.	.	.	*	.
du cristal	93	°	.	.	.	.	.	***	.	.	.	.	.	.
du verre	137	°	.	.	.	.	.	***	.	.	.	.	.	.
de la tôle	37	°	.	.	.	.	.	*	.	*	.	.	*	.
une corde en boyau	37	°	.	.	.	.	*	.	.	**	.	*	.	.
une corde métallique	80	°	.	.	.	.	*	.	**	*	***	*	.	.
une peau	95	°	.	.	.	.	.	.	.	*	.	.	**	.
de la peau et du métal	57	°	.	.	.	.	.	.	.	**	.	.	**	.
un matériau naturel	17	ns	.	.	.	*	*	**	*	.	.	.	*	*
un mat. synthétique	20	†	*	.	*	.	*	.	.	.	.	**	.	*
un matériau électrique	33	°	**	.	*	.	.	.	.	.	*	**	.	*
un mat. électronique	37	°	**	.	**	*	*	.	.	.	.	**	*	.

avec °:  $p < .001$ ; \*:  $p < .01$ , † :  $p < .05$ , ns : non significatif, et avec \*\*\* si plus de 3/4, \*\*: plus de 1/2 et \*: plus de 1/4 des sujets (". " de 1 à 4 sujets) ont choisi ce son comme correspondant au portrait considéré.

Pourtant, sur l'ensemble de ces 17 portraits, seuls 4 d'entre eux sont associés par plus de 3/4 des sujets à un des 12 sons. Il s'agit des portraits *du cuivre*, associé au son TBN, *du cristal* et *du verre*, au son VBR et *une corde métallique*, associé au son CVC. Ces associations ne sont pas particulièrement dues aux participants musiciens (en moyenne ils sont responsables de 51% de ces associations). En outre, plus de 3/4 des non musiciens associent *un matériau électronique* au son CRA et *du métal* au son CVC. Ce type de portrait présente les différences inter-sujets les plus faibles, en particulier en ce qui concerne la compétence musicale des participants. On retrouve dans ce groupe un des portraits les plus cohérents de l'ensemble : *du verre*.

### II.3.3. Noms de sources sonores

Toutes les répartitions des associations des sujets sur les 12 sons proposés sont faites de manière non aléatoire. On observe même 4 portraits pour lesquels le  $\chi^2$  est supérieur à 100, qui sont donc très spécifiques de quelques sons. Il s'agit de *on frappe du verre* ( $\chi^2=161$  : le maximum pour l'ensemble des 201 portraits testés), mais aussi *on souffle dans du bois*, *on souffle dans du cuivre* et *on percute une peau*.

Un seul portrait semble un peu plus dispersé sur les 12 sons, il s'agit de *une pièce de monnaie sur des cordes*. Pourtant, sur les 26 portraits de ce sous-groupe, seuls 8 d'entre eux sont associés à un même son par plus de 3/4 des sujets. Ainsi, *on souffle dans du cuivre* a été unanimement associé au son TBN (mais aussi au son TPT par plus de 50% des sujets), *on souffle dans du bois* et *on souffle dans un tube* sont associés à la CNT, même si le premier de ces deux portraits est plus spécifique de la CNT, *on frappe du verre* est associé au VBR, *on pince une corde* au CVC, *on étouffe le son*, *on percute une peau* et *double percussion sur une peau* au VBN.

Tableau 8.10. Timbres correspondant à chacun des portraits de noms de sources sonores :  
pourcentage de sujets et 2 des répartitions sur chaque portrait.

Portrait (partiels)	2	p	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
on souffle / cuivre	123	°	**	***	.	.	.	.	.	.	.	.	.	.
on souffle / bois	127	°	.	.	.	***	.	.	.	.	.	.	.	.
une anche vibre	33	°	*	.	*	*	.	.	.	.	.	.	.	.
on souffle / tube	82	°	*	**	*	***	.	.	.	.	.	.	.	.
du vent dans un tuyau	62	°	*	**	.	**	.	.	.	.	.	.	.	*
un frottement d'air	37	°	.	*	.	.	*	.	.	.	.	.	.	.
on frotte une corde	64	°	.	.	.	.	**	.	.	.	.	.	.	*
effleure / corde / métal	57	°	.	.	.	.	*	.	.	.	.	.	.	*
on frotte du verre	52	°	.	.	.	.	**	.	.	.	.	.	.	*
on frappe du bois	83	°	.	.	.	.	.	**	.	.	.	.	*	.
on frappe du verre	161	°	.	.	.	.	.	***	.	.	.	.	.	.
on casse du verre	34	°	.	.	.	.	.	*	.	.	.	*	.	.
corde vibre / verre	45	°	.	.	.	.	.	**	.	.	*	.	*	.
on frappe une corde	86	°	.	.	.	.	.	.	**	**	*	.	.	.
on percute du métal	55	°	.	.	.	.	.	*	**	*	*	.	**	.
on tape sur une tôle	87	°	.	.	.	.	.	*	.	**	.	.	**	.
on pince une corde	62	°	.	.	.	.	.	.	*	*	***	**	.	.
corde pincée / plaque	66	°	.	.	.	.	.	.	.	.	**	**	.	.
on étouffe le son	35	°	.	**	.	*	*	.	.	**	.	.	***	*
corde enrobée / tissu	67	°	.	.	.	.	.	.	.	**	.	.	**	.
bouge / plaque métal.	39	°	.	.	.	.	*	.	.	.	.	.	*	.
on percute une peau	121	°	.	.	.	.	.	.	.	.	.	.	***	.
métal / résonance	47	°	.	.	.	.	.	*	.	.	*	*	**	.
double percu. / peau	93	°	.	.	.	.	.	.	.	.	.	.	***	.
pièce monnaie /cordes	26	*	.	.	.	.	.	.	.	.	*	*	.	.
orgue électronique	36	°	**	.	**	**	.	*	.	*	.	.	.	.

avec °: p<.001; \*: p<.01, † : p<.05, ns : non significatif, et avec \*\*\* si plus de 3/4, \*\*: plus de 1/2 et \*: plus de 1/4 des sujets ("." de 1 à 4 sujets) ont choisi ce son comme correspondant au portrait considéré.

En moyenne, ces associations sont dues un peu plus souvent aux non musiciens (pour 52%), principalement l'association *on étouffe le son* / VBN, due pour 67% aux non musiciens (qui font cette association à l'unanimité). On peut noter de plus, que plus de 3/4 des musiciens associent *on frotte une corde* avec le son CFT et qu'ils associent à l'unanimité *on frappe du verre* avec le son VBR. Les non musiciens associent de plus (pour plus de 3/4 d'entre eux) *du vent dans un tuyau* au son CNT, *on frappe du bois* aux deux sons PNE et PNO et *un son produit par une touche d'orgue électronique* au son CNT. On obtient donc des grandes différences entre les deux groupes de sujets, en particulier

pour le portrait *on étouffe le son*, qui est unanimement associé au son VBN par les non musiciens et réparti sur les 12 sons par les musiciens (voir tableaux A8.II.30 et A8.II.31).

#### II.3.4. Familles d'instruments

Aucun des 6 portraits de ce sous-groupe n'a été associé aux 12 sons en les répartissant au hasard. De plus le  $\chi^2$  correspondant au portrait *famille des cuivres* est supérieur à 100. Ce portrait est donc particulièrement spécifique de quelques sons.

Tableau 8.11. Timbres correspondant à chacun des portraits de noms de familles d'instruments : pourcentage de sujets et  $\chi^2$  des répartitions sur chaque portrait.

Famille des ...	$\chi^2$	p	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
Cuivres	107	°	***	***	.	.	.	.	.	.	.	.	*	.
Bois	91	°	.	.	**	***	.	.	.	.	.	.	.	.
Vent	96	°	**	***	**	***	.	.	.	.	.	.	.	.
Cordes	61	°	.	.	.	.	**	*	**	**	***	**	*	*
Percussions	77	°	.	.	.	.	.	***	*	.	.	*	*	.
Claviers	79	°	.	.	.	.	.	.	**	***	*	.	.	.

avec °: p<.001; \*: p<.01, † : p<.05, ns : non significatif, et avec \*\*\* si plus de 3/4, \*\*: plus de 1/2 et \* : plus de 1/4 des sujets ("." de 1 à 4 sujets) ont choisi ce son comme correspondant au portrait considéré.

On constate que tous les portraits de ce sous-groupe ont été associés par 3/4 des sujets à au moins un même son, voire plusieurs. Notamment, le portrait *famille des bois* a été associé au son CNT, *famille des percussions* au VBR, *famille des claviers* au PNO et *famille des cordes* principalement au CVC. On peut aussi noter que les portraits *famille des vents* et *famille des cuivres* ont été associés par plus de 3/4 des sujets chacun à deux sons différents qui ne sont pas les mêmes. *Famille des cuivres* ayant été associé quasi exclusivement aux sons TPT et TBN, *famille des vents* aux sons TBN et CNT (et seulement dans une moindre mesure aux sons TPT et CRA).

On peut remarquer que ces 6 portraits n'ont pas permis de classer les 3 sons hybrides, ce qui est normal, vu qu'ils n'appartenaient à aucune de ces familles *a priori*, mais aussi les sons CFT et PNE. En effet, les participants ont compris par *famille des cordes* essentiellement corde pincée (le son CFT ayant été associé à ce portrait par seulement 12 des 20 participants). Le son PNE apparaissant lui aussi dans les 2 portraits *famille des cordes* et *famille des claviers* mais à chaque fois pour moins de 3/4 des participants. L'ensemble de ces associations est plus souvent dû aux musiciens (pour 54%), ce qui est particulièrement le cas pour l'association *famille des vents* / TBN, effectuée à l'unanimité par les musiciens. Ils effectuent d'ailleurs 3 autres associations à l'unanimité : *famille des vents* / CNT, et *famille des cuivres* / TPT et TBN. De plus, ils sont les seuls à associer pour plus de 3/4 d'entre eux les sons TPT et CRA au portrait *famille des vents*, et les sons PNE, PNO et OBC au portrait *famille des cordes*. Les non musiciens sont par contre plus nombreux à associer le PNO au portrait *famille des claviers*.

De plus, si on considère les sons attendus pour chacun de ces portraits, (soit les sons CNT, CRA, TPT et TBN pour la famille des vents, CNT et CRA pour la famille des bois, TPT et TBN pour la famille des cuivres, CVC, PNE, PNO et CFT pour la famille des cordes et VBR pour la famille des

percussions), plus de 75% des participants associent au moins un des sons attendus à chacun de ces portraits. En fait, tous les participants associent au moins un des sons attendus aux portraits *vent*, *cuivre* et *corde*, et seulement 95 et 85% d'entre eux au moins un son attendu respectivement aux portraits *bois* et *claviers*. Si on ajoute dans les sons attendus pour le portrait *famille des percussions* les sons VBN, PNO et PNE, 100% des participants y associent au moins un des sons attendus. Ces portraits sont donc globalement efficaces pour désigner les sons, même si les musiciens associent plus systématiquement tous les sons possibles à chacun des portraits.

Tableau 8.12. Timbres correspondant à chacun des portraits de noms d'instruments visés par le concepteur : pourcentage de sujets et 2 des répartitions sur chaque portrait.

Portrait	2	p	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
trompette	103	°	**	**	.	.	.	.	.	.	.	.	.	.
trombone	109	°	*	***	.	.	.	.	.	.	.	.	.	.
trompe	81	°	*	**	.	.	.	.	.	.	.	.	.	.
cor	83	°	*	**	.	.	.	.	.	.	.	.	.	.
tuba	49	°	.	**	.	*	.	.	.	.	.	.	.	.
cornet à piston	79	°	**	*	.	.	.	.	.	.	.	.	.	.
corne de brume	44	°	.	**	.	.	.	.	.	.	.	.	.	.
hautbois	71	°	.	.	**	*	.	.	.	.	.	.	.	.
cor anglais	56	°	*	**	*	.	.	.	.	.	.	.	.	.
clarinette	129	°	.	.	.	***	.	.	.	.	.	.	.	.
flûte	137	°	.	.	.	***	.	.	.	.	.	.	.	.
pipeau	59	°	.	.	.	**	.	.	.	.	.	.	.	.
accordéon	40	°	.	.	*	.	*	.	.	.	.	.	.	.
harmonica	31	°	.	.	*	.	.	.	.	.	.	.	.	.
harmonium	47	°	.	.	*	*	.	.	.	.	.	.	.	*
orgue	39	°	*	.	*	*	.	.	.	.	.	.	.	.
synthétiseur	10	ns	**	.	*	*	*	.	*	*	.	*	.	*
violon	148	°	.	.	.	.	***	.	.	.	.	.	.	.
violoncelle	148	°	.	.	.	.	***	.	.	.	.	.	.	.
alto	116	°	.	.	.	.	***	.	.	.	.	.	.	.
xylophone	208	°	.	.	.	.	.	***	.	.	.	.	.	.
timbale	35	°	.	.	.	.	.	*	.	.	.	.	*	.
marimba	64	°	.	.	.	.	.	**	.	.	.	.	.	.
métalophone	86	°	.	.	.	.	.	**	.	.	.	*	.	.
gong	43	°	.	.	.	.	.	*	.	*	.	.	.	.
cloche	134	°	.	.	.	.	.	***	.	*	.	.	.	.
tambour	65	°	.	.	.	.	.	.	.	.	.	.	**	.
woodbloc	43	°	.	.	.	.	.	**	.	.	.	.	.	.
piano	108	°	.	.	.	.	.	.	**	**	.	.	.	.
pianoforte	99	°	.	.	.	.	.	.	**	**	.	.	.	.
clavecin	85	°	.	.	.	.	.	.	*	.	**	.	.	.
guitare	63	°	.	.	.	.	.	.	*	*	**	.	.	.
mandoline	57	°	.	.	.	.	.	.	.	.	*	*	.	.
harpe	45	°	.	.	.	.	.	*	.	*	*	.	.	.
luth	37	°	.	.	.	.	.	.	.	.	*	*	.	.
cithare	89	°	.	.	.	.	.	.	.	.	**	**	.	.
banjo	69	°	.	.	.	.	.	.	*	.	*	**	.	.

avec °: p<.001; \*: p<.01, † : p<.05, ns : non significatif, et avec \*\*\* si plus de 3/4, \*\*: plus de 1/2 et \*: plus de 1/4 des sujets ("." de 1 à 4 sujets) ont choisi ce son comme correspondant au portrait considéré.

### II.2.5. Noms d'instruments

Pour ce qui est des 37 portraits de noms d'instruments, un seul se répartit au hasard sur l'ensemble des 12 sons proposés, il s'agit du portrait *synthétiseur*, ce qui peut se comprendre vu que l'ensemble des 12 sons a été produit par un TX802. Quelques participants ont sélectionné les 12 sons pour ce portrait.

On observe aussi 10 valeurs de  $\chi^2$  supérieures à 100, ce qui veut dire que les portraits correspondants sont particulièrement spécifiques de quelques sons. On retrouve ceux qui ont été notés comme ayant les plus forts %RC (voir §I.), c'est à dire les portraits de *xylophone*, *violon*, *violoncelle*, *flûte*, *cloche*, mais aussi ceux de *clarinette*, *alto*, *trombone*, *piano* et *trompette*.

Pourtant, sur ces 37 portraits, seuls 8 d'entre eux ont été associés par plus de 3/4 des sujets à un même son, et ce, bien que de manière générale, les choix des participants ne s'éloignent pas beaucoup de la diagonale. Ainsi, on constate que le portrait de *trombone* est associé au son TBN, ceux de *clarinette* et de *flûte* au son CNT, ceux de *violon*, *violoncelle* et *alto* au son CFT, ceux de *cloche* et de *xylophone* sont associés au son VBR. Il faut noter que l'association *xylophone* / VBR est faite à l'unanimité des participants, un seul d'entre eux ayant choisi un deuxième son à associer à ce portrait, le son PNO.

L'ensemble de ces associations est dû un peu plus souvent aux musiciens (pour 53%), particulièrement pour les associations *trombone* / TBN et *alto* / CFT, dues respectivement pour 67 et 56% aux musiciens. Par contre, les non musiciens sont un peu plus nombreux à faire l'association *violon*/CFT (un musicien manque à l'appel alors que les non musiciens font l'association à l'unanimité). De plus, les musiciens associent pour plus de 3/4 d'entre eux le portrait de *trompette* au son TPT, *cor* au TBN, *métalophone* au VBR, *clavecin* et *cithare* au CVC alors que les non musiciens n'ajoutent que l'association *piano* / PNE (effectuée par plus de 3/4 d'entre eux). On constate donc des différences de comportement en fonction des compétences musicales des participants.

De plus, si on regarde le timbre dominant pour chaque portrait (celui qui est le plus souvent associé au portrait considéré), on obtient très souvent le timbre visé par ce portrait. C'est en particulier le cas pour 21 des 37 portraits qui présentent clairement un timbre dominant : TPT pour *trompette* et *cornet à piston*, TBN pour *trombone*, *trompe*, *cor*, *tuba*, *corne de brume* et *cor anglais* (bien que ce dernier soit plus réparti sur TPT et CRA). C'est aussi le cas du son CRA pour *hautbois*, de la CNT pour *clarinette*, *flûte* et *pipeau*, de la CFT pour *violon*, *violoncelle* et *alto*, du VBR pour *xylophone*, *marimba*, *cloche* et *woodblock* (qui pourtant ne fait pas du tout le son du vibraphone !), du PNE et du PNO pour *piano* et *pianoforte*, et du CVC pour *clavecin*. Les trois sons hybrides ne sont pas dominants sur un de ces portraits, à part le son VBN fortement associé au portrait *tambour*.

Les autres portraits de ce groupe ont été associés à plus de sons différents, notamment *accordéon*, *harmonica*, *harmonium* et *orgue*, répartis surtout sur les imitations de sons d'instruments à vent et le

son SNO, les portraits *timbale*, *métalophone*, et *gong*, répartis sur les imitations de sons frappés et les hybrides, et les portraits *guitare*, *mandoline*, *harpe*, *luth*, *cithare* et *banjo*, principalement centrés sur le CVC et l'OBC même s'ils se répartissent parfois aussi sur les deux imitations de sons de piano.

Ce type de portrait est donc assez efficace pour désigner les sons de notre ensemble. Il faut tout de même rappeler que ces 37 noms d'instruments ont été produits spontanément pour décrire les 12 sons, mais que parfois ils sont assez éloignés d'une "réalité", définie comme un instrument qu'un de ces sons pouvait imiter de manière vraisemblable. Par exemple, le son émis par un woodblock ne peut "objectivement" se rapprocher d'aucun des 12 sons de notre ensemble.

### II.2.6. Noms d'instruments visés par le concepteur

Ce sous-ensemble reprend une partie des noms d'instruments du groupe précédent. Il n'est donc pas étonnant de ne trouver aucun des portraits répartis de manière aléatoire sur l'ensemble des 12 sons. Pourtant, les portraits de mélange de 2 sons présentent les 2 les plus faibles.

Tableau 8.13. Timbres correspondant à chacun des portraits de noms d'instruments visés par le concepteur : pourcentage de sujets et 2 des répartition sur chaque portrait.

Portrait	2	p	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
trompette	148	°	***	.	.	.	.	.	.	.	.	.	.	.
trombone	118	°	.	***	.	.	.	.	.	.	.	.	.	.
cor anglais	49	°	.	*	*	.	.	.	.	.	.	.	.	.
clarinette	141	°	.	.	.	***	.	.	.	.	.	.	.	.
corde frottée	102	°	.	.	.	.	**	.	.	.	.	.	.	.
vibraphone	118	°	.	.	.	.	.	***	.	.	.	.	.	.
piano	89	°	.	.	.	.	.	.	**	*	.	.	.	.
piano électrique	59	°	.	.	.	.	.	.	**	.	.	.	.	.
clavecin	102	°	.	.	.	.	.	.	.	.	***	.	.	.
hautbois/clavecin	34	°	.	.	.	.	.	.	.	.	.	*	.	*
vibraphone/trombone	47	°	.	*	.	.	.	.	.	.	.	.	*	.
corde frottée/ piano	50	°	.	.	.	.	.	.	.	.	.	.	*	*

avec °: p<.001; \*: p<.01, † : p<.05, ns : non significatif, et avec \*\*\* si plus de 3/4, \*\*: plus de 1/2 et \*: plus de 1/4 des sujets ("." de 1 à 4 sujets) ont choisi ce son comme correspondant au portrait considéré.

On ne peut que constater l'alignement des choix des participants le long de la diagonale du tableau précédent, les sons visés par chacun de ces 12 portraits, ou à défaut les sons les plus proches, sont très souvent choisis. Pourtant sur les 12 portraits, seuls 5 d'entre eux ont été associés au même son par plus de 3/4 des participants. Ainsi, comme nous l'attendions, le portrait *trompette* a été associé au son TPT, celui de *trombone* au TBN, *clarinette* à la CNT, *vibraphone* au VBR et *clavecin* au CVC. Ces associations sont plus souvent dues aux musiciens (pour 59% d'entre elles en moyenne), ce qui est d'autant plus vrai pour les associations *vibraphone/VBR*, *clavecin/CVC* et *trombone/TBN*, effectuées à l'unanimité par les musiciens qui sont donc responsables pour respectivement 67, 67 et 59% de ces associations. Les autres portraits de ce groupe sont moins clairement associés à un des sons de l'ensemble, même si les portraits *piano* et *piano échantillonnés* présentent tous les deux le son PNE comme timbre dominant (choisi par plus de 50% des sujets), et même si les 3 portraits de mélanges sont globalement plus souvent associés à un des sons hybrides qu'aux autres sons. De plus,



plus de 3/4 des musiciens font l'association *piano électrique*/PNE, les non musiciens répartissant plus leurs réponses sur l'ensemble des sons.

Si les sons imitant des instruments traditionnels peuvent être reconnus en général à partir des noms de ces instruments (sauf dans le cas du cor anglais, qui est un nom relativement peu connu chez les non musiciens), les noms de mélange d'instruments, décrivant la synthèse effectuée pour construire les sons hybrides, ne permettent pas de les reconnaître.

#### II.4. Comparaison de portraits proches dans des contextes différents

Certains portraits ou parties de portraits ont été testés dans plusieurs contextes différents. Il peut donc être intéressant de comparer la façon dont ils ont été associés aux sons. Ainsi, par exemple, si un portrait comportant trois qualités simples a été spécifiquement associé à un timbre, il peut être intéressant de vérifier si cette association est due vraiment à l'association des 3 adjectifs ou si seulement deux ou un seul de ces adjectifs est aussi efficace. De même dans le domaine des sources sonores, peut être que pour certains sons, le nom de la matière seule, ou le nom de l'action seule peut être aussi efficace que la description complète de l'action sur la matière.

##### **II.4.1. Qualités triples / doubles / simples**

Si on ne considère que les associations effectuées par plus de 3/4 des participants, on peut remarquer que les 4 portraits de 3 adjectifs qui ont été associés au même son par plus de 3/4 des participants ont souvent vu déjà un ou 2 des adjectifs les constituant déjà associés à ce même son. Ainsi, pour le portrait P3-CFT (*long, diffus et attaque lente*), associé au son CFT par 80% des sujets, on constate que déjà *diffus* et *attaquelente* présentés seuls avaient été associés principalement à ce son (par 75% des sujets). Par contre *long* seul avait été associé principalement au PNO (et seulement par 65% des sujets à la CFT) et *long et diffus* n'a été associé à la CFT que par 60% des sujets (même si c'est le timbre dominant de ce portrait).

De même, pour le portrait P3-VBR (*attaque rapide, clair et doux*), associé au son VBR par 85% des sujets, on constate que déjà *attaquerapide* et *clair*, présentés seuls ainsi que *clair et doux* avaient été principalement associés à ce même son. Par contre, *doux* seul avait été associé par 60% des sujets au son CNT et seulement par 40% des sujets au son VBR. De plus compte tenu d'une part du fait que *clair et rond* est principalement associé au VBR mais pas *rond* (même si le VBR fait partie des 3 sons associés le plus souvent par 70% des sujets à cet adjectif avec les sons TBN et CNT), et d'autre part que *attaque rapide* a été aussi associé par plus de 3/4 des sujets à 3 autres sons, on peut considérer que c'est surtout *clair* qui est spécifique du VBR dans notre contexte.

De même, pour le portrait P3-PNE (*attaque rapide, net et fort*) associé au son PNE par 90% des sujets, on constate que déjà *attaquerapide* et *fort*, présentés seuls, ainsi que *net et franc* avaient été principalement associés à ce même son. Par contre, *net* présenté seul a été associé principalement à 2

autres sons (VBR et CVC) même si le son PNE est dominant (avec 70% des sujets). *Pur et net* étant principalement associé au VBR, mais aussi au CVC et en troisième position au PNE par 45% des sujets.

Enfin, pour le portrait P3-CVC (*pincé, aigu et métallique*) associé au son OBC par 75% des sujets (mais aussi au CVC par 65% des sujets), on constate que déjà *aigu* présenté seul et *aigu et métallique* avaient été principalement associés à ce même son (et respectivement par 70 et 60% des sujets au son CVC). De plus, *pincé* et *métallique* présentés seuls sont principalement associés au son CVC (et respectivement par 60 et 65% des sujets au son OBC) et sur les 5 autres portraits composés de 2 adjectifs et contenant soit *pincé*, soit *métallique*, 2 d'entre eux sont principalement associés aux deux sons OBC et CVC (*métallique et froid* associé par 65% des sujets au OBC et *pincé et résonnant* associés respectivement par 70 et 65% des sujets aux sons OBC et de CVC). On constate donc que même si les portraits constitués de 3 adjectifs semblent plus efficaces pour atteindre un son visé, quand ils y arrivent, c'est souvent qu'un de leur composants suffit à distinguer le son voulu de l'ensemble des 12 sons du contexte.

Pour ce qui est des 5 portraits de 2 adjectifs qui ont été associés par plus de 3/4 des participants à un même son, on observe le même type de phénomène, bien que plus nuancé. En effet, nous avons déjà remarqué que si *aigu et métallique* est principalement associé au son OBC, c'est déjà le cas de *aigu* seul. *Métallique* seul est principalement associé au son CVC qui est le plus proche de lui. Ces deux sons sont d'ailleurs difficilement distinguables à l'aide de ces adjectifs. Par contre, nous avons vu que si *clair et doux* est principalement associé au son VBR, c'est déjà le cas de *clair* seul et que *doux* est déjà associé au son VBR par seulement 40% des sujets. *Doux* seul renvoie plutôt au son CNT. Ça n'est qu'associé avec *clair* qu'il désigne clairement le son VBR. Dans *clair et rond*, associé principalement au son VBR, il semble que ce soit la présence de *clair* qui spécifie le fait que *rond* désigne le VBR. En effet, *rond* présenté seul désigne aussi bien le son VBR que TBN et CNT. De même, si *pur et net* est associé principalement au son VBR, *pur* seul l'est déjà, même si *net* se répartit sur les 3 sons VBR, PNE et CVC. Enfin *net et franc* sont associés principalement au son PNE même si aucun des 2 adjectifs le composant présentés seuls ne désigne clairement ce son. En effet, *net* seul désigne aussi bien les sons VBR, PNE et CVC. *Franc* n'a pas été testé seul.

Il semble donc qu'en ce qui concerne les associations de 2 adjectifs, elles sont plus spécifiques d'un des sons, souvent du fait qu'un des deux adjectifs composant le portrait est déjà très spécifique de ce même son, l'autre pouvant désigner plusieurs sons.

#### II.4.2. Sources / action + matériau

De la même façon que précédemment, la plupart des noms de sources sonores ont été construits à partir des portraits de noms de modes d'excitation et de ceux de noms de matériaux. Ainsi par exemple, le portrait *on souffle dans du cuivre*, est constitué des deux autres : *on souffle* et *du cuivre*. Il peut donc être intéressant de comparer la spécificité de chacun des morceaux seuls et de l'ensemble. Nous allons considérer chacun des 8 portraits de noms de sources sonores qui ont été associés par

plus de 3/4 des sujets à un même son. Ainsi, pour le portrait *on souffle dans du cuivre* associé au son TBN par 100% des sujets, on constate que déjà *du cuivre* présenté seul était principalement associé au même son et *on souffle* seul était associé principalement à deux sons dont celui de TBN. Le nom du matériau semble donc ici plus spécifique que le nom du mode d'excitation.

On peut faire exactement la même observation avec le portrait *on souffle dans du bois* : il est principalement associé au son CNT, ce qui est déjà le cas de *du bois*, mais pas de *on souffle*, qui se partage entre les sons TBN et CNT. Nous ne pouvons faire la même observation avec *on souffle dans un tube* puisque un tube n'a pas été testé comme portrait de matériau. On peut faire le même genre de remarque pour le portrait *on pince une corde*, principalement associé au son CVC. On constate en effet, que *on pince* présenté seul renvoyait principalement aux deux sons CVC et OBC, une corde seule n'ayant pas été présentée mais *une corde métallique* renvoie déjà principalement au son CVC.

Pour le portrait *on percute une peau*, associé principalement au son VBN, on constate que *on percute* désignait plutôt les deux sons VBR et PNE, et que *une peau*, désignait déjà principalement le son VBN (pour 70% des sujets). Ici aussi on retrouve l'idée que le matériau renvoie plus spécifiquement au son que le mode d'excitation.

Par contre, pour le portrait *on frappe du verre*, principalement associé au son VBR, on constate que déjà *on frappe* et *du verre* présentés seuls renvoyaient sans équivoque principalement à ce même son. Enfin, pour le portrait *on étouffe le son*, principalement associé au son VBN, on étouffe n'a pas été présenté seul comme mode de production (par contre l'adjectif *étouffé* renvoyait déjà au son VBN). De même, pour le portrait *double percussion sur une peau*, principalement associé au son VBN, on constate que *une peau* lui était déjà principalement associé. Par contre, l'adjectif *double* se répartissait sur beaucoup de sons différents même si le son VBN était parmi les dominants avec 55% des sujets. Il semble donc que parmi les portraits de sources sonores, la partie désignant le matériau excité soit plus spécifique du son visé que la partie décrivant le mode d'excitation de ce matériau. Pourtant, les noms de mode d'excitation qui renvoyaient à la CFT lui étaient très spécifiques : aucun autre son de l'ensemble n'imitait une production par frottement, à part l'hybride de SNO, conçu comme mélange de corde frottée et de piano, mais qui s'est avéré très synthétique.

#### II.4.3. Matériau / Famille

Nous avons souligné chapitre 4 la difficulté que nous avons rencontrée lors du codage des verbalisations libres, de distinguer parfois dans les locutions produites par les participants celles qui désignaient un nom de matériau de celles qui désignaient une famille d'instruments (par exemple, quand un locuteur décrit un son par *corde*, ou *cuivre*). Nous avons donc souhaité comparer ces deux types de portraits.

Tableau 8.14. Nombre de sujets ayant associé un son à un portrait

Tous	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO
du cuivre	7	<b>16</b>	0	1	0	3	0	2	2	2	5	0
Famille des cuivre	<b>19</b>	<b>18</b>	4	3	2	0	0	0	0	1	6	3
du bois	1	2	3	13	0	3	1	0	1	2	5	0
Famille des bois	3	2	11	<b>17</b>	1	1	0	0	0	1	3	1
du vent	6	8	5	11	8	1	0	2	1	1	1	5
Famille des vents	13	<b>17</b>	14	<b>18</b>	2	0	0	0	0	1	4	3
une corde en boyau	0	0	0	0	5	2	4	10	1	7	4	4
une corde métallique	0	0	0	0	6	1	13	6	<b>17</b>	9	2	1
Famille des cordes	0	0	0	0	12	5	14	13	<b>16</b>	13	6	6

En gras, les associations effectuées par plus de 3/4 des sujets.

Ainsi, *du cuivre*, considéré comme matériau semble plus spécifique du son TBN que le portrait *famille des cuivres* qui désigne aussi bien le son TPT que TBN. De même, *du bois* ne présente que le son CNT comme portrait dominant (choisi par plus de 50% des sujets), alors que celui de *famille des bois*, s'il renvoie principalement à ce même son CNT, est aussi associé par plus de 50% des sujets au son CRA. Par contre, le portrait *du vent*, est associé à un plus grand nombre de sons (notamment le son CFT) alors que *famille des vents* fait bien ressortir les sons TBN et CNT, les deux seuls autres sons à être associés à ce portrait par plus de 50% des sujets étant TPT et CRA. Pour ce qui est du portrait *famille des cordes*, il est associé principalement au son CVC, même si les choix des sujets se répartissent principalement sur 5 sons différents. On ne peut le comparer directement à un portrait *une corde*, puisque la matière qui compose la corde a été précisée, *une corde métallique* renvoyant au son CVC de façon plus spécifique que *famille des cordes*. Le portrait *corde en boyau* semble plus flou, bien que 50% des sujets y associent le son PNO.

On peut remarquer que globalement les portraits de noms de matériaux semblent plus spécifiques d'un seul son que ceux des noms de familles d'instruments utilisant ces mêmes noms de matériaux.

#### II.4.4. Instruments / instruments visés

Certains noms d'instruments se retrouvent cités à la fois dans le sous-groupe des noms d'instruments et dans celui des noms d'instruments visés par le concepteur. Le comportement des sujets dans ces deux contextes peut donc être intéressant à observer. Ainsi, si on retrouve dans les deux sous-groupes les associations *trombone* / TBN et *clarinette* / CNT effectuées par plus de 3/4 des sujets, il est étonnant de remarquer que *trompette*, présent aussi dans les deux sous-groupes n'a été clairement associé au son TPT que dans le contexte des 12 noms d'instruments visés. Dans le contexte des 37 noms d'instruments de musique, il a été associé principalement aux deux sons TPT et TBN (respectivement par 65 et 60% des sujets). Il semble que le nom de trompette ait parfois été considéré par les participants comme un terme générique pouvant désigner les cuivres. De même, le nom de clavecin, présent dans les deux contextes n'a été clairement associé au son CVC que dans celui des noms d'instruments visés, même si c'est le timbre dominant de ce portrait (avec 70% des sujets et 100% des musiciens) dans le contexte des noms d'instruments.

Par contre, les comportements des participants concernant le portrait *coranglais*, présent lui aussi dans les deux contextes sont plus similaires. Dans le contexte des 37 noms d'instruments de musique, les sujets ont réparti leurs choix sur les trois sons TPT, TBN et CRA, principalement le son TBN pour 50% d'entre eux. Dans le contexte des 12 noms d'instruments visés, les sujets ont réparti leurs choix sur ces 3 mêmes sons, TBN et CRA étant associés à égalité par 35% des sujets. Les deux autres sons imitant des instruments traditionnels étaient désignés par des portraits différents dans les deux sous-groupes. Les comportements des sujets sont pourtant comparables : le son CFT a été choisi par plus de 3/4 des sujets pour être associé aux portraits de *violon*, *violoncelle* et *alto*, dans le contexte des noms d'instruments et au portrait de *cordefrottée* dans celui des noms de fichiers. De même, le son VBR a été associé par plus de 3/4 des sujets au portrait de *xylophone* dans le cadre des noms d'instruments et au *vibraphone* dans celui des noms d'instruments visés par le concepteur.

Les différences observées entre ces deux contextes, sur les mêmes participants, sont sans doute dues aux différences de nombre de portraits proposés dans chacun des contextes : 37 noms d'instruments contre seulement 12 noms de fichiers, ce qui a aidé les participants à être plus spécifiques dans le deuxième contexte.

#### II.4.5. Source sonore / qualité sonore

Parfois aussi quelques adjectifs de qualité renvoyaient à des matériaux ou des modes d'excitation possibles. Nous allons donc comparer les utilisations dans ces différents contextes. Le portrait *on pince* semble plus focalisé sur les deux sons CVC et OBC que *un son pincé*, réparti sur plus de sons différents et associé au OBC par moins de 3/4 des sujets. Par contre, les deux portraits *on souffle* et *un son soufflé* semblent associés de manière assez comparable aux sons proposés. Par contre, *un son métallique* est associé à beaucoup plus de sons différents que *du métal*, même si le timbre dominant est le même dans les deux cas : le CVC.

Tableau 8.15. Nombre de sujets non musiciens ayant associé un son à un portrait

Portraits	TPT	TBN	CRA	CNT	CFT	VBR	PNÉ	PNO	CVC	OBC	VBN	SNO
un son pincé	1	1	2	1	2	7	9	4	<b>16</b>	12	2	3
on pince	0	0	0	0	0	6	9	4	<b>18</b>	<b>15</b>	3	0
un son soufflé	13	<b>17</b>	8	<b>15</b>	6	0	0	1	0	0	6	4
on souffle	13	<b>16</b>	5	<b>16</b>	2	0	0	0	0	0	1	2
un son métallique	2	2	3	0	3	11	11	2	<b>16</b>	13	2	4
du métal	1	4	0	0	2	9	8	7	14	5	9	3
un son résonnant	1	3	3	1	8	12	2	12	3	5	14	5
morceau métal mis résonance	1	0	0	0	1	7	4	4	5	8	13	2
un son cuivré	11	<b>17</b>	3	1	4	1	0	0	1	1	8	0
du cuivre	7	<b>16</b>	0	1	0	3	0	2	2	2	5	0
un son double	1	4	1	1	5	3	3	4	5	9	11	4
double percussion sur une peau	0	1	0	0	0	2	2	3	1	1	<b>15</b>	0
un son étouffé	1	7	5	11	5	0	0	9	0	3	<b>16</b>	8
on étouffe le son	1	10	3	5	5	2	3	10	3	3	<b>15</b>	5

En gras, les valeurs dues à plus de 3/4 des non musiciens.

Une des principales différences entre les deux contextes concerne le son VBN, associé à *du métal* pour 9 participants alors que seulement par 2 participants le considèrent comme ayant *un son métallique*. De même, les sons VBR et PNO, considérés comme *résonnants* par plus de 50% des sujets ne sont pas pour autant associés au portrait *un morceau de métal mis en résonance*, même si le timbre dominant dans ces deux contextes est le son VBN. De même, si le son TBN est associé par plus de 3/4 des sujets à la fois comme *son cuivré* et comme *du cuivre*, le son TPT est considéré comme *cuivré* par plus de 50% des sujets même si seuls 7 d'entre eux l'associent à *du cuivre*.

De même encore, si *un son double* et *une double percussion* sont tous deux principalement associés au son VBN, le son OBC est considéré comme *double* par 9 sujets alors qu'il n'est associé à l'idée de *double percussion* que par un seul d'entre eux. Enfin, on retrouve le même phénomène avec un *son étouffé* et *on étouffe le son*, tous les deux principalement associés au son VBN, même si le son CNT est *étouffé* pour plus de 50% de sujets. Il semble donc que les adjectifs permettent d'associer plus de sons différents à une qualité que les noms de modes de production ou de matériaux correspondants. Ces derniers semblent plus spécifiques. Ainsi par exemple, un son peut paraître métallique, même si on n'associerait pas ce son avec l'idée de métal.

## II.5. Meilleurs portraits pour chacun des 12 sons

Nous allons reprendre ici tous les portraits qui ont été associés par plus de 3/4 des sujets à chacun des 12 sons de notre ensemble. En effet, le nombre et le type de ces portraits varient beaucoup d'un son à l'autre et peuvent être intéressants à analyser (voir tableau 8.16).

Tableau 8.16. Portraits associés par plus de 3/4 des sujets à chacun des 12 sons de notre ensemble

Sons	NB	Portraits
TPT	2	famille cuivre, "trompette"
TBN	9	soufflé, on souffle, cuivré, du cuivre, on souffle dans du cuivre, famille des cuivres, famille des vents, trombone et "trombone"
CRA	1	continu
CNT	10	soufflé, on souffle, constant, on souffle dans du bois, on souffle dans un tube, famille des bois, famille des vents, clarinette, flûte et "clarinette"
CFT	8	diffus, P3-CFT (long, diffus et attaque lente), on frotte, on glisse, on effleure, violon, violoncelle, alto
VBR	18	attaque rapide, clair, agréable, pur, clair et doux, clair et rond, pur et net, P3-VBR (attaque rapide, clair et doux), on frappe, on percute, on tape, du cristal, du verre, on frappe du verre, famille des percussions, xylophone, cloche et "vibraphone"
PNE	5	attaque rapide, fort, net et franc, P3-PNE (attaque rapide, net et fort), on percute
PNO	2	long, famille des claviers
CVC	8	attaque rapide, métallique, pincé, on pince, une corde métallique, on pince une corde, famille des cordes et "clavecin"
OBC	5	aigu, attaque rapide, aigu et métallique, P3-CVC (pincé, aigu et métallique), on pince
VBN	4	étouffé, on étouffe le son, on percute une peau, double percussion sur une peau
SNO	0	

Pour différencier les noms d'instruments des noms d'instruments visés par le concepteur, ces derniers seront entre "".

On peut en effet considérer 4 types de comportements différents correspondant à 4 groupes de sons. Dans chaque cas, il s'agit d'associations faites par au moins 3/4 des sujets.

**1- Ceux qui sont associés à très peu de portraits.** En particulier, le son SNO n'est associé à aucun des 201 portraits proposés. C'est un son hybride n'ayant pas de spécificité et ayant été moins décrit spontanément que les autres. Déjà il a été difficile de le caractériser de manière extrême à partir des 23 échelles sémantiques de l'expérience 2, le portrait résultant étant *assez long, assez doux et peu sec*. Puis, le son CRA est associé à un seul portrait, un des adjectifs de deuxième groupe testé. C'est un des sons qui imite le plus mal l'instrument visé par la synthèse. Enfin les sons TPT et PNO, sont chacun associés par plus de 3/4 des sujets à deux portraits différents. Pour le son TPT, ces deux portraits relèvent tous les deux du groupe décrivant une source sonore possible. Pour le PNO, il s'agit d'un adjectif et d'un nom de famille d'instruments. Ce qui rassemble ces sons est sans doute d'une part leur caractère très synthétique et artificiel, d'autre part leur absence de spécificités dans le groupe de sons.

**2- Ceux qui sont associés à entre 4 et 5 portraits différents.** On trouve ainsi le son VBN, pour lequel les portraits associés renvoient à une source sonore possible de ce son, et ce, malgré le fait que ce soit un son hybride. On pourrait résumer l'ensemble des portraits le désignant sous la forme *on percute une peau puis on étouffe le son*. On peut en effet noter que le seul portrait sous forme d'adjectif qui lui est associé (*étouffé*) renvoie lui aussi à un mode de production. Puis, le son OBC, lui aussi un hybride est par contre associé à des portraits formés essentiellement à partir d'adjectifs. De plus, le seul portrait sous forme de mode de production qui lui est associé, *on pince*, est lui aussi repris par un adjectif *pincé*. On pourrait résumer l'ensemble des portraits le désignant sous la forme *aigu, métallique et pincé*, soit justement le portrait P3-CVC auquel il est associé. Enfin, le son PNE est également associé essentiellement à des portraits sous forme d'adjectifs, bien que ce soit un son parmi les plus vraisemblables (piano échantillonné). On pourrait résumer l'ensemble des portraits le désignant sous la forme *attaque rapide net et fort*, qui est là aussi le portrait P3-PNE auquel il est associé.

**3- Ceux qui sont associés à entre 8 et 10 portraits différents.** On trouve ainsi le son CVC, principalement associé à des portraits décrivant la source de ce son. On pourrait résumer l'ensemble des portraits le désignant sous la forme *on pince une corde métallique*. Puis le son CFT, lui aussi plutôt associé à des portraits décrivant sa source, particulièrement sous forme de noms de modes d'excitation et de noms d'instruments à cordes frottées. On pourrait résumer l'ensemble des portraits le désignant sous la forme *corde frottée*. Puis le son TBN, essentiellement associé à des portraits décrivant sa source, les 2 portraits sous forme d'adjectifs qui lui sont associés, *cuivré* et *soufflé*, renvoyant eux aussi soit à un nom de matériau, soit à un nom de mode de production. On pourrait résumer l'ensemble des portraits le désignant sous la forme *on souffle dans du cuivre*. Enfin, le son CNT comme les autres sons de ce groupe est associé principalement à des portraits décrivant sa source. L'un des 2 portraits sous forme d'adjectifs qui lui est associé, *soufflé*, est lui aussi dérivé d'un mode de production. On pourrait résumer l'ensemble des portraits le désignant sous la forme *on souffle dans un tube en bois*.

4- **Le dernier groupe** est formé par le seul son VBR associé à 18 portraits différents, soit près de 2 fois plus que tous les autres sons. Il s'agit à la fois de beaucoup d'adjectifs qui permettraient de le décrire sous la forme *attaque rapide, clair, doux et pur*, et de beaucoup de noms de sources sonores, qui pourraient être résumées par *on frappe du verre*. On peut remarquer que beaucoup de noms de matériaux visaient ce son, qui a souvent été le son préféré par l'ensemble des sujets lors de l'écoute et a donc été beaucoup décrit. Il semble se distinguer de l'ensemble des 12 sons par son caractère agréable et doux, alors que tous les autres sont plutôt considérés comme artificiels et désagréables. Si vous voulez synthétiser un son plébiscité comme agréable, choisissez le vibraphone !

Dans notre ensemble de 12 sons, certains sont donc choisis beaucoup plus souvent que d'autres. On peut y voir deux raisons. D'une part, ce sont des sons qui présentent de fortes spécificités qui permettent de les distinguer dans le groupe, par exemple les sons CVC, CFT, VBR, chacun un unique exemplaire de leur mode de production. D'autre part, ce sont des sons qui ont été plus souvent décrits que les autres lors des verbalisations libres et qui ont donc mené à l'établissement de plus de portraits les "visant".

## II.6. Résumé

Pour chaque type de portrait, nous avons rassemblé les portraits qui ont été associés à au moins un même son par plus de 3/4 des sujets (voir tableau 8.17).

On observe globalement un plus grand nombre de portraits associés à un même son (au moins) pour les portraits décrivant le son que pour ceux décrivant la source, même si cette différence n'est pas significative ( $\chi^2(1)=0.61$ ,  $p=.43$ ). De plus cette différence est essentiellement due aux portraits construits à partir d'adjectifs négatifs, pour lesquels 67% des 21 portraits du sous-groupe sont associés à au moins un même son par plus de 3/4 des sujets. Dans le groupe des portraits décrivant le son, les différences entre les 5 sous-groupes ne sont pas significatives ( $\chi^2(4)=8.6$ ,  $p=.07$ ). Dans l'ordre de pourcentage décroissant, on trouve aussi les adjectifs positifs testés dans l'expérience 2 sous forme d'échelles sémantiques, puis les autres adjectifs simples, enfin les adjectifs groupés par 3 puis ceux groupés par deux. Comme nous avons vu que les adjectifs groupés par 3 ne sont pas *a priori* plus spécifiques d'un son que ceux groupés par deux (voir §II.4), la différence entre ces deux sous-groupes est peut-être due au plus faible nombre de portraits proposés dans le groupe des adjectifs groupés par trois.



Tableau 8.17. Pour chaque sous-groupe, portraits associés par plus de 3/4 des sujets à au moins un même son (% de chaque sous-groupe).

Type de portrait	nb	%	portrait (sons)
Total SON	91	42	(sans les adjectifs négatifs : 34%)
adjectifs positifs	21	43	aigu (OBC), attaque rapide (VBR, PNE, CVC et OBC), clair (VBR), diffus (CFT), fort (PNE), long (PNO), métallique (CVC), pincé (CVC) et soufflé (TBN et CNT)
adjectifs négatifs	21	67	attaque lente (CFT), pas brillant (CNT), pas doux (TPT, PNE, OBC), pas fort (CFT), pas métallique (CNT), pas nasillard (VBR), pas net (VBN), pas pincé (TPT, TBN, CRA, CNT, CFT), pas rond (OBC), pas résonnant (TPT, CRA), pas sec (TBN, CFT), pas soufflé (VBR, PNE, CVC, OBC), pas sourd (VBR, CVC), pas vibré (TPT)
Autres adjectifs	16	37	agréable (VBR), constant (CNT), continu (CRA), cuivré (TBN), étouffé (VBN), pur (VBR)
2 adjectifs	21	24	aigu et métallique (OBC), clair et doux (VBR), clair et rond (VBR), pur et net (VBR), net et franc (PNE)
3 adjectifs	12	33	P3-CFT: long, diffus et attaque lente (CFT), P3-VBR: attaque rapide, clair et doux (VBR), P3-PNE: attaque rapide, net et fort (PNE), P3-CVC: pincé, aigu et métallique (OBC)
Total SOURCE	11	36	
		0	
Excitation	12	67	on souffle (TBN et CNT), on frotte (CFT), on glisse (CFT), on effleure (CFT), on frappe (VBR), on percute (VBR, PNE), on tape (VBR), on pince (CVC, OBC)
Matériau	17	23	du cuivre (TBN), du cristal (VBR), du verre (VBR), une corde métallique (CVC)
Source	26	31	on souffle dans du cuivre (TBN), on souffle dans du bois (CNT), on souffle dans un tube (CNT), on frappe du verre (VBR), on pince une corde (CVC), on étouffe le son (VBN), on percute une peau (VBN), double percussion sur une peau (VBN)
Famille	6	100	famille des cuivres (TPT, TBN), des bois (CNT), des vents (TBN, CNT), des cordes (CVC), des percussions (VBR), des claviers (PNO)
Instruments	37	22	trombone (TBN), clarinette (CNT), flûte (CNT), violon (CFT), violoncelle (CFT), alto (CFT), xylophone (VBR), cloche (VBR)
Fichiers	12	50	trompette (TPT), trombone (TBN), clarinette (CNT), vibraphone (VBR), corde frottée (CFT), clavecin (CVC)

Par contre, dans le groupe des portraits décrivant la source, les différences entre les 6 sous-groupes en terme de nombre de portraits associés à au moins un même son par plus de 3/4 des sujets sont significatives ( $\chi^2(5)=21.3, p=7.10^{-4}$ ). Ceci est dû au fait que 100% des portraits de noms de famille d'instruments sont associés à au moins un même son par plus de 3/4 des sujets, et que c'est le cas pour 67% des portraits de noms de modes d'excitation. De même que pour les portraits décrivant la source, la différence entre les sous-groupes de noms d'instruments et des instruments visés par le concepteur, construits à base de noms d'instruments semble être due au plus faible nombre de portraits du groupe qui permet aux participants de concentrer leurs réponses sur les mêmes sons.



D'autre part, lorsque dans l'expérience 3, 20 participants ont eu à associer un ou plusieurs sons au portrait "un son de trompette", leurs choix ont abouti à 29 associations au total, dont 13 avec le son TPT, 12 avec le son TBN, etc... On peut envisager deux façons de comparer les verbalisations spontanées d'une part et le choix d'association de sons aux portraits lors de l'expérience 3.

1- A priori nos données sont des effectifs : nombre et nature des sons décrits ou associés par l'ensemble des sujets à chaque locution. On devrait donc s'orienter vers des mesures de  $\chi^2$  pour tester d'éventuelles différences de répartition des réponses des sujets sur l'ensemble des 12 sons possibles dans les 2 situations qu'on veut comparer. On s'attend à ce que si on associe les mêmes sons aux portraits que ceux qui ont été à l'origine de ces locutions, les différences de répartition testées au moyen des  $\chi^2$  seraient non significatives.

2- On peut étudier la corrélation existante entre le nombre de sujets qui ont fait chacun des associations locution-son, de manière spontanée ou lors de l'association aux portraits. On s'attend à ce que si les participants associent les mêmes sons aux portraits que ceux qui ont été à l'origine de ces descriptions, les corrélations entre les nombres de gens ayant fait chaque association seraient d'autant plus fortes et significatives. Ceci sous entend que nos données puissent être considérées comme continues et suivant une loi normale... ce qui n'est pas le cas !

Plusieurs problèmes se présentent si on calcule des  $\chi^2$  :

1- Lors du calcul de  $\chi^2$ , on attend des différences de répartition non significatives, ce qui nous oblige à garder notre hypothèse nulle initiale dans les cas qui nous intéressent, c'est à dire ceux où on observe la même répartition d'une locution spontanée sur les 12 sons que lors de l'association de ces sons au portrait.

2- Nos répartitions ne sont pas exclusives : d'une part, un même sujet peut choisir de décrire avec le même terme des sons différents (par exemple "trompette" sert à décrire 7 sons différents sur l'ensemble des sujets de l'expérience 1). D'autre part, un même sujet peut associer un ou plusieurs sons à chacun des portraits proposés.

3- Plus une description et une reconnaissance vont être "spécifiques" d'un son, plus il va y avoir de "cases vides", c'est à dire de sons qui n'ont jamais été associés à cette locution (c'est le cas de 5 sons dans notre exemple de trompette). Le problème c'est que cela entraîne lors du calcul du  $\chi^2$  des 0 dans les effectifs théoriques attendus qui rendent le calcul du  $\chi^2$  impossible sur l'ensemble du tableau, bien que ces 0 correspondent à une égalité entre effectifs théoriques et observés (0 dans les deux cas). Le paradoxe, c'est que le calcul complet n'est possible que dans les rares cas où la même description a été associée au moins une fois à chacun des sons, c'est-à-dire les cas les moins intéressants.

Nous avons donc choisi de limiter le calcul du  $\chi^2$  aux seules cases non vides. On réduit du même coup le nombre de degrés de liberté, le calcul est alors possible, mais est-il vraiment représentatif de nos données ? En effet, si seuls 2 sons ont été décrits avec une locution et que lors du test de la reconnaissance de ce même portrait verbal, les sujets n'associent que les 2 mêmes sons, on ne prend pas en compte le fait que tous ces participants ont choisi de décrire et d'associer aucun des 10 autres sons à ces verbalisations. En effet, ces 10 "non choix" sont identiques dans les deux situations et devraient aboutir à une participation au  $\chi^2$  nulle. Ainsi, dans l'exemple ci-dessus, le  $\chi^2$  qu'on peut calculer est égal à 5.52, suivant si on considère qu'il est représentatif de l'ensemble des données (dl=11) ou seulement des cases non vides (dl=6), le seuil de significativité associé à ce  $\chi^2$  sera de  $p=.90$  ou  $p=.48$ . Dans les deux cas ici, le  $\chi^2$  n'est donc pas significatif, mais cette diminution du nombre de degrés de liberté peut rendre certaines valeurs significatives. Nous avons calculé les  $\chi^2$  même dans les cas où le pourcentage d'effectifs théoriques inférieur à 5 était supérieur à 20%, conscient que ceci menait à une sur-évaluation des  $\chi^2$ , donc de leur possibilité d'être significatifs (erreur de type I). Néanmoins, comme nous cherchons à montrer que la plupart d'entre eux ne sont pas significatifs, cela ne nous avantage pas. Ce test est donc conservateur par rapport à nos attentes. De plus, des corrections de Yates ont été appliquées au calcul du  $\chi^2$  dans les cas des faibles effectifs (moins de 40 associations en tout).

Nous sommes donc conscients des limites de validité de ces résultats. Il semble qu'une analyse bayésienne de tableau de contingence (cf. Jean-Marc Bernard) conviendrait mieux à nos données. Une collaboration avec l'auteur est envisagée dans un futur proche. L'ensemble des résultats de ces calculs de  $\chi^2$  est présenté en annexe (tableaux A.8.56 à A.8.65). Nous allons en faire une description ci-dessous.

### III.1. Portraits décrivant le son

Dans cette partie, nous allons donc analyser pour les sous-groupes de portraits décrivant le son, d'éventuelles différences de répartition entre les associations faites entre chaque locution et chaque son d'une part dans le cadre des verbalisations spontanées et d'autre part, dans celui des associations aux portraits. Comme des associations spontanées de 3 adjectifs n'ont pas été observées, nous ne testerons pas ce sous-groupe de portraits, mais seulement ceux construits à partir d'un ou de 2 adjectifs.

#### **III.1.1. Qualités simples testées dans l'expérience 2**

Pour ce sous-groupe de portraits, les analyses sont en annexe (tableaux A.8.56. et A.8.57). Nous avons comparé les associations locution-son effectuées pour les 21 adjectifs positifs et négatifs, ainsi que pour 11 autres types de locutions qui pouvaient se rapprocher des portraits considérés.

##### Comparaison Spontanée E1 - Portraits E3

On observe pour les 21 adjectifs positifs globalement 71% des différences entre les associations spontanées extraites de l'expérience 1 et les associations faites dans les portraits qui sont non

significatives. C'est le cas de 84% des différences pour les adjectifs négatifs. En particulier, la manière dont les sujets décrivent spontanément un son comme *fort* est significativement différente de la façon dont ils associent ces mêmes sons au portrait *un son fort*. C'est aussi le cas pour les adjectifs *doux*, *riche*, *long*, *métallique* et *large*, mais aussi pour *pas doux*, *pas long* et *pas sec*.

Tableau 8.19. Nombre total d'associations locution-son (nT) spontanées dans l'expérience 1 (S1) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	CVC	PNE	PNO	OBC	VCN	SNO	nT
P- fort	<b>11</b>	11	<b>9</b>	2	0	11	2	<b>38</b>	2	7	7	<b>0</b>	45
S1- fort	30	15	0	6	3	7	2	15	0	2	7	14	88
P- doux	0	4	<b>0</b>	24	18	16	6	0	16	<b>0</b>	6	<b>8</b>	49
S1- doux	0	7	12	10	15	25	2	5	12	10	2	0	59
P- riche	4	13	4	0	10	15	10	3	10	13	15	<b>3</b>	71
S1- riche	11	15	2	3	12	9	5	<b>12</b>	5	5	6	<b>15</b>	65
P- long	4	8	4	5	17	8	3	1	20	4	<b>17</b>	9	76
S1- long	8	8	7	7	30	3	7	3	10	3	<b>2</b>	11	98
P- métallique	3	3	4	0	4	<b>16</b>	23	<b>16</b>	3	19	3	<b>6</b>	69
S1- métallique	9	5	2	2	5	7	24	<b>5</b>	7	13	8	15	123
P- large	2	17	2	9	15	15	2	4	11	6	15	4	54
S1- large	<b>19</b>	30	7	22	4	4	0	4	4	4	<b>0</b>	4	27
P- pas sec	7	<b>21</b>	8	<b>17</b>	<b>21</b>	8	1	3	15	0	<b>18</b>	10	71
S1- pas sec	0	<b>0</b>	0	<b>2</b>	<b>3</b>	0	0	2	2	3	<b>0</b>	0	65
P- pas long	12	6	11	12	2	10	12	12	<b>0</b>	11	3	4	93
S1- pas long	0	0	29	14	0	0	0	43	<b>29</b>	29	14	0	7
P- pas fort	3	2	5	10	17	7	9	0	6	2	14	8	88
S1- pas fort	0	0	<b>64</b>	9	27	<b>45</b>	36	<b>9</b>	45	9	9	0	11

En gras, les participations au 2 les plus importantes

Ainsi, les différences observées pour l'adjectif *fort* sont essentiellement dues à une plus forte association aux sons CRA et PNE lors de la reconnaissance de portraits que spontanément, et une plus forte association spontanée aux sons SNO et TPT que lors de la reconnaissance de portraits. De même, les différences observées pour l'adjectif *doux* sont essentiellement dues à l'existence d'associations spontanées aux sons CRA et OBC alors qu'ils ne sont jamais associés à ce portrait. Par contre, le son SNO est associé au portrait *doux* alors qu'il n'est jamais spontanément décrit avec cet adjectif. Pour l'adjectif *riche*, les différences sont essentiellement dues au fait que le son SNO est spontanément beaucoup plus décrit comme riche qu'il n'est associé à ce portrait, ce qui est aussi le cas, dans une moindre mesure pour le son PNE. Pour *long*, les différences sont dues quasiment essentiellement au fait que le son VCN est beaucoup plus associé à ce portrait qu'il n'est spontanément décrit comme *long*. Pour *métallique*, les différences sont dues au fait que les sons PNE et VBR sont plus souvent associés à ce portrait qu'ils ne sont spontanément décrits comme métalliques, et qu'au contraire, le son SNO est plus souvent décrit comme *métallique* qu'il n'est associé à ce portrait. De plus, le son TPT est plus souvent décrit comme *large* qu'il n'est associé au portrait correspondant alors que le son VCN est au contraire beaucoup associé à ce portrait malgré le fait qu'il n'est jamais spontanément décrit comme *large*.

Le son TBN, est souvent associé au portrait *pas sec* alors qu'il n'est jamais décrit spontanément ainsi. C'est aussi le cas des sons VBN, CFT et CNT. Les différences observées dans le cas du portrait *pas long* sont essentiellement dues au fait que le son PNO est beaucoup décrit spontanément comme *pas long* alors qu'il n'est jamais associé à ce portrait. Enfin, les sons CRA, VBR et PNE sont plus souvent décrits spontanément comme *pas fort* qu'ils ne sont associés au portrait correspondant.

Parmi les 11 autres locutions spontanées comparées aux adjectifs positifs correspondant, 3 sont significatives. Ceci nous montre que même quand on rassemble l'ensemble des locutions traitant de la longueur des sons (*long, longtemps, se prolonge*) les correspondances avec les associations au portrait *long* ne sont pas améliorées. Il en est de même pour les objets considérés comme métalliques ou l'ensemble des locutions décrivant soit un son métallique soit un objet métallique.

Quatre autres locutions se référant aux qualités d'attaques ont été comparées aux locutions *attaque rapide* et *attaque lente* (respectivement un ensemble regroupant des locutions proches de *attaque forte* ou *une attaque*, et *attaque faible* ou *pas d'attaque*). Les locutions spontanées *attaque forte* ou *une attaque* présentent des différences significatives de répartition sur les 12 sons avec l'utilisation du portrait *attaque rapide*. De même, la différence d'association spontanée entre les locutions regroupées sous *attaque faible* et le portrait *attaque lente* est significative, alors que ce n'est pas le cas des locutions spontanées de type *pas d'attaque*.

#### Comparaison Spontanée E3 - Portraits E3

Sur les 21 adjectifs positifs proposés comme portraits, seul 90% d'entre eux ont été spontanément verbalisés dans l'expérience 3, mais toutes les différences entre les associations spontanées extraites de l'expérience 3 et les associations faites dans les portraits sont non significatives. C'est aussi le cas de 71% des différences pour les adjectifs négatifs, sachant que seulement 33% des 21 portraits d'adjectifs négatifs proposés ont été spontanément verbalisés dans l'expérience 3. La manière dont les sujets décrivent spontanément un son comme *pas clair* ou *pas résonnant* est donc significativement différente de la façon dont ils associent ces mêmes sons aux portraits correspondants.

Tableau 8.20. Nombre total d'associations locution-sons (nT) spontanées dans l'expérience 3 (S3) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	CVC	PNE	PNO	OBC	VBN	SNO	nT
P-pas clair	6	6	11	10	16	0	3	3	9	9	16	10	79
S3-pas clair	0	0	0	0	0	<b>100</b>	0	0	0	0	0	0	1
P-pas résonnant	20	9	17	16	5	2	10	5	1	7	2	5	86
S3-pas résonnant	0	0	0	0	0	0	0	0	<b>100</b>	0	0	0	1

En gras, les participations au 2 les plus importantes

Ces différences sont dues au fait que seul le son VBR a été spontanément décrit comme *pas clair* et le son PNO comme *pas résonnant*, alors que les associations aux portraits correspondant ont été beaucoup plus réparties sur plus de sons.

Parmi les 11 autres locutions spontanées comparées aux adjectifs positifs correspondant, une seule présente des différences significatives, la locution *clairement*, qui n'est donc pas utilisée spontanément

vis à vis des mêmes sons que l'adjectif *clair*, ni que l'utilisation du portrait *un son clair*. De même, parmi les 4 autres locutions comparées aux portraits *attaquerapide/lente*, une seule présente des différences d'utilisation significatives entre les verbalisations spontanées et les associations aux portraits. Il s'agit de *attaqueforte*, qui n'est donc pas utilisé spontanément dans le même sens que *attaque rapide*.

### III.1.2. Autres qualités simple

Pour ce sous-groupe de portraits, les analyses sont en annexe, tableau A.8.58. Nous avons comparé les associations locution-son effectuées pour les 16 autres adjectifs correspondant aux portraits, ainsi que pour 9 autres types de locutions qui pouvaient se rapprocher des portraits considérés.

#### Comparaison Spontanée E1 - Portraits E3

Sur les 16 adjectifs proposés comme portraits, on observe pour seulement 60% d'entre eux des différences entre les associations spontanées extraites de l'expérience 1 et les associations faites dans les portraits qui sont non significatives. En fait, les différences significatives concernent 6 adjectifs : *étouffé*, *constant*, *précis*, *court*, *agressif* et *éloigné* (voir tableau 8.21). Ainsi, les sons CFT et VBN (et, dans une moindre mesure, le son PNO) sont plus souvent associés au portrait *un son étouffé* qu'ils ne sont spontanément qualifiés de cette qualité (le son CFT n'est jamais décrit comme *étouffé*). Par contre, le son CRA, souvent décrit comme *étouffé* est peu associé au portrait correspondant. De même, les sons SNO et, dans une moindre mesure, CFT souvent décrits spontanément comme *constant*, ne sont presque jamais associés au portrait correspondant, alors que le son CRA, souvent associé au portrait *un son constant* était peu décrit avec cette qualité. Le son SNO, décrit comme *précis* dans 9% des utilisations spontanées de cet adjectif n'a pourtant jamais été associé au portrait correspondant, alors qu'au contraire, les sons PNE et VBR, souvent associés au portrait *un son précis*, ont rarement été définis spontanément avec cet adjectif.

Tableau 8.21. Nombre total d'associations locution-sons (nT) spontanées dans l'expérience 1 (S1) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	CVC	PNE	PNO	OBC	VBN	SNO	nT
P- étouffé	2	11	<b>8</b>	17	<b>8</b>	0	0	0	<b>14</b>	5	<b>25</b>	12	65
S1-étouffé	5	18	26	10	0	4	7	8	3	6	8	4	115
P- constant	19	6	20	26	0	6	7	6	7	3	0	<b>1</b>	70
S1-constant	21	11	<b>6</b>	27	<b>6</b>	0	3	2	3	0	2	<b>18</b>	62
P- précis	5	2	8	10	0	<b>22</b>	21	<b>17</b>	5	8	3	<b>0</b>	63
S1-précis	7	6	12	3	3	7	15	<b>4</b>	13	16	3	<b>9</b>	67
P- court	<b>12</b>	0	13	10	2	15	8	<b>19</b>	<b>2</b>	10	10	0	52
S1-court	1	6	7	3	1	18	17	6	<b>17</b>	10	14	1	72
P- agressif	21	2	19	5	5	0	0	<b>21</b>	0	26	2	<b>0</b>	43
S1-agressif	<b>46</b>	8	5	0	5	0	3	<b>0</b>	3	14	5	<b>11</b>	37
P- éloigné	4	4	4	8	21	0	2	0	17	4	25	12	52
S1-éloigné	<b>40</b>	0	<b>40</b>	0	0	0	0	0	0	<b>20</b>	0	0	5

En gras, les participations au 2 les plus importantes

De même, le son PNO, décrit comme *court* dans 17% des locutions utilisant cet adjectif spontanément ne représente que 2% des associations au portrait correspondant, alors qu'au contraire, les sons TPT et PNE, souvent associés au portrait *un son court* ont été très peu décrits spontanément à l'aide de cet adjectif. Au contraire, le son PNO représentant 21% des associations au portrait un son agressif n'a pourtant jamais été décrit spontanément avec cet adjectif, alors que les sons de SNO, et dans une moindre mesure de TPT, ont été moins souvent associés au portrait *un son agressif* qu'ils n'ont été spontanément considérés comme tel. En effet, le son SNO n'a jamais été associé au portrait *un son agressif* et le son TPT était associé à 46% des utilisations spontanées de cette locution. Enfin, les sons TPT, CRA, et dans une moindre mesure OBC, représentant l'essentiel des associations spontanées aux locutions *un son éloigné* (respectivement pour 40, 40 et 20% de ces associations), ont été beaucoup moins souvent associés au portrait correspondant, dont 25% des associations concernent le son VBN.

De plus, sur les 8 autres locutions spontanées testées, 2 seulement mènent à des différences non significatives avec les portraits proches. En effet, alors que la différence entre les associations spontanées et réalisées lors des portraits pour l'adjectif *chaleureux* n'est pas significative, elle le devient si on considère les locutions spontanées de *chaud*. Il en est de même pour les locutions *attaque franche*, dont la répartition sur les 12 timbres est significativement différente de l'association au portrait *un son franc*. De même, les locutions spontanées référant à l'idée de *loin* ou *lointain*, sont significativement différentes de l'association au portrait *un son éloigné*. Alors que les occurrences d'*un son cuivré* ne présentent pas de différence de répartition avec les associations au portrait correspondant (de même que les locutions spontanées font référence au matériau *du cuivre*), les différentes locutions spontanées se référant à l'idée de *cuivre* (celles rassemblant les locutions de *famille des cuivres* et de l'ensemble des locutions où le mot *cuivre* apparaît) présentent des différences significatives avec le portrait proche.

#### Comparaison Spontanée E3 - Portraits E3

Sur les 16 adjectifs proposés comme portraits, seuls 75% d'entre eux ont été spontanément verbalisés dans l'expérience 3, mais 83% des différences entre les associations spontanées extraites de l'expérience 3 et les associations faites dans les portraits sont non significatives. En fait, les différences significatives concernent seulement les 2 adjectifs *précis* et *continu* (voir tableau 8.22).

Tableau 8.22 Nombre total d'associations locution-son (nT) spontanées dans l'expérience 3 (S3) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	CVC	PNE	PNO	OBC	VBN	SNO	nT
P- précis	5	2	8	10	0	22	21	17	5	8	3	0	63
S3-précis	0	0	0	0	<b>33</b>	0	0	33	0	33	0	0	3
P- continu	18	7	22	19	4	5	4	3	7	5	5	1	74
S3-continu	0	0	0	0	0	0	0	<b>100</b>	0	0	0	0	1

En gras, les participations au 2 les plus importantes



Ainsi, les sons CFT et PNE sont beaucoup plus souvent décrits spontanément respectivement comme *précis* et *continu* qu'ils ne sont associés aux portraits correspondants. Ces descriptions spontanées paraissent d'ailleurs assez incongrues. L'importance de la différence concernant *continu* est sans doute essentiellement due au fait qu'une seule occurrence spontanée de *continu* a été observée lors de l'expérience 2. Par contre, sur l'ensemble des 9 autres locutions testées, une seule présente des différences significatives avec son portrait proche. Il s'agit de *attaque franche*, ici aussi associée spontanément différemment des associations au portrait *un son franc*.

### III.1.3. Association de deux qualités

Pour ce sous-groupe de portraits, les analyses sont en annexe, tableau A.8.59. Nous avons comparé les associations locution-son effectuées pour les 21 paires d'adjectifs correspondant aux portraits. Il faut noter que les productions spontanées correspondant à ces couples d'adjectifs présentent très peu d'occurrences (en général une ou deux seulement). Ces résultats sont donc à considérer avec précaution.

#### Comparaison Spontané E1 - Portraits E3

Sur les 21 paires d'adjectifs proposées comme portraits, pour 90% d'entre eux les différences entre les associations spontanées extraites de l'expérience 1 et les associations faites dans les portraits sont non significatives. En fait, les différences significatives concernent seulement les 2 couples d'adjectifs *aigu et métallique* et *net et franc* (voir tableau 8.23).

Tableau 8.23. Nombre total d'associations locution-son (nT) spontanées dans l'expérience 1 (S1) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	CVC	PNE	PNO	OBC	VBN	SNO	nT
P-aigu et métal.	5	0	10	0	5	2	29	12	0	36	2	0	42
S1-aigu et métal.	0	<b>100</b>	0	0	0	0	0	0	0	0	0	0	2
P- net et franc	11	2	5	5	0	22	20	27	0	7	0	0	55
S1-net et franc	0	<b>50</b>	0	0	0	0	50	0	0	0	0	0	2

En gras, les participations au 2 les plus importantes

En fait, toutes ces différences sont essentiellement dues à 3 productions spontanées assez incongrues au premier abord et concernant le son TBN. En effet, un même participant a spontanément parlé 2 fois de *la vibration aiguë et métallique* du son TBN, alors que lors des associations à ce même portrait, personne n'y a associé le son TBN, mais plutôt les sons OBC et CVC. De même, le son TBN a été une fois décrit spontanément comme *net et franc*, alors que lors des associations à ce même portrait, ce sont plutôt les sons PNE, VBR et CVC qui y ont été associés.

#### Comparaison Spontané E3 - Portraits E3

Sur ces 21 paires d'adjectifs proposées comme portraits, seul 29% d'entre eux ont été spontanément verbalisés dans l'expérience 3, mais toutes les différences entre les associations spontanées extraites de l'expérience 3 et les associations faites dans les portraits sont non significatives. 7 autres locutions spontanées ont été ici comparées à ces portraits, notamment celles des adjectifs *pincé*, *froid*, *mou* et

*rapide*, qui n'ont pas été spontanément associées à l'autre adjectif du couple présenté dans le portrait. Une seule de ces différences est significative : celle concernant la comparaison de la production spontanée de *rapide* avec le portrait *rapide et précis*. Ceci est dû au fait que spontanément les sons CFT et SNO ont été chacun une fois qualifiés avec l'adjectif *rapide* alors qu'ils n'ont jamais été associés aux portraits *rapide et précis*. Les autres comparaisons ont tendance à montrer que les associations aux portraits *doux et froid*, *froid et agréable*, *métallique et froid*, correspondent ici à l'utilisation spontanée de l'adjectif *froid*, même produit seul, qui caractérise en effet essentiellement le son VBR. De même, le portrait *pincé et résonant* est associé aux mêmes sons que ceux qui sont spontanément décrits comme *pincés*.

### III.2. Portraits décrivant une source sonore possible

Dans cette partie, nous allons donc analyser pour les sous-groupes de portraits décrivant une source sonore possible d'éventuelles différences de répartition entre les associations faites entre chaque locution et chaque son, d'une part dans le cadre des verbalisations spontanées et d'autre part, dans celui des associations aux portraits. Il s'agit donc de 6 sous-groupes de portraits : des noms de modes d'excitation, de matériaux, de sources sonores conçues comme association d'une action sur un matériau, de noms de familles d'instruments, d'instruments ou des noms d'instruments visés par le concepteur aux 12 sons de notre groupe.

#### **III.2.1. Noms d'action**

Pour ce sous-groupe de portraits, les analyses sont en annexe, tableau A.8.60. Nous avons comparé les associations locution-son effectuées pour les 12 noms de modes d'excitation correspondant aux portraits, ainsi que pour 9 autres groupements de locutions spontanées correspondant à l'ensemble de l'utilisation de chacun des verbes principaux testés.

#### Comparaison Spontané E1 - Portraits E3

Sur les 12 noms de modes d'excitation proposés comme portraits, 91% d'entre eux présentent des différences entre les associations spontanées extraites de l'expérience 1 et les associations faites dans les portraits sont non significatives. En fait, les différences significatives concernent seulement un des noms de modes d'excitation *on percute* (voir tableau 8.24).

Tableau 8.24. Nombre total d'associations locution-sons (nT) spontanées dans l'expérience 1 (S1) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO	nT
P- on percute	0	2	0	0	0	26	28	18	3	10	13	0	61
S1- on percute	0	0	0	0	0	25	0	0	<b>50</b>	25	0	0	4

En gras, les participations au 2 les plus importantes

Cette différence est essentiellement due au fait que le son CVC est plus souvent spontanément décrit avec le mode de production *on percute* qu'il n'est associé au portrait correspondant. Il semble donc

qu'une éventuelle confusion de mode de production entre l'idée de pincer et celle de frapper soit plus facile lors des verbalisations spontanées que lorsqu'on peut choisir parmi tous les sons de l'ensemble. De plus, sur les 9 autres ensembles de locutions contenant toutes les occurrences des principaux verbes concernés par ces portraits, 4 ensembles présentent des différences significatives entre les productions spontanées et les associations aux portraits. En effet, on observe spontanément peu d'occurrences du verbe *frotter* en rapport avec les sons VBN et SNO, alors qu'ils sont plus souvent associés au portrait *on frotte* (particulièrement le son SNO qui représente 26% des associations à ce portrait). Les différences correspondant au portrait *on percute* sont essentiellement dues au fait que spontanément, le son TPT a été beaucoup décrit comme ayant une *attaquepercutée* ou étant *percutant* (sans forcément de référence avec l'action mécanique de frapper, mais peut-être parce qu'une attaque de cuivre peut être très "franche" et paraître ainsi percutée). Il apparaît donc ici deux sens possibles de ce verbe *percuter* : lors de l'association aux portraits, seul le sens de frapper a été retenu alors que spontanément, un autre sens est aussi utilisé. On observe aussi des différences dues au fait que les sons CFT et SNO ont été plus souvent décrits comme *percutés* spontanément, alors qu'ils ont très peu été associés au portrait correspondant. On observe le même type de comportement pour les occurrences spontanées du verbe taper, beaucoup utilisé pour décrire le son OBC alors que ce son a été très peu associé au portrait correspondant. Enfin, l'essentiel des différences concernant les occurrences du verbe *pincer* sont dues au fait que le son SNO n'est jamais associé au portrait *on pince*, alors que spontanément, il est assez souvent décrit comme par exemple *un son pincé superposé à un son soufflé*.

#### Comparaison Spontané E3 - Portraits E3

Sur ces 12 noms de modes d'excitation proposés comme portraits, seul 50% d'entre eux ont été spontanément verbalisés dans l'expérience 3, et seulement 67% des différences entre les associations spontanées extraites de l'expérience 3 et les associations faites dans les portraits sont non significatives. En fait, les différences significatives concernent seulement les 2 noms de modes d'excitation *on souffle* et *on pince* (voir tableau 8.25).

Tableau 8.25. Nombre total d'associations locution-sons (nT) spontanées dans l'expérience 3 (S3) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO	nT
P- on souffle	24	29	9	29	4	0	0	0	0	0	2	4	55
S3- on souffle	6	19	13	<b>6</b>	<b>25</b>	0	0	0	0	0	<b>13</b>	<b>19</b>	16
P- on pince	0	0	0	0	0	11	16	7	33	27	5	0	55
S3- on pince	0	0	0	<b>13</b>	0	13	13	13	13	25	0	<b>13</b>	8

En gras, les participations au 2 les plus importantes

Qu'on regarde les productions spontanées correspondant strictement à *on souffle*, ou l'ensemble des locutions dérivées du verbe *souffler* (*on souffle*, *soufflé*, *un souffle*, *soufflant...*), la différence de répartition entre associations spontanées ou par portrait est significative. En effet, dans les deux cas, les productions spontanées des sujets des locutions contenant le verbe *souffler* se répartissent sur beaucoup plus de sons différents que les choix des sujets dans la situation d'association de sons au

portrait *on souffle*. Par exemple, le son SNO est beaucoup plus souvent spontanément décrit à l'aide du verbe *souffler* qu'il n'est associé au portrait correspondant. C'est aussi le cas du son CFT. Il est intéressant de remarquer qu'aucun de ces deux sons n'imitent une production par un souffle mais plutôt par un frottement. Il semblerait donc que la confusion entre frottement et souffle se fasse plus facilement lors de productions spontanées que dans le contexte d'association de sons avec un portrait verbal. Au contraire, le son CNT est plus souvent associé au portrait *on souffle* qu'il n'est spontanément décrit à l'aide du verbe *souffler*. Par contre, les différences concernant le portrait *on pince*, sont dues au fait qu'on observe des occurrences spontanées concernant les sons CNT et de SNO alors que ces deux sons ne sont jamais associés au portrait correspondant.

Parmi les 9 ensembles de locutions rassemblant toutes les occurrences des principaux verbes testés seules deux différences sont significatives : celle concernant les occurrences du verbe *souffler* (voir ci-dessus) mais aussi celles concernant le verbe *percuter*. On observe en effet le fait que le son SNO est décrit à l'aide de ce verbe (lors de la description de mélanges par exemple), alors qu'il n'est jamais associé au portrait correspondant.

### III.2.2. Noms de matériaux

Pour ce sous-groupe de portraits, les analyses sont en annexe, tableau A.8.61. Nous avons comparé les associations locution-son effectuées pour les 17 noms de matériaux correspondant aux portraits, ainsi que pour 16 autres groupements de locutions spontanées proches des portraits testés.

#### Comparaison Spontané E1 - Portraits E3

Sur les 17 noms de matériaux proposés comme portraits, les différences entre les associations spontanées extraites de l'expérience 1 et les associations faites dans les portraits sont non significatives pour 94% d'entre elles. En fait, les différences significatives concernent un seul nom de matériau *de la tôle* (voir tableau 8.26).

Tableau 8.26. Nombre total d'associations locution-son (nT) spontanées dans l'expérience 1 (S1) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO	nT
P- de la tôle	0	3	3	0	6	16	10	23	0	6	<b>29</b>	3	31
S1- de la tôle	0	6	0	0	0	0	0	<b>0</b>	0	0	<b>94</b>	0	16

En gras, les participations au 2 les plus importantes

Les différences sont dues essentiellement à un grand nombre d'occurrences spontanées décrivant le son VBN comme *on tape sur une tôle*, qui fait que 94% des locutions spontanées de *tôle* ont été faites pour décrire ce son, alors que lors de l'association au portrait correspondant, le son VBN n'y est associé que dans 29% des cas. Au contraire, le son PNE, associé dans 23% des cas au portrait *de la tôle*, n'est jamais spontanément décrit ainsi.

De plus, parmi les 16 autres ensembles de locutions testées, on observe 5 différences significatives. En effet, alors que les locutions spontanées de *du cuivre*, ainsi que *cuivré* se font de la même façon

que l'association au portrait *du cuivre*, si on ajoute les occurrences de *cuivre* correspondant à la *famille des cuivres*, on observe une différence significative due au fait que les sons VBR, de PNO et OBC ne sont pas décrits ainsi spontanément. De même, si on compare les locutions décrivant *un son naturel* au portrait *un matériau naturel*, on observe des différences dues au fait que le son CVC (et dans une moindre mesure PNE) est souvent considéré spontanément comme *naturel* alors qu'il est beaucoup moins associé au portrait *un matériau naturel*. Au contraire, le son CNT n'est jamais spontanément décrit comme *naturel*, alors qu'il est associé au portrait correspondant dans 15% des cas. On observe de plus que le son PNO est beaucoup plus souvent décrit comme *un son électrique* alors qu'il est moins souvent associé au portrait *un matériau électrique*. De même, le son SNO est plus souvent décrit comme *un son électronique*, et le son CRA représente 45% des locutions spontanées contenant le mot *électronique*, alors qu'ils sont beaucoup moins souvent associés au portrait *un matériau électronique*. Au contraire, le son TPT, souvent associé à ce portrait n'a jamais été décrit comme *électronique* spontanément.

#### Comparaison Spontané E3 - Portraits E3

Sur les 17 noms de matériaux proposés comme portraits, seuls 65% d'entre eux ont été spontanément verbalisés dans l'expérience 3, mais 73% des différences entre les associations spontanées extraites de l'expérience 3 et les associations faites dans les portraits sont non significatives. En fait, les différences significatives concernent les 3 noms de matériaux *du cuivre*, *du métal* et *une corde métallique* (voir tableau 8.27).

Tableau 8.27. Nombre total d'associations locution-sons (nT) spontanées dans l'expérience 3 (S3) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO	nT
P- du cuivre	18	42	0	3	0	8	0	5	5	5	13	0	38
S3- du cuivre	67	0	0	0	<b>33</b>	0	0	0	0	0	0	0	3
P- du métal	2	6	0	0	3	15	13	11	23	8	15	5	62
S3- du métal	0	0	0	0	0	<b>100</b>	0	0	0	0	0	0	4
P- une corde métalli.	0	0	0	0	11	2	24	11	31	16	4	2	55
S3- une corde métalli.	0	0	0	0	0	<b>25</b>	0	0	50	0	0	<b>25</b>	4

En gras, les participations au 2 les plus importantes

Les différences concernant *du cuivre* sont essentiellement dues au fait que le son CFT est décrit en référence à ce matériau dans 33% des occurrences spontanées, alors qu'il n'est jamais associé au portrait correspondant. Celles concernant le portrait *du métal* sont essentiellement dues au fait que le seul son décrit spontanément en référence à ce matériau est le son VBR, qui ne représente que 15% des associations au portrait correspondant. De même, les sons SNO et VBR sont beaucoup plus souvent décrits spontanément comme fabriqués à partir de *cordemétallique* qu'ils ne sont associés au portrait correspondant.

Pour ce qui est des 16 autres ensembles de locutions testées, seules 2 différences sont significatives. Elles concernent les locutions d'*objetélectrique* et de l'ensemble de toutes les locutions spontanées

contenant le mot *électrique* comparées au portrait *unmatériauélectrique*. En effet, dans les 2 cas, les sons PNO et PNE (de même que dans une moindre mesure le son CNT) sont beaucoup plus souvent décrits spontanément comme des *objetsélectriques* (vraisemblablement par exemple *pianoélectrique*) qu'ils ne sont associés au portrait correspondant. Par contre, le son OBC est plus souvent associé à ce portrait qu'il n'est spontanément décrit comme *électrique*.

### III.2.3. Noms de sources sonores

Pour ce sous-groupe de portraits, les analyses sont en annexe, tableau A.8.62. Nous avons comparé les associations locution-son effectuées pour les 26 noms de sources sonores correspondant aux portraits, ainsi que pour 10 autres groupements de locutions spontanées proches des portraits testés. On peut remarquer que cet ensemble de noms de source sonore peut se scinder en deux groupes par rapport à leur nombre d'occurrences spontanées. En effet, si la majorité des locutions testées représentent moins de 3 occurrences spontanées, 3 d'entre elles (*on frotte une corde*, *on pince une corde* et *on frappe une corde*) présentent entre 20 et 113 occurrences spontanées dans l'expérience 1.

#### Comparaison Spontané E1 - Portraits E3

Sur les 26 noms de sources sonores proposés comme portraits, les différences entre les associations spontanées extraites de l'expérience 1 et les associations faites dans les portraits sont non significatives pour 85% d'entre eux. En fait, les différences significatives concernent les 4 noms de sources sonores *on pince une corde*, *on frappe du bois*, *on frappe une corde* et *on frotte une corde* (voir tableau 8.28). Il s'agit donc globalement des locutions les plus souvent produites spontanément dans ce sous-groupe.

Tableau 8.28. Nombre total d'associations locution-son (nT) spontanées dans l'expérience 1 (S1) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO	nT
P- pince / corde	0	0	0	0	0	<b>45</b>	0	14	<b>0</b>	14	<b>27</b>	0	22
S1- pince / corde	0	1	0	1	0	13	19	10	27	17	7	4	113
P- frappe / bois	0	0	0	0	0	7	33	33	14	7	5	2	43
S1- résonne /bois	0	0	0	0	<b>50</b>	0	0	0	50	0	0	0	2
P- frappe / corde	0	0	0	0	0	6	15	15	36	23	4	0	47
S1- frappe / corde	0	3	0	0	3	<b>24</b>	14	34	<b>10</b>	<b>3</b>	7	0	29
P- frotte / corde	0	0	0	0	<b>42</b>	3	0	6	10	6	10	23	31
S1- frotte / corde	0	0	5	0	<b>90</b>	0	0	0	0	0	0	5	20

En gras, les participations au 2 les plus importantes

Ainsi, les différences concernant le portrait *on pince une corde* sont essentiellement dues au fait que le son VBR (et dans une moindre mesure VBN) est beaucoup plus souvent associé à ce portrait qu'il n'est spontanément décrit ainsi, ce qui laisse supposer la possibilité de confondre l'idée de frapper et de pincer lors de l'association aux portraits. Au contraire, le son PNO n'est jamais associé à ce portrait alors qu'il représente 27% des productions spontanées proches de *cordepincée*. On remarque en fait que les productions spontanées (très nombreuses) se répartissent sur un plus grand nombre de sons différents que les associations à ce portrait. Il est donc possible que les locutions spontanées de

*corde pincée* ou *on pince une corde* soient une manière plus générale de parler de son d'instrument à cordes.

Pour *on frappe une corde*, les différences sont essentiellement dues au fait que les sons CVC et OBC sont beaucoup moins souvent spontanément décrits par l'idée de *frapper une corde*, alors qu'ils représentent respectivement 36 et 23% des associations à ce portrait. On retrouve l'idée de confusion entre une émission par frappement et par pincement, cette fois ci plus présente lors de l'association aux portraits, alors que spontanément, les locuteurs décrivent l'idée de *frapper une corde* le plus souvent en référence au son VBR. Il se peut que spontanément, les locuteurs se focalisent plus sur l'idée de *frapper*, associée au son VBR alors que lors de l'association aux portraits, ce soit la présence de *corde* qui oriente les choix (le VBR imitant un frappement sur du métal). Pour *on frotte une corde*, les différences sont dues au fait que spontanément, les sujets associent les 20 occurrences quasiment exclusivement au son CFT, alors que lors de l'association aux portraits, le son CFT ne représente que 42% des associations. Enfin, pour le portrait *on frappe du bois*, que nous avons comparé aux 2 occurrences de locutions spontanées de *résonance dans du bois*, les différences doivent être dues aux différences entre ces deux locutions. En effet, ce sont les sons CFT et CVC qui ont été décrits avec l'idée de *résonance dans du bois* alors que le son CFT n'est jamais associé au portrait *on frappe du bois*, ce qui est assez logique.

Sur les 10 autres locutions spontanées testées (contenant en général seulement une partie de la description de la source), 2 présentent des différences significatives. En effet, les locutions spontanées de *plaquemétallique*, correspondent pour 33% chacune, aux sons VBR et PNE qui ne sont presque jamais associés au portrait *une corde pincée sur une plaque métallique*. En effet ce portrait, pour l'essentiel associé aux sons CVC et OBC, semble orienté par l'idée de corde pincée. De même, les occurrences de *tuyau* correspondent, pour 33% chacune, aux sons CRA et VBR qui ne sont presque jamais associés au portrait *du vent dans un tuyau*. Ceci est dû au fait que, pour le VBR, la résonance se fait dans un tuyau mais qu'on ne souffle pas vraiment dedans, et pour le CRA ce son a été rarement associé à l'idée de souffle à cause de son caractère trop artificiel.

#### Comparaison Spontané E3 - Portraits E3

Sur les 26 noms de sources sonores proposés comme portraits, seuls 46% d'entre eux ont été spontanément verbalisés dans l'expérience 3, mais 92% des différences entre les associations spontanées extraites de l'expérience 3 et les associations faites dans les portraits sont non significatives. En fait, les différences significatives concernent un seul nom de source sonore *on pince une corde* (voir tableau 8.29). On retrouve ici les différences concernant le portrait *on pince une corde*, dues ici aussi principalement au fait que les sons VBR et VBN sont beaucoup plus souvent associés à ce portrait qu'ils ne sont spontanément décrits ainsi, mais aussi au fait que les sons CVC et PNE ne sont jamais associés à ce portrait, ce qui est particulièrement étonnant pour le son CVC.

Tableau 8.29. Nombre total d'associations locution-sons (nT) spontanées dans l'expérience 3 (S3) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO	nT
P- pince / corde	0	0	0	0	0	45	<b>0</b>	14	<b>0</b>	14	<b>27</b>	0	22
S3- pince / corde	0	0	0	2	0	<b>4</b>	20	10	28	32	4	0	50

En gras, les participations au 2 les plus importantes

Parmi les 10 autres locutions testées, on trouve aussi 2 différences significatives qui concernent encore le fait que le son VBR est décrit comme vibration dans *un tuyau* alors qu'il n'est pas associé au portrait *on souffle dans un tuyau*. De plus, le son OBC est décrit spontanément avec l'idée d'*un tube* alors qu'il n'est jamais associé au portrait *on souffle dans un tube*. Ces différences sont donc dues au fait que des parties des descriptions de ces sources sonores peuvent différer trop de l'ensemble de la description. En particulier elles peuvent renvoyer à des instruments différents. Des analyses effectuées sur des mots et non sur des locutions plus complètes peuvent donc amener à des contresens.

### III.2.4. Familles d'instruments

Pour ce sous-groupe de portraits, les analyses sont en annexe, tableau A.8.63. Nous avons comparé les associations locution-son effectuées pour les 6 noms de familles d'instruments correspondant aux portraits, ainsi que pour 4 autres groupements de locutions spontanées proches des portraits testés.

#### Comparaison Spontané E1 - Portraits E3

Sur les 6 noms de familles d'instruments proposés comme portraits, les différences entre les associations spontanées extraites de l'expérience 1 et les associations faites dans les portraits sont non significatives pour 83% d'entre eux. En fait, les différences significatives concernent un seul nom de famille d'instruments *famille des cordes* (voir tableau 8.30).

Tableau 8.30. Nombre total d'associations locution-son (nT) spontanées dans l'expérience 1 (S1) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO	nT
P- F. Cordes	0	0	0	0	<b>14</b>	6	16	15	19	15	7	7	85
S1- F. Cordes	0	0	1	0	<b>38</b>	9	16	6	9	10	3	9	69

En gras, les participations au 2 les plus importantes

Ces différences sont essentiellement dues au fait que le son CFT est plus souvent décrit spontanément comme appartenant à la famille des cordes alors qu'il ne représente que 14% des associations à ce portrait. Il semble que dans le contexte des portraits à reconnaître, le plus grand nombre de sons imitant des cordes pincées ou frappées ait orienté le portrait *famille des cordes* vers un sens plus spécifique, alors que lors des productions spontanées, les sujets pensaient plus souvent à un *instrument à cordes* quand ils entendaient le son CFT, souvent associé aux instruments de violon,



violoncelle (peut-être aussi parce que spontanément, le piano n'est pas forcément considéré comme un instrument à cordes).

#### Comparaison Spontané E3 - Portraits E3

Les 6 noms de familles d'instruments proposés comme portraits ont tous été spontanément verbalisés dans l'expérience 3 et toutes les différences entre les associations spontanées extraites de l'expérience 3 et les associations faites dans les portraits sont non significatives. Il en est de même des 4 autres locutions proches testées. Il semble donc que ce type de portrait soit extrêmement concordant par rapport aux locutions spontanées.

### III.2.5. Noms d'instruments

Pour ce sous-groupe de portrait, les analyses sont en annexe, tableau A.8.64. Nous avons comparé les associations locution-son effectuées pour les 37 noms d'instruments correspondant aux portraits.

#### Comparaison Spontané E1 - Portraits E3

Sur les 37 noms d'instruments proposés comme portraits, les différences entre les associations spontanées extraites de l'expérience 1 et les associations faites dans les portraits sont non significatives pour 95% d'entre elles. En fait, les différences significatives concernent seulement 2 noms d'instruments *orgue* et *trombone* (voir tableau 8.31).

Tableau 8.31. Nombre total d'associations locution-sons (nT) spontanées dans l'expérience 1 (S1) et dans le cadre des portraits (P) et pourcentage pour chaque son

	TPT	TBN	CRA	CNT	CFT	VBR	PNE	PNO	CVC	OBC	VBN	SNO	nT
P- trombone	<b>28</b>	60	0	4	0	0	0	0	0	0	8	0	25
S1- trombone	<b>4</b>	60	0	4	0	0	0	0	0	0	32	0	25
P- orgue	<b>25</b>	9	16	28	3	0	3	6	3	3	0	<b>3</b>	32
S1- orgue	<b>4</b>	0	43	32	0	0	0	0	0	0	0	<b>21</b>	28

En gras, les participations au 2 les plus importantes

La différence concernant la locution trombone est due au fait que spontanément, le son TPT est moins souvent décrit comme un son TBN alors qu'il représente 28% des associations à ce portrait. L'autre incongruité concerne le mot *orgue*, utilisé spontanément pour décrire moins de sons différents alors que quasiment tous les sons ont été associés au moins une fois au portrait correspondant. En particulier, il est surtout utilisé spontanément pour décrire les sons CRA, CNT et SNO (et très peu TPT) alors que les associations de ce portrait au son TPT représentent 25% des cas.

#### Comparaison Spontané E3 - Portraits E3

Sur ces 37 noms d'instruments proposés comme portraits, 78% d'entre eux ont été spontanément verbalisés dans l'expérience 3 mais toutes les différences entre les associations spontanées extraites de l'expérience 3 et les associations faites dans les portraits sont non significatives. Il semble donc qu'en général les noms d'instruments soient suffisamment spécifiques de quelques sons pour que ce type de

portrait soit extrêmement concordant par rapport aux locutions spontanées, souvent dues aux musiciens.

### III.2.6. Noms d'instruments visés par le concepteur

Pour ce sous-groupe de portraits, les analyses sont en annexe, tableau A.8.65. Nous avons comparé les associations locution-son effectuées pour les 12 noms de fichiers correspondant aux portraits, ainsi que pour 4 autres groupements de locutions spontanées proches des portraits testés. Sur les 12 noms d'instruments visés proposés comme portraits, 100% des différences entre les 10 associations spontanées verbalisées lors de l'expérience 1 et les associations faites dans les portraits correspondant sont non significatives. Ces 12 noms de fichiers sont verbalisés spontanément pour 83% d'entre eux dans l'expérience 3. Toutes les comparaisons effectuées ici montrent des différences non significatives entre les distributions de sons correspondant d'une part à ceux qui ont motivé des locutions spontanées et d'autre part ceux qui ont été associés aux portraits verbaux correspondants. A propos des portraits de mélanges qui correspondent effectivement à la composition de ces sons, seule une production spontanée correspondait directement à ces portraits (et encore : c'est problématique puisque que c'est un membre du laboratoire qui a dit *mélange de hautbois et de clavecin...*) Nous avons donc essayé de comparer les associations faites par les sujets pour ces trois portraits de mélanges avec les locutions spontanées des sujets qui induisaient le fait qu'il y avait mélange et nous avons ainsi testé 4 ensembles de locutions : celles qui induisaient des mélanges *corde - vent*, comparé au portrait *mélange de hautbois et de clavecin*, *frappe et souffle* comparé au portrait *mélange de vibraphone et de trombone* et les 2 associations *frotté et souffle* et *soufflé/frotté et pincé* au portrait *mélange de piano et de corde frottée*. Toutes les différences sont non significatives.

### III.3. Résumé

Malgré toutes les réserves émises vis à vis de l'indicateur utilisé pour comparer ces données, on constate que globalement, peu de comparaisons entre la production spontanée de locutions et leur association lors de la reconnaissance de descriptions sous forme de portraits verbaux des sons mènent à des différences significatives. En effet, pour l'ensemble des 201 portraits proposés aux participants, les répartitions des sons associés aux portraits d'une part et lors des productions spontanées de ces locutions d'autre part mènent à des différences non significatives pour 86% des verbalisations recueillies dans l'expérience 1 et 91% de celles correspondantes recueillies dans l'expérience 3 (voir tableau 8.32). Pour l'ensemble des portraits décrivant le son, 78% des comparaisons avec les productions spontanées de l'expérience 1 et 91% des comparaisons possibles avec les productions spontanées de l'expérience 3 sont non significatives. C'est le cas de 91% des comparaisons dans le cas de l'expérience 1 et de 90% des comparaisons possibles dans le cas de l'expérience 3, pour l'ensemble des portraits décrivant une source sonore possible.

**Tableau 8.32.** Pour chaque groupe et sous-groupe de portraits : nombre de portraits (nb), pourcentage de comparaisons non significatives (%ns), pourcentage verbalisé spontanément dans l'expérience 3 (%V) et noms des portraits menant à des différences significatives entre l'association spontanée et lors de la reconnaissance des portraits

Spontané Portrait	nb	Expérience 1		Expérience 3		Lesquels
		% ns	Lesquels	% V	% ns	
Total Son	91	78	(17 ns sur 76 verbalisés)	48	91	(4 ns sur 44 verbalisés)
adjectifs +	21	71	fort, doux, riche, long, métallique, large	90	100	
adjectifs - autres adj.	21 16	84 60	pas sec, pas long, pas fort étouffé, constant, précis, court, agressif, éloigné	33 75	71 83	pas clair, pas résonant précis, continu
2 adjectifs	21	90	aigu et métal., net et franc	29	100	
Total Source	110	91	(9 ns sur 106 verbalisés)	67	90	(7 ns sur 74 verbalisés)
Excitation	12	91	on percute	50	67	on souffle, on pince
Matériau	17	94	de la tôle	65	73	du cuivre, du métal, une corde métallique
Source	26	85	pince/corde, frappe/bois, frappe/corde, frotte/corde,	46	92	pince/corde
Famille	6	83	famille des cordes	100	100	
Instrument	37	95	orgue, trombone	78	96	hautbois
Fichiers	12	100		83	100	
TOTAL	201	86	(26 ns sur 182 verbalisés)	59	91	(11 ns sur 118 verbalisés)

Pour l'expérience 1, globalement les différences de pourcentage de portraits ayant des différences d'associations spontanées / portrait non significatives entre les deux groupes "son" et "sources" sont très significatives ( $\chi^2(1)=6.96, p=.008$ ). Par contre, à l'intérieur du groupe des portraits décrivant le son, tout comme ceux décrivant la source, les différences entre sous-groupes de portraits ne sont pas significatives (respectivement  $\chi^2(3)=5.62$  et  $\chi^2(5)=3.60, p=.13$  et  $.61$ ).

Pour l'expérience 3, globalement ces différences entre les deux groupes "sons" et "sources" ne sont pas significatives ( $\chi^2(1)=4.10-3, p=.95$ ). Par contre, si les différences entre sous-groupes de portraits décrivant le son ne sont pas non plus significatives ( $\chi^2(3)=6.55, p=.09$ ), les différences entre les sous-groupes de portraits décrivant la source sont significatives ( $\chi^2(5)=14.6, p=.01$ ), à cause d'une sur représentation des différences significatives dans le sous-groupe des portraits de noms de matériaux.

Enfin, si on considère pour chaque groupe ou sous-groupe de portraits d'éventuelles différences de comportement entre les comparaisons avec les verbalisations spontanées de l'expérience 1 et celles de l'expérience 3, une seule de ces différences est significative. Il s'agit du sous-groupe des adjectifs positifs ( $\chi^2(1)=6.39, p=.01$ ), pour lequel on observe une sous représentation du nombre de portraits, pour lesquels on observe des différences significatives entre les associations spontanées de l'expérience 3 et celles dues aux portraits.

## **IV. Conclusion**

Les résultats présentés dans ce chapitre sont une des originalités de notre travail et permettent de tester la possibilité de reconnaître un son à partir d'une de ses descriptions en langue. L'intérêt de continuer d'une part à étudier la façon dont on parle des sons et d'autre part d'utiliser des descriptions verbales pour les indexer, par exemple dans des bases de données sonores, en dépend. En effet, s'il est de toute façon impossible de comprendre de quel son quelqu'un a parlé, il est vain de tenter de décrire les sons.

Il ressort de cette étude que lorsque des participants doivent choisir d'associer un ou plusieurs sons à une des 201 descriptions proposées, on obtient en moyenne 78% de Réponses Communes et 39% de Choix Communs. Cet accord entre participants est en moyenne plus important pour une description de la source sonore du son que sur ceux décrivant le son lui-même. On peut aussi noter globalement une plus forte cohérence entre les sujets quand ils sont musiciens que lorsqu'ils sont non-musiciens. Pour la plupart des portraits, les choix des participants se focalisent sur un petit nombre de sons, qui correspondent pour la grande majorité des cas aux sons décrits spontanément avec ces locutions. Ainsi, pour l'ensemble des 201 portraits proposés aux participants, les répartitions des sons associés aux portraits, d'une part, et des productions spontanées de ces locutions, d'autre part, mènent à des différences non significatives pour 86% des verbalisations recueillies dans l'expérience 1, et 91% de celles correspondantes recueillies dans l'expérience 3.

Il semble donc possible de communiquer verbalement sur la nature, la source sonore ou les qualités des sons que l'on perçoit de manière à être compris par d'autres personnes. Et ce, malgré le fait qu'on n'utilise pas, pour décrire les sons, de vocabulaire spécifique du domaine sonore et que les locuteurs ayant à décrire des sons sont rarement satisfaits des mots qu'ils utilisent pour les décrire, les jugeant souvent pas adaptés.

## **Partie III : Discussion**

Nous avons vu que lorsqu'on écoute un être humain parler des sons qu'il perçoit, cela nous donne des informations à la fois : 1- sur ce que représentent les sons pour lui : comment il les perçoit, quel sens le sonore a pour lui ; 2- sur le langage : comment il utilise cet outil pour rendre compte de ses perceptions, les décrire, communiquer ; 3- sur lui : comment s'organisent en général ses interactions avec le monde qui l'entoure à travers ses perceptions, ses représentations mentales, sa mémoire (ou ses mémoires ?).

Nous allons donc essayer de tirer les fruits de ce travail dans ces trois directions, à travers les trois paragraphes suivants intitulés : les sons pour nous, le langage pour décrire nos perceptions et notre perception du monde. Il s'agira de comprendre la portée des résultats obtenus dans ce travail expérimental et de faire une tentative d'élaboration d'un contexte théorique qui puisse rendre compte de ceux-ci. Nous décrirons donc nos hypothèses sur les "mécanismes mentaux" effectués lors des deux types d'expériences proposées ici : description de sons écoutés (de façon libre ou à travers une série d'échelles sémantiques) et association de son(s) à une description en langage naturel.

### **I. Les sons pour nous**

Ce paragraphe va essentiellement discuter des résultats analysés dans le chapitre 6 (Description des sons), mais aussi quelques résultats extraits du chapitre 8 (Données d'association de sons aux portraits). En effet, nous allons essayer de comprendre, à partir de l'analyse des champs sémantiques utilisés par les locuteurs lors des verbalisations libres de description des sons que nous leur avons proposées lors des expériences 1 et 3, quels sont les liens les plus systématiques liés à notre expérience du monde sonore.

Ceci pourrait nous aider à comprendre à quoi nous sert l'audition, ce qu'elle nous évoque. Les principaux résultats discutés ci-dessous seront donc le fait que : 1- on a tendance à décrire souvent une source sonore quand on nous demande de décrire un son, et 2- le son semble avoir une forme d'ordre temporel. Nous discuterons enfin la présence de deux attitudes d'écoute liées aux deux principaux types de verbalisations, selon qu'elles décrivent le son ou sa source.

#### **I.1. Un son perçu comme une source sonore**

Un examen des résultats analysés chapitre 6 nous montre que les sujets auxquels on demande de décrire verbalement les similarités et les différences entre les deux sons entendus, décrivent bien souvent non pas le son lui-même mais ce qui aurait pu le produire : un instrument, une action, un matériau, un mécanisme, comme s'ils revenaient à la définition même de ce qu'est un son : l'effet d'une vibration transmise par un milieu et produite par une action sur un matériau.

En effet, les auditeurs décrivent la source sonore dans 60% de l'ensemble des verbalisations, ce qui représente entre 12% et 98% des productions de chacun des sujets, sans distinction de compétence musicale. Ces pourcentages sont importants, surtout si on prend en compte le fait que ce type de verbalisations n'est pas induit par la consigne.

Ceci est d'autant plus frappant dans le cas des sons hybrides : ils n'ont pas été produits mécaniquement, pourtant les sujets s'évertuent à leur chercher un mode de production plausible. Par exemple, le son hybride appelé *vibrone* (mélange de vibraphone et de trombone) est décrit comme "ressembl(ant) à un coup sur une tôle ou une double percussion sur une peau", voire même comme "un tromboniste qui fait une bêtise ou une blague sur une paire de timbales" pour un des musiciens professionnels qui avait donc retrouvé l'idée du mode de synthèse. Les sons les plus réalistes sont rapidement identifiés. Les sons qui le sont moins sont plus longuement décrits. Les hybrides sont parfois reconnus comme tels ("bizarres", "mélanges"... ) et surtout la description des sons s'accompagne souvent de gestes mimant les modes de production : frappes, frottements, souffles. Cette tendance à décrire la source de nos perceptions se retrouve dans d'autres domaines sensoriels peu lexicalisés comme celui des odeurs (David et al., 1997).

Le champ sémantique de la source sonore qui apparaît dans nos données comprend avant tout des noms d'objets : instruments, parties d'instrument, familles d'instruments, objets courants ou matériaux ; des noms d'actions effectuées au début des sons qui provoquent leur émission : "pincer", "frapper"... ou à la fin du son, ce qui induit une terminaison qui est perçue comme libre ("laisse vibrer"... ) ou forcée ("on arrête"... ). Apparaissent aussi des descriptions plus complètes, décrivant des actions sur une matière, en particulier : "on pince une corde", "on frappe une corde", ainsi que des descriptions plus générales par exemple sur les qualités des objets sonores supposés en jeu, l'interprétation du musicien qui a produit le son, les lieux de provenance ou d'exécution...

### Niveaux de catégorisation

On retrouve donc d'une certaine manière 2 des 3 niveaux de catégorisation décrits par Guyot et al. (1998) lors de la description de bruits domestiques (grattements, frottements, froissements...) : le niveau qu'ils considèrent comme le second niveau, qu'ils définissent comme le niveau de base, et qui correspond à l'identification du mouvement ou du geste responsable de l'excitation d'une part, et d'autre part leur troisième niveau, pour lequel la source elle-même est identifiée. Par contre, le premier niveau qu'ils définissent comme le plus abstrait et qui correspondrait à la description du "type d'excitation" (mécanique versus électronique) est difficile à retrouver ici. Ceci est sans doute dû au type de sons écoutés par les auditeurs, qui apparaissent d'une certaine manière comme tous électroniques (ils sont tous produits par un synthétiseur), mais qui semblent aussi renvoyer tous à des productions mécaniques (on rappelle que les hybrides sont aussi décrits comme résultats d'excitations de type mécanique). Nous pourrions ainsi avancer l'hypothèse que les sons électroniques sont perçus de la même façon (c'est-à-dire à travers les mêmes *a priori*) que les sons produits mécaniquement.

De plus, il nous est difficile d'organiser nos verbalisations de manière hiérarchique, considérant par exemple, qu'il est plus abstrait d'identifier le mouvement ou le geste que la source elle-même. En effet, on note que les auditeurs en situation de manque du mot vont effectuer des gestes qui miment un

mode de production possible du son. Il est donc possible que le geste corresponde au niveau de base de l'identification des sons. De plus, du point de vue de la perception directe, la notion d'affordance nous mène à penser que la source contient toutes ses utilisations possibles : de même qu'une chaise contient la possibilité de l'utiliser (principalement pour s'asseoir, mais aussi pour monter dessus, construire des cabanes... en fonction de la taille et de l'imagination des utilisateurs !), un piano contiendrait l'idée de frapper (et parfois, dans certaines utilisations contemporaines, de frotter, ou de couper...). Il serait ainsi possible de considérer que "piano" serait à un niveau hiérarchique sous "frapper".

Il est néanmoins très difficile de systématiser cette mise en hiérarchie. En effet, nous avons remarqué que si l'action semblait décrite facilement, le matériau excité l'était aussi : "corde" se situerait hiérarchiquement sous les divers gestes possibles pour l'exciter (frapper, pincer ou frotter ?) mais aussi sous métal, et sous boyau. De plus, il est difficile de déterminer, quand un auditeur identifie "une corde" s'il se situe au niveau "concret" de l'objet ou au niveau plus abstrait de la famille d'instruments.

On remarque aussi que la notion de niveau de base ne correspond pas toujours à la classification musicologique : pour un grand nombre d'auditeurs (et il semble que ce soit d'autant plus important que ceux-ci sont néophytes en musique), "trompette" sert à identifier l'ensemble des cuivres, pour d'autres, l'ensemble des sons obtenus en soufflant (équivalent des vents), alors que certains utilisent bien ce terme pour identifier uniquement le son TPT, c'est-à-dire au niveau de l'instrument de musique précis. Ainsi, il faudrait pouvoir considérer, pour chaque personne, son arbre hiérarchique afin de positionner à quel niveau se situe "trompette", par exemple. On peut noter que certains sujets ont réparti l'ensemble des 12 sons proposés en deux catégories : trompette ou guitare, en fonction de la perception d'instrument à vent ou de cordes frappées ou pincées.

De plus, nous avons eu des difficultés à positionner les identifications de "piano" ou de "clavecin" : s'agit-il pour l'auditeur d'un niveau plus spécifique que pincer ou frapper (pour certains, ces deux modes d'excitations sont confondus à l'écoute), ou un niveau plus spécifique que "corde" ou encore que "clavier"... Ces objets complexes peuvent être classés de plusieurs façons, ce qui rend leur catégorisation difficilement modélisable par un système de classement purement hiérarchique où chaque objet appartient à une seule classe au niveau supérieur. Cette complexité peut être prise en compte par l'arborescence STONE, grâce au treillis de Gallois sous-jacent qui "relie" entre elles toutes les feuilles co-occurentes (c'est-à-dire les feuilles utilisées ensemble pour tous les objets). Malheureusement, dans le cadre de l'analyse de verbalisations libres, il est rare d'avoir des données systématiques et il faudra attendre la version utilisant la logique floue, en cours de réalisation (liant avec une force variable les feuilles utilisées ensemble souvent à très souvent), pour pouvoir véritablement profiter de cet avantage de STONE avec des données de ce type.

### Réminiscence d'une fusion Perception / Motricité

Le timbre apparaît donc à travers nos verbalisations comme lié à des expériences motrices ayant engendré du son... comme si le mouvement était inscrit, contenu dans le son.

On peut rappeler ici notre observation que les locuteurs en manque de mots pour décrire les sons proposés, font souvent le geste d'un mode d'excitation possible (dans 12% des descriptions). Cette tendance semble utilisée très précocement puisque nous avons eu la surprise de l'observer de manière fortuite chez des enfants entre 18 et 30 mois en crèche parentale lors d'un jeu de loto des sons : non seulement ces enfants associent facilement le son entendu à l'objet dont la photographie est reproduite sur un carton, mais même ceux incapables de dire le mot qui correspond miment la source, par exemple en faisant mine de se laver les dents ou de scier une planche...

On peut rappeler à ce sujet les résultats cités par Dumaurier (1982) que le fait de considérer un son comme l'attribut d'un objet est très précoce dans le développement, l'association entre un son et l'objet qui le produit étant observée dès 4 mois et le fait de tourner la tête en direction de la source sonore étant observé dès la naissance dans le cas de l'audition de la voix de la mère.

Il se pourrait ainsi que la forte présence du geste dans le son écouté soit une sorte de réminiscence de la période de fusion entre motricité et perception observée dans la petite enfance (Delalande, 1982). La découverte du domaine sonore et de son association avec nos mouvements lors de l'expérimentation "musicale" des jeunes enfants (par exemple lancer sa fourchette par terre jusqu'à ce que le partenaire ramassant la fourchette se lasse) aurait produit des associations en mémoire qui "coloreraient" de gestes toute écoute par la suite. On pourrait alors rejoindre Delalande quand il dit que "le geste laisse une empreinte dans la forme sonore elle-même et contribue à lui donner une signification" (p 163). Le son serait donc une forme de trace ou de signature d'une action sur une matière.

### Le son est avant tout signe ?

Les psychologues ont coutume de diviser les perceptions sonores en deux types selon qu'elles servent ou non à véhiculer le langage à travers la parole. Or une caractéristique première des sons de parole est qu'ils servent de code : dans une conception référentielle du langage, ils sont des signes conventionnels (voir discussion en partie théorique) qui renvoient à des "objets" du monde. On pourrait donc reprendre Arnauld et Nicole (1970) et dire que la nature du son de parole est d'exciter l'idée de l'objet auquel il renvoie à travers sa propre matérialité. Et de même que Condillac souligne le fait que l'association entre le mot considéré comme signe verbal et la chose qu'il représente est due à une co-occurrence au moins une fois, de la perception de cette chose et de l'audition du mot la désignant (Condillac, 1749), de même, on pourrait considérer que le son produit par une chose est une sorte de signe naturel de cette chose. Les signes naturels étant définis par le fait qu'ils semblent véhiculer une information portant sur ce dont ils sont les signes. La fumée est ainsi le signe du feu, et le nombre de cercles que l'on peut observer sur la souche d'un arbre, le signe de son âge. On pourrait ainsi reprendre l'assertion de Ludwig (1997, p 13) "rien n'est plus naturel que de faire abstraction de la matérialité du signe pour laisser vagabonder l'esprit vers la chose représentée".

C'est ainsi que les sons utilisés pour prononcer le mot "chaise" peuvent être considérés comme un signe parce qu'ils renvoient à un objet du monde réel sur lequel on peut s'asseoir. Mais on peut aussi considérer que si le son "pouet" renvoie à un objet du monde réel extérieur au monde sonore, comme par exemple une trompette, ce son est lui aussi considéré comme signe naturel de cet objet.



Cette tendance à servir de signe d'autre chose serait-elle une particularité du son ? Le fait que le timbre d'un son est perçu comme le signe de sa source serait-il une des raisons pour lesquelles le langage utilise principalement le canal auditif ? En effet, même si le langage peut être écrit, signé (dans la langue des signes) ou peut utiliser des pictogrammes, la langue parlée est sans doute son mode fondateur, tant du point de vue de l'espèce que de la majorité des individus (Segui & Ferrand, 2000). D'après Ludwig (1997), le visuel aurait plus de difficulté à faire signe. "Comme Berkeley l'a fait remarquer, on voit mal comment une image, par nature particulière, pourrait véhiculer la signification d'un terme abstrait" (p 13). Pourtant, les pictogrammes, notamment ceux du code de la route, sont bien des signes malgré leur nature particulière. Et si on considère la matérialité sonore, on est là aussi confronté à une multiplicité d'occurrences particulières. Une preuve en serait la difficulté pour une machine de catégoriser des "mots" prononcés au sein d'un discours. La catégorisation ne semble pas évidente. Le domaine sonore serait donc particulièrement facile à catégoriser, dans le sens d'oublier la particularité de l'exemplaire pour ne voir que la catégorie à laquelle il renvoie ?

### Un objet de mémoire

La prédominance de descriptions de la source ou d'un mode de production possible peut se comprendre si on réalise que le son est avant tout un objet de mémoire : sa composante temporelle est constitutive du son. Il paraît donc évident que quand on parle d'un son, il est déjà passé : on ne rapporte que ce que la mémoire en a retenu. Au contraire, on a l'impression que l'image perçue visuellement est fixe. On peut la décrire sans faire appel à la mémoire. Ce caractère permanent la fait apparaître comme une restitution "objective" du monde.

Si de plus on prend en compte le fait que nous ne sommes capables d'enregistrer et de reproduire le son que depuis un siècle, nous avons été longtemps obligés, pour le décrire de manière "objective", de donner les moyens à notre interlocuteur de le re-produire, pour lui permettre de l'entendre à nouveau ou d'imaginer un moyen de le produire. La stratégie de décrire la source sonore semble donc, dans cette perspective, un très bon moyen de permettre à l'autre de "re-produire" le son, et donc la seule façon de pouvoir le ré-entendre.

Il est évident que les pratiques en ce domaine sont en train de changer radicalement. Le fait de pouvoir enregistrer et reproduire le son à volonté en change son ontologie. Il devient ainsi plus "objectif", et prend un caractère permanent. C'est sans doute les nouvelles pratiques liées à ces nouvelles possibilités techniques qui ont produit des banques de sons, donc des façons de les mémoriser à long terme et qui explique l'essor des études actuelles sur la façon dont on parle des sons. Mais cette évolution de l'ontologie du sonore va engendrer aussi une évolution de la façon dont on en parle.

On retrouve ainsi une analogie entre le domaine sonore et l'olfaction pour laquelle on n'observe pas ou peu de lexicalisation : les noms d'odeurs étant toujours construits en référence à une source, peu de qualités olfactives ayant été abstraites.

Ainsi, on peut peut-être prophétiser que le temps d'intégrer complètement notre nouvelle possibilité de rendre le sonore "permanent", notre façon d'en parler va évoluer vers une abstraction des qualités sonores, de même qu'à partir du moment où le visuel a pu être rendu permanent, notamment par la peinture, on a abstrait des qualités visuelles qu'on considère comme des objets à part entière (par

exemple la couleur : on considère maintenant que le rouge existe alors qu'il n'existe que des objets rouges. Rouge n'est qu'une qualité d'une chose).

## **I.2. Qualités des sons**

Il est intéressant de discuter le fait que parmi les catégories de verbalisation les plus fréquentes quand les sujets parlent des sons sont des descriptions d'ordre temporel et des descriptions de la "forme" des sons. En effet, nous verrons que cette prépondérance d'une forme temporelle nous renvoie une nouvelle fois à l'identité de la source sonore.

### Une forme temporelle

Les locuteurs décrivent la forme temporelle du son, son déroulement en quelque sorte : montée ou attaque, plateau puis fin ou résonance, dans 79% des descriptions. On rappelle que ces sons sont pourtant très courts : 836 ms en moyenne. Les descriptions en question concernent donc des évolutions ou variations perçues à l'intérieur de ce laps de temps, donc inférieures à la seconde.

On rappelle en particuliers que les sujets décrivent souvent le début des sons : leur attaque, début, commencement, impact... ; mais aussi le "milieu" avec des descriptions de leur évolution, progression, ou au contraire de leur non-évolution, et de leur caractère constant, continu ... La fin des sons est aussi souvent décrite selon qu'elle est perçue comme progressive, avec résonance, écho, ou qu'elle est perçue comme brutale, étouffée, arrêtée... On trouve aussi des descriptions de qualités sonores qui relèvent du domaine temporel, concernant la durée (long / court ) ou la vitesse perçue (lent / rapide) des sons.

Ces descriptions nous semblent à rapprocher d'une part, aux descriptions de forme des sons entendus, et d'autre part, à certains gestes visant eux aussi à décrire les sons.

On trouve en effet des descriptions de forme des sons dans 44% des verbalisations. Il s'agit par exemple de considérer les qualités du "trait" (net, précis, franc, diffus, flou...), du "dessin" (rond, plat, pointu...), de la "composition" (simple, structuré...), de la "taille" (large, ample...) ou de la "densité" (resserré, ramassé...). Or, la plupart de ces descriptions de forme peuvent tout à fait être comparées à l'enveloppe temporelle de ces sons, visualisée avec un oscilloscope à stockage.

### Correspondance avec les descriptions acoustiques

On trouve en effet de nombreuses correspondances entre la présence d'un certain type de verbalisation et les caractéristiques acoustiques des sons décrits. Par exemple, les sujets décrivent plus souvent un début de son progressif quand un des sons de la paire est le son imitant la corde frottée, dont le temps d'attaque est le plus long, ou encore les sujets décrivent une fin brutale correspondant à l'imitation du phénomène de retour du sautereau plus souvent quand un des sons de la paire est le son imitant le clavecin, qui est le seul à posséder cette qualité temporelle qui dure quelques ms.

Il apparaît donc que les micro-variations temporelles des sons sont extrêmement saillantes et accessibles à tous les auditeurs, quelle que soit leur compétence musicale. Notre capacité à décrire

avec précision des phénomènes se déroulant dans un laps de temps de moins d'une seconde est intéressante à souligner. Ceci nous fait penser que, tout comme l'image a une forme qui se déroule dans l'espace, le son aurait une forme qui serait d'ordre temporel, et que c'est sans doute cette forme perçue par tout humain qui a été "captée" et représentée par les acousticiens quand ils ont "inventé" un moyen de visualisation d'une enveloppe temporelle des sons.

D'autre part, les verbalisations recueillies nous permettent parfois de découvrir la stratégie employée par le sujet pour réaliser sa tâche. Cette stratégie peut être explicite (trois sujets ont expliqué à la fin de l'expérience comment ils avaient jugé de la dissemblance entre les sons) ou transparaître dans la manière systématique d'aborder différents points pour chacune des paires.

Or, la plupart des sujets se sont focalisés en premier lieu sur l'attaque ou le début du son, ou sur une différence entre instruments à cordes et instruments à vent, ce qui correspond dans les deux cas à l'organisation des timbres suivant la première dimension. Certains ont fait ensuite des considérations de hauteur qui peuvent être liées au centre de gravité spectral des sons. On retrouve donc les deux premières dimensions perceptives. Dans un deuxième temps, beaucoup de sujets se sont focalisés sur la structure du son ou sur son évolution dans le temps. Enfin ils se focalisaient sur la manière dont le son se terminait ou sur sa longueur.

#### Rôle de l'attaque ?

Si on considère à nouveau l'importance des descriptions de la source sonore, on peut se poser la question du lien entre ces résultats et le fait que l'ensemble de sons étudiés ici est décrit de manière perceptive par un espace à trois dimensions (McAdams et al., 1995) dont la première, qui correspond donc à l'explication du maximum de la variance, est fortement corrélée au paramètre acoustique décrivant le temps d'attaque de ces sons. En effet, l'idée que l'identification des sons est grandement due à l'attaque ou aux transitoires a été beaucoup étudiée (Berger, 1964). Il semble que son importance soit d'autant plus grande pour des sons comme les nôtres qui ne sont pas affectés de variations dynamiques ou harmoniques au cours de leur durée (vibrato par exemple) (Schaeffer, 1966, p 230).

Si les sons de notre ensemble se différencient principalement par leur attaque, ceci peut avoir induit une plus grande tendance à les identifier.

### **I.3. Parler du son / parler de la source**

Nous avons donc découvert deux façons de décrire les sons, selon que les locuteurs décrivent directement le son ou "indirectement" une source sonore possible. Ces deux types de verbalisations renvoient d'une part aux deux types de définition du timbre présentés en introduction : l'un en référence à la source, l'autre analytique (le timbre est défini par ce qu'il n'est pas), mais aussi aux deux attitudes d'écoute décrites par Schaeffer (1966) et à sa suite Castellengo (1994) et Guyot (1998). De manière plus générale, on retrouve aussi la controverse entre les partisans d'une perception

d'objets contre les partisans d'une perception de qualités. Un examen de l'ensemble de nos résultats nous mène à conclure, tout comme Miller (1962) à une pluralité d'activités perceptives, les deux attitudes d'écoute étant révélées par les verbalisations des participants. Ces résultats nous permettent de valider chacune des deux théories de la perception, que ce soit l'idée de perception directe proposée par Gibson (1950) ou celle d'une perception de type "analytique". La situation dans laquelle le sujet est placé (dans un laboratoire de psychoacoustique, dans un studio d'enregistrement, lors du travail d'un instrument de musique ou dans la vie quotidienne) peut donner l'avantage à l'une ou l'autre des attitudes...

### Influence des compétences musicales

De manière générale, peu de résultats de notre étude se sont révélés liés à la compétence musicale des participants. Pourtant, on peut rappeler les résultats analysés dans le chapitre 8 concernant les différences de jugement de sens (précis, vague, pas de sens) observées lors de l'association de sons aux portraits verbaux proposés. En effet, on observe des différences en fonction du type de portrait proposé, mais aussi de la compétence musicale des participants.

Plus de portraits sont jugés n'ayant "pas de sens" s'ils décrivent la source sonore plutôt que le son (respectivement 12% contre 7%). Or, ceci semble essentiellement dû à une différence de comportement liée à l'expertise musicale des participants : les musiciens semblent plus sensibles aux sens des portraits décrivant la source sonore, les non-musiciens à ceux décrivant le son.

En effet, les musiciens considèrent que plus de portraits décrivant le son n'ont pas de sens (9% contre 5% pour les non musiciens), ces différences sont particulièrement importantes pour les adjectifs des groupes de qualité 1 (*aigu*) et qualité 2 (*doux et froid*). Les musiciens considèrent aussi que plus de portraits décrivant la source ont un sens précis (56% contre 49% pour les non-musiciens), ces différences étant particulièrement significatives pour les noms d'"instruments" électroniques et les noms de familles d'instruments (83% sont jugés précis par les musiciens contre 58% pour les non-musiciens).

Ceci peut s'expliquer par le fait que les adjectifs qui sont à l'origine des portraits décrivant le son, sont les moins spécifiques du domaine sonore que les noms utilisés pour construire les portraits décrivant la source sonore. En effet, les descripteurs spécifiques du champs sémantique lié au domaine sonore représentent 28% des noms contre seulement 12% des verbes et 4% des adjectifs. Dans ce contexte, on pourrait comprendre que les musiciens soient plus à l'aise avec les descriptions de sources qu'avec les adjectifs proposés.

Est-ce à dire que les musiciens ont une plus forte tendance à une perception directe de la source qu'à une écoute analytique des qualités sonores ? Ceci est contraire aux attentes de Schaeffer notamment, mais peut s'expliquer plutôt par le fait que les adjectifs testés sont ceux qui ont été le plus fréquemment produits dans la phase de description libre des sons. Pour avoir été produits par le plus grand nombre de participants, dans le plus grand nombre de situations, ils sont donc sans doute les plus vagues et polysémiques. Il se pourrait que les musiciens témoignent d'une écoute des qualités sonores par la production d'adjectifs plus "personnels" et plus variés, spécifiques de chaque situation d'écoute. Ces adjectifs n'auraient pas été sélectionnés avec comme objectif de sélectionner ceux

totalisant le plus grand nombre d'occurrences. Il s'agit donc d'une limitation du type d'expérimentation que nous avons choisi.

### Axes de classification

Nos résultats sont intéressants à comparer avec ceux obtenus par David (1997) à partir de définitions écrites de son, odeur et bruit. En effet, elle définit trois axes de classification les plus utilisés, qui permettent de classer environ 95% de ses réponses et qui sont les trois propriétés sur lesquelles se construisent les dénominations. Il s'agit tout d'abord d'un axe de classification lié à la source et construit soit en empruntant le nom de cette source, par exemple "odeur de cuisine, gaz d'échappement, nourriture, téléphone, bruit de la ville voisine", soit en utilisant un adjectif "dénominal", c'est-à-dire dérivé d'un nom, par exemple : "odeur corporelle, odeur épicée, bruit naturel, son personnel". Vient ensuite un axe lié à l'effet qui implique un rapport entre l'observateur et l'objet émetteur et qu'elle définit par "la manifestation d'un phénomène qui accompagne l'existence d'un objet, dont l'apparence ne renvoie pas directement à une cause productrice" (p 233), par exemple : "odeur piquante, odeur insupportable, bruit désagréable, son agressif". Le troisième axe de classification étant considéré comme lié aux propriétés physiques, "renvoyant au domaine des sciences de la nature (physique, chimie) : odeur forte, bruit sourd, son aigu, son grave, etc." (ibid.). La distinction entre les deux derniers axes de classification ne nous paraît pas aussi évidente : en quoi odeur "forte" et bruit "sourd" ne renvoient pas à des effets sur l'observateur ? impression de force, ou de surdité ?

Ce qui est néanmoins intéressant dans cette classification, c'est la comparaison effectuée entre les différents domaines à partir des définitions d'odeur, bruit et son. En effet, "pour l'olfaction, les réponses des sujets construites avec des dénominations liées à l'effet représentent 70% de l'ensemble. Pour plus de 40% d'entre elles, il s'agit même du seul axe utilisé. Pour le bruit et le son, les pourcentages sont très inférieurs et tombent respectivement à 6.5 et 4 % si l'on s'attache aux réponses construites exclusivement sur cet axe. L'axe source, quant à lui, est le plus représenté dans les dénominations de bruit (70.5%), mais les dénominations d'odeur et de son liées à cet axe présentent des pourcentages qui restent élevés (46.5 et 52 %). Enfin, pour les propriétés physiques, on retrouve le son (45.5%) et 17% des réponses sont uniquement organisées par cet axe (2% pour le bruit et aucune pour l'odeur (16.5% et 8%))." (p 235).

On retrouve donc bien l'idée que le domaine sonore est beaucoup décrit en terme de sources, comparé au domaine olfactif beaucoup décrit en terme d'effet. Néanmoins, cette étude ayant été effectuée sur des étudiants (donc pas particulièrement experts dans le rapport de leurs perceptions olfactives ou auditives), on peut se demander si la même question posée à des nez (parfumeurs) ou à des oenologues n'aurait pas abouti à plus de description des odeurs en terme de sources. En effet, le vocabulaire utilisé par les oenologues par exemple (Paris-Grignon, 1996) est essentiellement défini en terme de source ("banane", "poire", ou nom de la parcelle, de l'année de la récolte...).

## Littérature française

Enfin, on peut comparer l'ensemble du vocabulaire récolté ici avec le gigantesque travail effectué par Frize (2000 a, b) pour le Ministère de l'Aménagement du territoire et de l'Environnement. Dans son analyse de la manière dont les sons sont décrits dans la littérature française (à travers une exploration exhaustive des sept volumes des Lagarde et Michard) ainsi qu'à travers des entretiens enregistrés, il retrouve l'idée que pour parler des sons on donne le nom de l'objet, mais pas forcément de la partie de l'objet qui a produit le son. Par exemple, on dit plutôt "un bruit de voiture" que "le crissement des pneus d'une Peugeot sur une route humide". Frize en déduit que les auditeurs décrivent le son à travers leur relation à l'objet producteur. Il retrouve aussi l'abondance des descriptions métaphoriques des sons dans la littérature (plus de 200 adjectifs différents qui ne relèvent pas du domaine sonore apparaissent pour qualifier les sons) alors que ce type de description est absent des entretiens qu'il enregistre auprès de sujets variés. Il en conclue d'une part, que les auteurs n'écoutent pas les sons, au contraire du commun des mortels, mais inventent le monde et d'autre part, que la littérature ne semble pas influencer notre vocabulaire potentiel pour décrire les sons. Nous nous inscrivons en faux contre cette conclusion. En effet, les verbalisations que nous avons recueillies se rapprochent des descriptions métaphoriques des auteurs de la littérature, et ce, bien que nous ayons enregistré des participants de compétences variées. Les différences observées avec les entretiens effectués par Frize nous paraissent dues au fait que d'une part nous questionnons les participants sur des sons qu'ils entendent au moment où ils doivent les décrire et d'autre part au côté répétitif de l'expérience : les participants sont ainsi encouragés à développer leurs descriptions, ce qui peut faire apparaître plus de métaphores, alors qu'une description rapide se contenterait peut-être de nommer la source. Il semble donc bien que le commun des mortels puisse décrire les sons de manière métaphorique, à la manière des auteurs de littérature française.

## **Conclusion**

Le timbre apparaît donc à travers l'ensemble des descriptions que nous avons recueillies à la fois comme la signature de la source sonore et comme un ensemble de qualités sonores. Si les fonctions "écologiques" principales des sons étaient de localiser et d'identifier les objets qui les produisent, ce renvoi quasi-systématique à la source permettrait, à travers le procédé rhétorique de métonymie, de décrire directement le son comme un "objet" du monde, en disant par exemple en entendant un son (dans une cabine insonorisée à travers un casque) : "ça, c'est un piano".

On pourrait ainsi s'inscrire en faux contre toutes les "évidences" qui font étudier l'identification et la constitution des objets essentiellement dans le domaine visuel (Dokic et al., 2000) en affirmant qu'une des fonctions principales du timbre serait de reconnaître les objets du monde. Nous pourrions ainsi nous joindre à Casati et Dokic (1994) pour critiquer l'hégémonie du modèle visuel dans la philosophie de la perception.

## **II. Le langage pour décrire nos perceptions**

Utilisant des verbalisations comme mode d'accès à ce que représente le timbre pour nos sujets, nous pensions n'utiliser le langage que comme moyen et non pas comme sujet d'étude. Pourtant, les difficultés que nous avons rencontrées pour extraire le sens de ce que disaient les locuteurs nous ont révélé nos a priori sur le langage et obligé à les reconsidérer. En caricaturant, il nous paraissait possible de trouver un vocabulaire qui correspondrait directement à nos perceptions, comme une sorte de bijection entre deux mondes : perceptifs et sémantiques. Nous avons donc été d'abord surpris par le "feu d'artifice verbal" découvert en écoutant nos participants, puis désappointés devant nos difficultés à faire rentrer ce vocabulaire dans des catégories bien hiérarchisées. Et plus nous avons poursuivi notre investigation, plus le lien entre les mots et leur sens nous est apparu comme mouvant, avant d'en comprendre la logique qui semble s'inscrire dans la communication intersubjective.

### **II.1. Un vocabulaire inadapté ?**

En demandant de décrire 12 sons à 20 participants de compétences musicales variées, nous avons récolté un ensemble de 1126 descripteurs différents. Or ce nombre se rapproche de l'ordre de grandeur du nombre de mots nécessaires pour parler un français courant. De plus, seuls 14% du vocabulaire récolté pouvaient être considérés comme relevant directement d'un champ sémantique lié au domaine sonore. Il s'agit essentiellement d'un vocabulaire utilisé en musique (attaque, aigu, timbre...), de noms de sons (son, écho, sonneries...), de noms d'instruments de musique (corde, trompette, instrument...), mais aussi de noms d'objets dont la fonction première est de produire des sons (sonnette, Klaxon...).

Il semble donc que la majorité du vocabulaire recueilli n'est pas spécifique du domaine sonore, bien que ce soit le même que celui qui est utilisé par les musiciens (Berlioz, 1855 (1987); Scherchen, 1953; Piston, 1955; Berlioz, 1993) ou celui relevé par Frize (2000) dans la littérature française. Ceci rejoint aussi les remarques de David (1997), soulignant le peu de termes lexicalisés en français aussi bien pour l'olfaction que pour le domaine sonore.

#### Un vocabulaire non pertinent ?

De plus, le vocabulaire utilisé spontanément pour décrire des sons n'est pas forcément considéré comme pertinent, même par leur locuteur.

En effet, il est extrêmement inattendu d'observer qu'en moyenne, chacune des échelles sémantiques construites à partir du vocabulaire le plus fréquent spontanément n'est jugée pertinente que par 62% des sujets (dont 64% des anciens sujets qui ont pourtant produit spontanément ces termes). De plus, si on considère uniquement les sujets ayant participé aux deux expériences, on peut remarquer que leurs jugements de pertinence des échelles proposées ne coïncident pas avec leur propre utilisation spontanée de ce même vocabulaire pour décrire ces mêmes sons. En effet, si on trouve des mots à la fois utilisés spontanément et considérés comme pertinents par un grand nombre de sujets, on observe aussi des échelles considérées comme pertinentes par moins de 50% des sujets bien que plus de 70%

d'entre eux les ont utilisées spontanément. A l'inverse, certains termes peu employés spontanément peuvent paraître au contraire globalement adapté pour beaucoup de sujets.

### Impossibilité de décrire une perception ?

Doit on en conclure qu'il n'y a pas forcément de mots pour exprimer une perception, que le son fait partie de l'ineffable et doit être laissé aux poètes, Le son serait-il donc, tout comme la manne céleste, impossible à nommer ? "La première fois qu'ils virent la manne, les Hébreux furent bien évidemment étonnés. C'était quelque chose de menu, de granuleux, de fin comme du givre sur le sol. Lorsque les Israélites virent cela, ils se dirent l'un à l'autre : Qu'est-ce cela ? Cette interrogation (...) demeura sans réponse et les Hébreux durent apprendre à accepter qu'une réalité existe sans recevoir de nom, résistant ainsi au besoin de nommer et de comprendre toute chose. La Manne, c'est l'effacement préventif d'un nom qui vient enfermer les objets et les êtres dans la certitude illusoire de posséder un outil qui nous offrirait le monde." (Payeur, 1995, p 217).

### Utilisation commune de ce vocabulaire

On remarque que lors de l'utilisation des échelles sémantiques, les corrélations inter-sujets montrent que les différents participants en font une utilisation similaire. De plus, les échelles sont utilisées de la même façon que spontanément. En effet, si on compare d'une part les positions relatives des paires de sons sur les échelles et d'autre part la proportion de verbalisations libres "A est plus XXX que B", 19 des 23 échelles sémantiques sont significativement corrélées à la proportion de verbalisations "A est plus XXX que B".

De même, on observe de bonnes corrélations inter-sujets dans l'utilisation des portraits. Sur l'ensemble des 201 portraits proposés, on obtient 78% de Réponses Communes (RC) et 39% de Choix Communs (CC). Ceci est d'autant plus fort quand il s'agit de portraits décrivant la source sonore (sur 110 portraits décrivant la source 84% RC et 40% CC contre 72% RC et 38% CC pour les 31 portraits décrivant le son).

Il semble donc que ce vocabulaire, si difficile à formuler et jugé si imparfait, est pourtant utilisé de façon similaire en situation contrainte de positionnement des sons sur des échelles sémantiques ou d'association aux sons. Et ce, même par des participants nouveaux qui n'ont pas été les locuteurs directs des descriptions testées.

## **II.2. Sans vocabulaire spécifique, on se comprend**

Est-ce pour autant que ce vocabulaire permet de comprendre de quoi un autre a parlé ? L'enjeu de notre troisième expérience était bien de répondre à cette question. Si on demande après coup à d'autres participants quel son a été décrit avec une locution donnée, est-ce qu'ils peuvent retrouver les sons dont les locuteurs avaient parlé ainsi ? Nos résultats nous permettent de répondre par l'affirmative.

En effet, si on compare pour chaque "portrait" testé, d'une part la répartition de l'utilisation spontanée d'une description sur les 12 sons entendus et d'autre part le nombre de sujets lui ayant associé chacun des sons, on obtient en général des répartitions identiques.



On peut donc questionner cette correspondance avec les sons à l'origine des descriptions, correspondance inattendue dans la mesure où le vocabulaire semblait peu adapté à cette perception.

#### Le sens des mots est ajusté dans le contexte ?

Si on considère les commentaires écrits des participants sur leur utilisation des échelles sémantiques au terme de notre deuxième expérience, on peut remarquer que plus une échelle est globalement considérée comme non-pertinente par un grand nombre de sujets, plus ceux-ci ajoutent des commentaires pour nuancer leur propos. Ainsi par exemple l'échelle liée à *fort*, considérée globalement comme pertinente par le moins de sujets (23%), suscite de nombreux commentaires du genre : "fort dans la présence sonore, l'attaque", ou encore "évoque la puissance et la persistance du son".

Il semblerait donc que les sujets essayent de définir une acception ou un contexte qui permettrait d'utiliser quand même ce mot, une sorte de mode d'emploi possible dans ce contexte. Comme s'ils pouvaient à la fois témoigner du fait que l'utilisation de ce mot dans ce contexte n'était pas "normale" ou conventionnelle, mais qu'ils pouvaient néanmoins comprendre comment un autre locuteur avait pu l'utiliser et dans quel sens spécifique.

#### Empathie ?

La tâche de description des sons demandée était inhabituelle et difficile à réaliser pour les participants. Pourtant, même si les locuteurs semblent insatisfaits par leurs propres descriptions, placés dans le même contexte qu'eux, d'autres participants semblent comprendre et retrouver facilement de quel son ceux-ci parlaient.

Ceci pourrait présenter une mise en évidence expérimentale du fait que notre langage n'est pas uniquement référentiel mais que le sens qu'on attribue aux mots peut se construire en contexte à partir de la nécessité de communiquer quelque chose. En effet, le vocabulaire utilisé par quelqu'un d'autre que soi mais dans le même contexte (le même ensemble de sons, donc par comparaison, les mêmes références) peut se comprendre si c'est celui que l'on aurait produit soi-même dans son cas. On retrouve en effet une idée que l'interprétation pragmatique des énoncés repose sur l'empathie (Récanati, 1979). Si l'autre est comme moi, je peux comprendre ses productions, il suffit que je me mette à sa place pour comprendre pourquoi il a utilisé ce mot. On pourrait par exemple imaginer un "dialogue intérieur" d'un des participants : "Aucun des sons que je suis en train d'écouter n'est objectivement 'bleu', mais si quelqu'un (qui aurait pu être moi) a choisi de décrire un des 12 sons que je peux entendre tout comme lui comme 'bleu', alors cela ne peut être que celui-là".

Autrement dit, ce n'est pas tellement le mot qui est compris, ou le son qui est reconnu, mais c'est **le lien** fait par un autre (humain comme moi) entre un son et un mot qui est reconnu.

On pourrait tester cette idée de manière encore plus extrême : chacun des hybrides a été nommé par leur créateur avec des noms plus ou moins "parlant", par exemple, dans notre ensemble : *striano* ou *obochord* qui nécessitent de connaître *string* et *harpsichord* qui sont anglais, mais aussi *vibrone*, de vibraphone et trombone. On devrait pouvoir prédire qu'un grand nombre de sujets associeraient le son VBN au portrait "vibrone", même si c'est un néologisme. On pourrait aussi demander à des locuteurs

de baptiser chacun des sons selon leur fantaisie et vérifier auprès d'un autre panel de participants si ces noms ont été choisis de manière suffisamment conformiste pour être compris par un autre membre du groupe.

### Mots, non-mots et néologismes

C'est ainsi qu'on peut comprendre les néologismes inventés par les enfants (comme *pesette* pour une balance) qui nous font rire parce que nous comprenons à la fois leur côté incongru et le sens que l'enfant veut leur donner. L'apparition de ces néologismes pourrait correspondre à la compréhension par l'enfant d'une logique interne au langage : par exemple le fait qu'on construit des mots dérivés d'autres ou que certains mots peuvent être utilisés dans des contextes différents.

### Faire résonner du sens chez l'autre

Nos résultats s'inscrivent donc dans l'étude de l'aspect *collectif* de la faculté de langage ou du moins intersubjectif. Cet aspect collectif du langage est aussi abordé dans la linguistique historique, pour laquelle le langage d'un groupe est considéré comme une forme de mémoire collective de celui-ci (Nyckees, 1997). On comprend donc que l'émergence et l'évolution du lien entre les mots et les choses, vu comme convention linguistique au sein d'une population, peut ainsi être modélisée de manière informatique (Steels & Kaplan, 1999). On comprend ainsi l'assertion de Schaeffer (1966) p 481 : "quand on écoute un son, on est en présence de la société (qui donne le nom de la chose écoutée)".

Si même sans vocabulaire précis ou spécifique on se comprend, c'est que nos mots peuvent se plier aux choses d'une façon qui soit prévisible ou compréhensible par les autres utilisateurs de cette langue dans un contexte donné. Pour reprendre Bergson (1941), "le mot est mobile (...) il chemine d'une chose à une autre" et c'est cela même qui fait qu'il nous permet d'exprimer la pensée... ou la perception.

On peut ainsi concevoir la communication, à l'instar de Karl Pribram, comme un moyen de faire résonner du sens chez l'autre, à la limite peu importe le véhicule matériel utilisé. Et pourtant, il y a des véhicules qui sont plus facilement "marqués" pour communiquer une chose plutôt qu'une autre. Utiliser un code (donc figé) mais dans une situation toujours nouvelle et particulière, n'est-ce pas très exactement notre utilisation du langage. Cet état de chose subit un effet grossissant dans le travail des poètes, qui selon Eco (1999) "font de l'ambiguïté substantielle du langage la matière même de leur travail. Ils cherchent à exploiter cette ambiguïté pour en faire sortir, non pas un surplus d'être mais un surplus d'interprétation. (...) A travers cette réinvention permanente du langage, [ils] nous incitent à reprendre à chaque instant le travail d'interrogation et de reconstruction du monde..." (p 38).

Cette idée d'utiliser l'ambiguïté du langage est d'ailleurs professée par Verlaine (1938) dans son "Art poétique" :

"De la musique avant toute chose,  
Et pour cela préfère l'Impair  
Plus vague et plus soluble dans l'air,  
Sans rien en lui, qui pèse ou qui pose.

Il faut aussi que tu n'aïles point  
Choisir tes mots sans quelque méprise :  
Rien de plus cher que la chanson grise  
Où l'Indécis au Précis se joint.

C'est des beaux yeux derrière des voiles,  
C'est le grand jour tremblant de midi,  
C'est, par un ciel d'automne attiédi,  
Le bleu fouillis des claires étoiles !

Car nous voulons la nuance encore,  
Pas la couleur, rien que la nuance !  
Oh ! la nuance seule fiancée  
Le rêve au rêve et la flûte au cor ! (...)"

### III. Notre perception du monde

Nous allons maintenant essayer de discuter l'apport d'un autre ensemble de verbalisations recueillies dans ce travail. Il s'agit de toutes les expressions qu'on peut considérer comme métaphoriques. Elles sont en effet nombreuses et nous semblent révéler une partie d'un vaste réseau qui associerait langage / mémoire / action / perceptions multimodales / émotions...

Après avoir rappelé l'ampleur de ces expressions métaphoriques, discuté de leur origine synesthésique possible, et envisagé la structure de la mémoire qui permettrait de telles verbalisations, nous essayerons d'élaborer un modèle du sens compatible avec nos données.

#### III.1. Métaphores

Quand il s'agit de décrire les qualités du son, les locuteurs empruntent du vocabulaire à d'autres domaines sensibles (qualifiant par exemple les sons de *brillant, clair, chaud, riche, doux, sec...*). Ces descriptions des qualités des sons représentent 75% des verbalisations, et se répartissent comme suit : 56, 20 et 18% dans les domaines sensoriels respectivement auditif, tactile ou visuel.

Le domaine de l'audition est principalement représenté par la notion de "hauteur" (*aigu, grave, tonalité, harmoniques...*) et de qualités comme *sourd, nasillard, sonore*, auxquelles on a joint les verbalisations de *timbre, écouter, entendre...* Le domaine du toucher est principalement représenté par les notions de "texture" (*doux, piqué*), de "température" (*chaud, froid*), ainsi que des opposés *plein / creux* et *dur / mou*. Le domaine de la vision est principalement représenté par les notions de "luminosité" (*clair, sombre, brillant*) et de *couleur* (7 noms de couleurs différentes cités). Enfin, le domaine du goût est essentiellement représenté par les notions *acide* et *fade*.

L'ensemble des verbalisations relevant de perceptions d'autres sens serait encore plus conséquent si on y joignait toutes les descriptions qui renvoient au sens kinesthésique, par exemple les verbes d'actions, ainsi que les gestes effectués.

### Vocabulaire des musiciens

Il est intéressant de remarquer que c'est ce même vocabulaire qui est utilisé dans les traités d'orchestration ou de direction d'orchestre (Scherchen, 1953; Piston, 1955; Berlioz, 1994/1841). Un vocabulaire courant, flou, éminemment polysémique, applicable à un grand nombre d'expériences : *doux* peut par exemple se dire d'un goût, d'une lumière, d'une impression tactile, d'un son, d'un sentiment, d'un moment de la journée...

C'est aussi le vocabulaire mis au jour par Schaeffer dans son Traité des objets musicaux (1966). Il classe (p 507) l'ensemble de ce vocabulaire dans les quatre catégories qu'il a constituées dans son travail :

"1- Nous ne retrouvons pas bien entendu, des termes typologiques, qui correspondent à une abstraction du sonore toute artificielle. En revanche, nous avons choisi nous-mêmes des termes analogiques pour la pourvoir : cellule, échantillon, accumulation, etc. 2- Rond, pointu, gros, petit, menu, fin sont autant de termes morphologiques, aussi bien que tendu et plat. Ils désignent tout autant des formes que des *factures*, précisées par incisif, appuyé, perçant, etc. 3- Métallique, claironnant, criard, tout comme la tautologie *timbré - détimbré*, évoquent des genres de sons, en insistant sur l'allusion instrumentale. 4- Clair et sombre, comme chaud et froid, sont évidemment des analogies qui recherchent un équivalent des dimensions musicales dans les registres de la couleur ou de la température. Ainsi la *hauteur* et le *niveau* des physiciens sont aussi des analogies spatiales ; l'intensité et la force, des analogies kinesthésiques."

### Couleurs sonores

On peut aussi comparer les verbalisations recueillies aux études sur l'audition colorée (Karwoski & Odbert, 1938; Odbert et al., 1942; Pecjak, 1970). Le principe de ces études est de faire écouter des extraits d'oeuvres musicales et de demander aux auditeurs à quelles couleurs cela leur fait penser. On peut y retrouver le principe des Correspondances de Baudelaire ou des Voyelles de Rimbaud (A Blanc, I Rouge, O Bleu...). La notion de timbre étant très souvent liée à celle de couleur sonore chez les musiciens. Ainsi par exemple : "Les quatre bois peuvent *colorer* une succession d'accords selon le timbre des flûtes, des hautbois, des clarinettes ou des bassons" p134 (Scherchen, 1953). Schönberg, intitule même une de ses cinq pièces pour orchestre op. 16, *Farben* (Couleurs). Et Messian déclare que "la couleur du son est au coeur de ses préoccupations" (Surrans, 1993, p 55).

Cette idée d'associer des couleurs aux sons, peut se voir comme le reflet d'un réseau d'associations habituelles. Par exemple, pour Ponge le jaune est acide, parce que jaune -> citron -> acidité. Mais cette association peut être véritablement vécue comme des synesthésies, c'est-à-dire un rapport physiologique son / couleur. Messian décrit ce phénomène comme le fait : "lorsque j'entends de la musique et aussi lorsque je la lis, de voir intérieurement par l'oeil de l'esprit, des couleurs qui bougent avec la musique." (Surrans, 1993, p 71). Il indique d'ailleurs parfois ces correspondances sur ses partitions.

Dans les verbalisations spontanées recueillies à propos de nos 12 sons, on observe la présence de sept noms de couleurs différents (bleu, rouge, vert, jaune, orange...). Ceux-ci ont été principalement

produits par trois participants différents, tous néophytes, alors que le mot *couleur* est utilisé plus souvent, autant par des musiciens que par des non musiciens. On peut donc considérer que l'idée de parler de la couleur des sons, mais de façon relativement abstraite de tout nom de couleur réel, fait partie d'un vocabulaire "courant", notamment celui des musiciens. Par contre le fait de témoigner véritablement de ce que les chercheurs ont appelé l'audition colorée ne concerne ici que trois participants, soit 15% d'entre eux.

#### Lumières, goût et impressions tactiles

Mais le principe de synesthésie peut prendre des proportions plus grandes : l'expérience du goût lors de la stimulation auditive par exemple. Luria (1970) rapporte de nombreux comportements de type synesthésique dans son étude de celui qu'il a appelé le mnémoniste : "Le phénomène synesthésique apparaissait chaque fois qu'on donnait à Veniamin à entendre certains sons" (p 28) qu'il décrit de la façon suivante : "Chaque son faisait naître spontanément une impression de lumière et de couleur, des sensations gustatives et tactiles," (p 29) ou encore "le son et la voix possèdent chacun sa couleur et son goût, ils font surgir des *nuages de vapeur*, des *éclaboussures*, des *tâches*; certains sont blancs et lisses, d'autres sont oranges et acérés comme des flèches" (p 72). On pourrait citer quelques exemples des productions du mnémoniste : "votre voix est jaune et friable" (p 29), ou "la sonnerie retentit... une petite boule bondit, je sens sous mes doigts quelque chose de rêche, comme une ficelle, puis un goût d'eau salée... et une blancheur..." (p 69).

Si les descriptions des sons produites par le mnémoniste reflètent de façon systématique des emprunts aux domaines de la vision, du toucher et du goût, on n'y trouve, pas plus que chez nos participants, de référence à l'odorat. En effet, les verbalisations recueillies contiennent elles aussi du vocabulaire emprunté aux domaines de la vision, du toucher et du goût (respectivement dans 20, 18 et 2% des protocoles), mais pas d'emprunt au domaine de l'odorat.

#### Hauteur sonore

De plus, on peut, comme Schaeffer (1966) considérer que décrire la hauteur d'un son est un emprunt au domaine de la localisation. En effet, l'idée d'associer *haut* et *aigu* n'est pas forcément évidente, même si cet usage est tellement courant qu'il nous en fait oublier l'arbitraire. Pourtant, on rencontre cette association aussi bien dans les langues latines que dans les langues arabes et chinoises. De plus, on peut l'observer chez de très jeunes enfants (Wagner et al. 1981, cité dans Hamilton 1998), qui semblent capables, dès 9 mois, d'associer l'image d'une flèche qui pointe vers le haut avec une mélodie dite "ascendante" (les tons deviennent plus aigus). Pourtant, certains formateurs en initiation musicale remarquent que si la "position" relative des sons les uns par rapport aux autres est intuitive pour les enfants, ce n'est pas forcément le cas du sens de l'axe (montée ou descente), qui doit leur être donné. Willems (1985) note ainsi que parmi les enfants qui ne peuvent évaluer la hauteur des sons, certains n'ont pas la moindre notion de l'acuité sonore alors que d'autres confondent seulement le haut et le bas et d'autres encore ne semblent distinguer que les grandes différences. Il considère que les mots *haut*, *bas*, *grave*, *aigu* correspondent à des abstractions et que l'éducateur musical doit aider l'enfant à concrétiser ces notions. Ainsi, une enfant qui confondait les sons haut et bas l'expliquait

ainsi : "c'est haut parce que c'est gros, et puis parce que c'est fort (...) comme une maison... et puis un éléphant" (ibid. p 95). Pour corriger cette association fautive du point de vue de la musique, il a demandé à l'enfant de tendre la main et y a déposé un livre gros et lourd en lui faisant remarquer qu'en musique tout comme avec ce livre, ce qui est grand, gros et lourd va vers le bas.

Dans le cas de la description de nos sons, l'importance et le côté systématique de leur comparaison en terme de "aigu / grave" (respectivement 98 et 80% des locuteurs ont utilisé spontanément ces qualificatifs) apparaît comme encore plus étrange dans la mesure où ce n'est pas la hauteur tonale qui est jugée (tous les sons étant présentés sur la même note). Les fortes corrélations entre les utilisations des échelles sémantiques correspondantes avec celles construites à partir de "haut / bas", ainsi que les corrélations entre ces échelles et le centre de gravité spectral des sons montrent pourtant que "aigu" et "grave" sont bien utilisés dans le sens d'une "localisation" haute ou basse des partiels perçus dans chaque son. Cette utilisation indique d'une part que nous sommes capables de percevoir la localisation des fréquences les plus importantes dans des sons produits sur la même note. D'autre part, cette utilisation rejoint celles répertoriées par Hamilton (1998), qui remarque que les particules *up* et *down* participent d'une série de métaphores dans lesquelles une échelle de positions verticales se trouve projetée sur une échelle d'autres valeurs : quantitatives ou affectives par exemple. Ceci induit qu'en général, "la particule *up* a une valeur méliorative, augmentative, positive ou d'achèvement" (p 35). Lakoff (1987) attribue ce groupe de métaphores à plusieurs phénomènes dont la kinesthésie : "un état d'esprit positif a un effet sur la posture : tendance à se tenir plus droit (à la verticale), à porter la tête haute, etc." (ibid.). Dans le cas des sons, les qualifications hautes et basses peuvent être liées aux différentes parties du corps qui résonnent quand on produit des sons de différentes hauteurs, ce qui rejoint une remarque d'un des participants : "ils ont l'air d'être plus aigus, enfin, c'est pas vraiment aigu, mais, on a l'impression qu'on les entend, dans une partie de l'oreille qui est différente, (...) c'est comme si on les entendait un peu plus haut dans la tête".

On peut tout de même noter que dans le cas de la qualification des timbres, c'est "grave" et non pas "aigu" qui est associé à chaleureux et agréable et qui a donc le sens positif. On peut relier à ceci le fait que les bébés enrichissent leurs pleurs d'harmoniques aiguës quand ils veulent leur donner la valeur communicative d'expression de désagrément (Cyrulnik, 1995). Cette modulation par les bébés des fréquences de leur voix est peut-être aussi due à leur expérience intra-utérine : "quand la mère parle, le bébé perçoit les basses fréquences qui émanent de cette parole, filtrée par la poitrine, le diaphragme et l'utérus. La voix de sa mère lui parvient lointaine, douce et grave" (p 65).

### **III.2. Perception**

Si on considère l'ensemble des verbalisations de qualités de type auditives, visuelles, tactiles et du domaine du goût ainsi que celles décrivant la "forme" des sons, (qui relèvent de la vision ?), les verbalisations de ce type sont présentes dans 82% des verbalisations de nos participants.

Ce "feu d'artifice verbal" de termes analogiques, artistiques ou physiologiques cache en fait une impossibilité dénoncée par Schaeffer (1966), comme "l'impossibilité foncière où nous sommes de

décrire un objet *en soi*, hors de toute structure. Si nous refusons de référer ce son, cette voix à d'autres objets sonores, point d'autre solution que d'aller chercher des références ailleurs, dans d'autres domaines de la perception ou de la pensée : la vue, le sens kinesthésique, le toucher, etc." (p 506). Et si cette impossibilité foncière était en fait un reflet de notre perception du monde, qui loin de capter chaque stimulus par un canal sensoriel différent et indépendant des autres, ne pourrait percevoir que des situations globales à travers l'ensemble de ses sens (vue, ouïe, toucher, etc.) et de leur sens (quel sens à pour moi cette situation ?).

### Métaphores ou synesthésie

On peut se demander si ces emprunts à des champs sémantiques *a priori* différents du domaine sonore sont de l'ordre de la synesthésie, de l'analogie ou de la métaphore. Autrement dit, si un auditeur décrit un son comme *fade* ou *brillant*, est-ce parce que ce son entraîne chez lui la perception d'un goût fade ou d'une impression visuelle brillante, ce qui relèverait du phénomène de la synesthésie ? Est-ce plutôt parce que ce son a "réveillé" chez lui toute une chaîne d'associations en mémoire où cette perception sonore serait associée à une situation comprenant une perception gustative ou visuelle, ce qui relèverait plutôt de l'analogie ? Ou bien est-ce que ce n'est qu'un effet de langage : le mot *fade* ou *brillant* étant associé de manière principale à un ressenti de l'ordre respectivement du goût ou de la vision, mais peut être associé à toute autre perception qui comporterait des traits communs avec ces "idées" de fade ou de brillant ?

On rejoint ici des questions présentes dans le champ de recherche de la compréhension de la métaphore (Tijus, 1996; Pudelko, 1999), les différences de définition entre analogie et métaphore pouvant être présentées à partir des deux exemples suivant : "l'essence est à la voiture ce que le lait est au bébé" est une analogie qui correspondrait à la métaphore : "la voiture a soif" (Tijus, 1996, p 94).

### Des modalités sensorielles indépendantes

Le fait d'obtenir des descriptions globales faisant appel à l'ensemble de nos sens peut mener à remettre en question l'évidence de modalités sensorielles indépendantes. Cette idée dérive de l'idée d'envisager l'organisation du cerveau en modules indépendants les uns des autres (Fodor, 1983). Pourtant, la modularité n'exclut pas la coopération entre différents modules par synchronisation entre différentes populations de neurones par exemple. Elle peut sembler "raisonnable" en ce qui concerne les aires primaires associées à chacun de ces sens. Il est pourtant intéressant de rappeler ici que dans le cas de la vision, 80% des informations qui arrivent au CGL viennent du cerveau et non pas du nerf visuel (Varéla 1989). Mais dès qu'il est question de donner un sens à ces perceptions, l'indépendance ne paraît plus du tout évidente. Elle n'est même peut-être plus qu'un reflet de l'organisation de la recherche en laboratoires ou équipes de spécialités différentes !

Ainsi, Claude Cadoz (cité par Risset, 1994) considère que l'apprentissage auditif est inséparable des autres modalités sensorielles (sonores, visuelles et tactiles) dans le cadre de ce qu'il appelle l'expérience instrumentale.

### Sons de parole - Reste de l'audition

Cette conception modulaire mène aussi souvent à considérer que les sons prononcés à but de langage sont perçus de manière totalement différente des autres sons. On trouve ainsi beaucoup d'études montrant une différence de latéralisation hémisphérique si les sons perçus sont du langage (à gauche) ou de la musique (à droite) (Melher & Bertoncini, 1979). S'il semble indéniable que les sons non verbaux n'activent pas les aires de Broca, on peut pourtant considérer qu'une grande partie du traitement des différents sons est identique. Ceci peut d'ailleurs amener un autre éclairage sur les données verbales. Ainsi, par exemple, la plupart des études sur le processus menant "du son au sens" pour les sons de langage (Segui & Ferrand, 2000) considèrent que le sens d'un mot est atteint par un auditeur quand celui-ci est capable de le répéter. Il serait ainsi évident qu'une fois qu'on a un mot on a son sens (coté irrépressible de cet accès au sens) mais aussi que le sens du mot est le mot lui-même. Par exemple, le sens de la perception auditive résultant de la prononciation de "coucou" serait ainsi atteint quand l'auditeur aurait reconnu le mot "coucou". Pourtant, le sens de "coucou" pourrait être : "alors, surpris de me trouver ici ?" ou pourrait aussi faire référence à l'oiseau du même nom, ou à son chant qui annonce le printemps. Le fait de n'étudier souvent que des mots isolés mène en effet à évacuer le phénomène de polysémie. Ces études ne prennent donc pas en compte les résultats des linguistes pour qui le sens d'un mot dépend de ses voisins en contexte. Transposé dans le domaine de l'audition non verbale, où on observe aussi un accès irrépressible à la source productrice du son (son sens ?), ce genre de résultat correspondrait au fait de considérer que le sens de "schting", serait de pouvoir imiter "schting", alors que les auditeurs disent : "un clavecin" ou "on pince une corde" ou "un son très bref et aigu"... Le fait qu'on ne peut rendre compte du sens sans faire appel aux mots du langage apparaît ainsi comme un handicap dans la compréhension de celui-ci.

### Perception pure ?

Nous allons voir dans quelle mesure les résultats obtenus dans ce travail nous permettent de discuter la notion d'indépendance entre perception et langage, c'est-à-dire entre ce qu'on perçoit et ce qu'on en dit, pense, attend...

En effet, si on considère les résultats obtenus dans le cadre de la recherche de corrélats verbaux de dimensions perceptives, on rappelle que les verbalisations n'altèrent ni la structure, ni la nature de l'espace perceptif tel qu'il peut être modélisé par analyse multidimensionnelle. Pourtant, il a été impossible de trouver une structure d'adjectifs qui pourrait se superposer aux dimensions perceptives. En effet, l'aspect composite des attributs verbaux vis à vis de ces dimensions perceptives nous mène à concevoir un non-parallélisme direct entre représentations sémantiques et représentations perceptives, ce qui va dans le sens d'une indépendance entre ces deux niveaux de représentation (ou d'une relation complexe entre les deux).

Pourtant, lire une description semble influencer notre manière de la percevoir. Ce résultat apparaît notamment lors de l'analyse des choix d'association de sons aux portraits dans l'expérience 3 (chapitre 8 ?). On observe en effet le fait que, quand un son a fait l'objet de plusieurs descriptions différentes et disjointes (et donc *a priori* incompatibles), les sujets associent le même son à chacun de ces portraits. Ainsi, par exemple pour le son de *vibrone*, les descriptions libres nous ont permis de construire plusieurs portraits dont "on étouffe le son" et "on percute une peau". Or respectivement



75% et 85% des sujets ont associé à ces portraits le son de vibrone dans les deux cas, en gardant l'impression bizarre de passer leur temps à être influencés (ballottés) par ce qu'ils lisent. A chaque nouvelle description, ils sont nombreux à dire : "mais oui, c'est sûr, c'est ce son là !" et si, peu après, on leur propose une autre description différente du même son, ils disent à nouveau : "mais oui, c'est bien ça, je me suis trompé avant". Plusieurs sujets ont décrit l'impression bizarre ressentie en effectuant cette expérience par "il semble que le terme influence notre perception" ou par "j'ai vraiment l'impression que les sons ne sont pas les mêmes en fonction du mot donné, comme si le mot imposait un éclairage particulier, et faisait scintiller une partie différente du même son".

Ces résultats semblent montrer qu'une perception pure n'existe pas. Ce qu'on perçoit est modifié par nos attentes, notamment par la lecture d'une description langagière. Ceci rejoint l'interrogation : "Peut on regarder sans mots ? (...) Avec quels yeux regardons nous ? Les seuls yeux des mots ? Les seuls yeux de la mémoire ?"

### **III.3. Mémoire**

Nos attentes étant un reflet de nos connaissances préalables, c'est-à-dire de notre mémoire, on peut essayer de comprendre comment nos résultats peuvent apporter des arguments à propos de la structure de la mémoire. Pour Bergson (1919) : "A celui qui aborde, sans idée préconçue, sur le terrain des faits, l'antique problème des rapports de l'âme et du corps, ce problème apparaît bien vite comme se resserrant autour de la question de la mémoire, et même plus spécialement de la mémoire des mots". On peut ainsi reposer ici la question de savoir si la perception peut s'affranchir de la mémoire.

Les descriptions synesthésiques du mnémoniste de Luria (1970) ont été attribuées au fonctionnement anormal de sa mémoire qui ne pouvait s'empêcher d'enregistrer chacune des situations perçues de manière spécifique et indélébile (au risque de gêner les processus de catégorisation qui demandent de regrouper comme identiques des événements différents). Les observations faites sur le mnémoniste nous permettent peut-être un effet grossissant sur la façon dont les gens normaux fonctionnent. La conjonction temporelle de stimulations de l'ensemble de nos sens fixerait des associations en mémoire. Chaque perception étant complexe et ne relevant pas d'un seul canal sensoriel, des percepts de chacun des différents domaines sensoriels sont reliés entre eux et les mots utilisés pour les décrire contiendraient des nuances relevant des différents sens...

#### Une mosaïque pour mémoire

Lorsqu'on a commencé à étudier les parties du cerveau activées lors de l'utilisation de la mémoire se rapportant à différentes classes d'objets, on a pu observer que la mémoire des outils active des zones motrices, celle des objets "visuels" active les zones visuelles, celle des objets sonores active les zones sonores. Doit-on en conclure que la mémoire est morcelée et distribuée dans toutes les zones du cerveau ?

Ceci rejoint l'hypothèse de Damasio (1989) pour lequel on aurait des localisations différentes en mémoire en fonction de sémantiques différentes. Ainsi, par exemple, les verbes d'actions seraient

stockés dans des zones proches de la zone motrice. Sachant qu'en imaginant effectuer un mouvement on active des zones du cerveau semblables à celles activées lors de l'exécution réelle du même mouvement (Berthoz, 1997), il se pourrait que le fait d'énoncer un verbe d'action reviendrait à activer la représentation mentale d'un mouvement. Martin et al. (1996) obtiennent des résultats TEP allant dans ce sens : le fait de nommer des images actives des zones différentes si celles-ci représentent des animaux, des outils ou des figures de points répartis aléatoirement. Dénommer des outils activant notamment les aires prémotrices gauches activées aussi lors de l'imagination de mouvements des mains.

On comprendrait ainsi l'idée d'envisager la mémoire comme une collection de fragments de soi, dans le sens d'expériences passées alliant notamment perceptions et actions. La description de la perte d'une mémoire conçue ainsi a été imaginée par Makyo (1994) : "Ecoute, chaque souvenir, chaque pensée, chaque souvenir de pensée sont des parties de toi ! (...) Sous le choc de ces petites images qui me fascinaient, m'aspiraient, j'eus très vite une sensation extrêmement pénible de décomposition... Et ce n'étaient pas mes souvenirs ou mes pensées que je perdais.. mais... mes doigts.. puis.. mes mains, mes bras, mes jambes, mes oreilles, mes cheveux, mon nez... mon corps entier éclatait en morceaux pour ne laisser place qu'à une sorte de brouillard, d'obscurité où j'étais moi-même à ma recherche... et impossible de me retrouver..." (p 36).

#### De la mémoire en toutes choses

Paul Valéry affirmait : "ce qui ne ressemble à rien n'existe pas". Ceci peut nous permettre de comprendre la manière dont on utilise un mot qu'on connaît déjà quand on présente un objet nouveau, ce qui d'ailleurs peut modifier le sens de ce mot si l'usage se perpétue. Ainsi, la façon dont on perçoit les sons hybrides fait appel à des modes de productions connus et envisageables par l'auditeur. On peut ainsi expliquer les descriptions du type : "le son de gargouillis est terrible, (...) un peu... comme un tromboniste qui fait une bêtise sur une paire de timbales" (voir Compact Disque en Annexe), face à un son inconnu.

Cette perspective nous permet aussi de replacer la notion d'abstrait, considérant que les idées les plus abstraites ont un contenu sensoriel (Filliozat, 1997). Cette idée rejoint les résultats de Sanders (1997) qui observe par exemple que la majorité des gens répondent "non" à la question : "Est-ce qu'un carré est un trapèze ?" Il en déduit qu'un concept mathématique abstrait n'existe pas : il est toujours réalisé dans un exemplaire particulier. Par exemple, nous avons en général une représentation d'un trapèze typique qui n'est pas carré. On comprend aussi de cette façon le fait que la résolution de problèmes de géométrie de secondaire passe par la construction d'une figure. C'est un moyen de rendre concret un problème abstrait. De la même façon, l'apprentissage du langage se fait toujours dans le concret. La notion de concret prenant un sens plus large à la lumière des résultats montrant des similitudes entre l'imagination d'un mouvement et sa réalisation réelle. En effet, on ne peut oublier de reconnaître à l'imagination humaine la capacité de créer purement et simplement des similitudes. Ainsi, on n'est pas obligé de se cantonner, pour expliquer des analogies à la notion de "similitude expérientielle" (due à l'expérience), mais y ajouter toutes les similitudes imaginables (Lakoff & Johnson, 1985, p 163).

### Vibrones, licornes, et autres chimères

On peut ainsi comprendre le fait que les chimères auditives que nous avons présentées à nos participants ont été entendues et "traitées" de la même façon que les autres sons se voulant plus vraisemblables. Ainsi, comme le souligne Risset (1994) : "rien ne tape sur rien pour produire un son de synthèse entendu comme *percussif*(...) [même si l'oreille] tend à rattacher un timbre entendu à un timbre qu'elle connaît, ou plus précisément dont elle connaît la genèse" (p 104).

Il nous paraîtrait normal d'envisager de la même façon que les chimères visuelles utilisées dans les tests neurologiques pour dépister une aphasie soient en fait traitées de la même façon que les images d'objets réels. Ainsi par exemple le dessin d'un "ciseau-tenaille" (Eustache & Lambert, 1998) réputé un non-sens, ne m'a absolument pas intriguée au premier abord et que j'ai eu à faire un effort pour trouver ce non-sens qui m'était exposé oralement, peu après cette vision. Je réalisais alors qu'en fait, j'avais accepté cet objet comme un outil possible de la trousse des chirurgiens comme d'une sorte de "ciseau pour tenir". N'est-ce pas un indice (introspectif certes !) à défaut d'une preuve que nous traitons les chimères de la même façon que les images d'objets réels ? Est-il alors possible d'utiliser ces chimères pour montrer l'existence de deux voies dans le mode d'attribution de sens aux perceptions visuelles ? Ne serait-ce pas au contraire très intéressant de considérer les dénominations de ce genre d'objets par les patients aphasiques comme une trace visible, dans ce cas bizarre, du mode d'"accès au sens" utilisé habituellement à notre insu ?

On traiterait les "non-choses" comme les "choses", soit en ne se rendant même pas compte que ce sont des chimères qui n'appartiennent donc pas à la réalité (ex : dessin de ciseaux-tenaille et autres licornes), soit en cherchant quel sens a bien pu nous échapper. Les verbalisations recueillies lors de l'audition de chimères comme le son de *obochord* montrent que les deux types de processus peuvent être observés chez les sujets : considérer *a priori* que c'est un son plausible et le décrire ("on percute une peau qui vibre") et rechercher une explication à son étrangeté ("c'est quoi ce truc ?").

### **III.4. Vers un modèle du lien percept-concept (ou du sens ?)**

#### Théorie de l'information

A l'origine, notre approche s'est inspirée de la théorie de l'information. Nous envisagions donc, lors de la description des sons, une propagation *séquentielle* et *hiérarchique* d'une information qui se transformait à chaque étape : de son, elle devenait représentation perceptive puis représentation sémantique, celle-ci activait un lexique qui était verbalisé par les sujets.

son -> représentation perceptive -> représentation sémantique -> lexique -> verbalisation
---

Figure 1

De la même façon, on pouvait concevoir que lors de la lecture d'une description en langage naturel on pouvait avoir une activation d'une représentation sémantique à travers le lexique et que cette

représentation sémantique était comparée à celle activée par les sons et conduisait au choix d'un ou plusieurs d'entre eux comme correspondant à cette description.

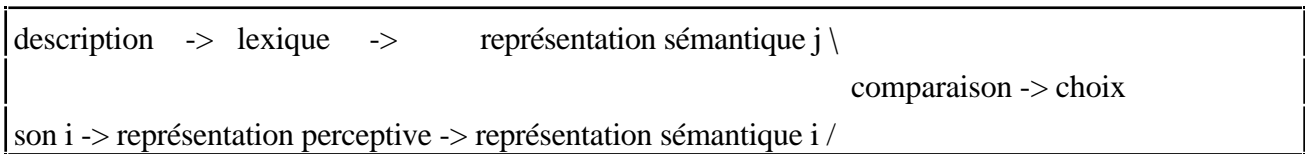


Figure 2

Ce type d'hypothèse sur le fonctionnement de notre cerveau pose beaucoup de problèmes. En particulier, cela sous-entend que les processus sont séquentiels. D'autre part, et bien que cela ne soit pas forcément indispensable à ce modèle, cette conception a tendance à sous-entendre que ce soit des sous-parties distinctes du cerveau qui traiteraient ces différents niveaux d'analyse. Ce qui sous-entend localisation fonctionnelle.

### Mosaïque perceptive

On constate que, si on leur fait écouter un son de marimba, par exemple, les descriptions produites par les sujets nous semblent refléter l'immensité des associations produites par la perception de ce son. Les descriptions vont être d'ordre motrices : "on frappe sur du verre" (dans certains cas, les sujets vont aussi mimer cette action), émotives "c'est un son chaleureux, très agréable", tactiles "doux, chaud", colorées "un son bleu"... Des souvenirs plus personnels peuvent aussi être évoqués... par exemple des situations, des moments de la journée... Bien que les conditions de l'expérience soient plutôt hostiles (cabine sombre, casque, longueur et aspect répétitif de l'expérimentation...) la richesse de ce que peut évoquer un son surprend au premier abord.

On a l'impression que la perception d'un stimulus (le son ici) active une multitude de zones du cerveau qui peuvent prendre l'apparence d'une mosaïque. Si on se réfère à des travaux sur la vision par exemple, on a montré que des neurones de régions différentes du cerveau étaient "chargés" de "répondre" à des inclinaisons de barres différentes, comme si chacun des traits d'un dessin étaient décomposés et traités en des lieux différents.

### Entrées - Sorties ?

Toute "entrée" dans ce système dynamique complexe (que ce soit un son, ou la lecture d'une description dans le cas de notre expérimentation, mais cela peut aussi bien être une stimulation venant de l'extérieur que de l'intérieur...) "bouscule" l'ensemble du système, ce qui l'amène à prendre une succession d'états liés à des émotions, des activités, des "idées", des "sens" qui peuvent être verbalisés...

Si on regarde à nouveau les résultats obtenus dans le domaine de la vision, on sait que 80% au moins des entrées dans les zones visuelles primaires ne viennent pas des yeux mais d'autres zones du cerveau (Varela, 1989). Ce que l'on perçoit est fortement influencé par ce que l'on connaît et ce que l'on attend. Il en est de même en ce qui concerne l'audition. Le timbre par exemple, serait "construit" après une décomposition fréquentielle de certains événements sonores mais ces événements sonores

sont organisés comme tels à partir de nos connaissances, de nos attentes, d'indices comme... le timbre ! (McAdams, 1993). Où est la poule et où est l'oeuf !

La perception ne peut donc plus être vue comme ayant un sens de l'extérieur vers l'intérieur de nous, mais plutôt comme une zone d'interface et de questionnement ou de mise en équilibre avec l'extérieur... Cette suite dynamique de furtifs états d'équilibre de l'ensemble des neurones de notre cerveau serait donc en constant "équilibre" avec nos zones d'interfaces, donc avec l'extérieur, ce dans quoi nous vivons et dont nous sommes une partie.

### Unité du sens

Pourtant, si la perception d'une "chose" active un grand nombre d'aires différentes dans le cerveau comme si cela effectuait une sorte d'analyse, qui éclaterait chacun des traits de ce signal pour le traiter à chaque fois dans une aire différente, nous avons bien conscience de percevoir un tout, une forme, une "chose", un objet, un son... A quoi donc servirait d'éclater en millions de morceaux s'il était ensuite nécessaire de reconstruire l'unité pour en comprendre le sens ?

Et si, au lieu de concevoir "une boîte" spécialisée dans la représentation sémantique et qui donnerait un sens à cette mosaïque éclatée, on essayait de penser que c'est justement CETTE mosaïque qui EST le sens de notre perception. Cette mosaïque serait comme une sorte *d'empreinte* sur l'ensemble de notre cerveau de ce que nous sommes justement en train de percevoir, cette empreinte étant à la fois d'ordre perceptive, motrice, langagière, émotionnelle... Et la mémoire serait juste une façon de relier entre elles certaines de ces zones dues à leur association lors de perceptions (ou imaginations) passées. Elle nous permettrait en quelque sorte d'enrichir chacune de nos perceptions nouvelles grâce à notre vécu antérieur qui serait à nouveau associé.

On rejoindrait ainsi la conception de Ninio (1989) : "Quelque part, dans une mémoire acoustique du cerveau, se trouve une représentation du mot bateau, qui permet de le repérer dans une conversation orale. Quelque part, dans une mémoire graphique, est représentée l'orthographe du mot. Autre part, dans une mémoire motrice, gutturale, sont représentés les ordres de commandes musculaires à accomplir pour prononcer *bateau*. Ailleurs, dans d'autres mémoires, sont conservées des traces de perceptions se rapportant aux navires : le mélange caractéristique d'odeurs de cordage, de mazout, de peinture fraîche et d'air marin dans la mémoire olfactive, le hurlement des sirènes dans la mémoire acoustique, des images de bateau dans la mémoire visuelle, des souvenirs de croisière dans la mémoire des épisodes. Le mot bateau lui-même, existe-t-il autrement qu'éparpillé en ses multiples représentations ? Est-il autre chose que cette toile d'araignée qui relie les unes aux autres toutes les parcelles de mémoire qui le concerne ?" (p 255)

A la différence près qu'il considère que c'est le mot bateau qui est une toile d'araignée, nous dirions plutôt que c'est son sens.

### Emergence d'une forme cérébrale ?

De nombreux résultats actuels, notamment en neuropsychologie, montrent que des activités considérées auparavant comme indépendantes activent en fait les mêmes circuits neuronaux. Par exemple, Rizzolatti a montré que ce sont les mêmes neurones qui sont activés lorsqu'un singe regarde

faire une action et lorsqu'il la fait lui même. (Il est intéressant de souligner que la localisation de ces neurones chez le singe se trouve dans la zone qui correspondrait à l'aire de Broca (c'est-à-dire du langage) chez l'homme). Il est étonnant de remarquer que, dès 1975, Krishnamurti disait que "voir et faire sont une seule et même chose" (Krishnamurti & Bohm, 1975/1999) en se basant sur une expérience dérivée de la pratique de la méditation.

De la même façon, certains ont montré que perception, simulation et action relevaient d'un même mécanisme (Berthoz, 1997), d'autres ont montré qu'imaginer et faire (cf. visualisations des sportifs de haut niveau) étaient la même chose, d'autres encore ont montré que se souvenir et percevoir activaient des zones communes du cerveau... (cf. expériences sur la mémoire sémantique)... Nous allons donc sans doute vers une conception plus unifiée de l'ensemble de nos activités.

L'idée de modularité de l'esprit a permis de comprendre beaucoup de choses mais ne peut exister sans le concept de boîte noire centrale impossible à élucider, et c'est justement dans cette boîte noire qu'elle place le sens. Faisons éclater cette boîte noire, elle ne sert à rien : l'apparence de morcellement des informations n'est qu'une apparence. De même qu'un trait de crayon de papier sur une feuille n'est qu'une multitude de petits éclats de graphites accrochés sur les aspérités du papier, de même l'apparent morcellement des activations des neurones de notre cerveau constituerait une "empreinte neuronale", une "forme cérébrale", dont le sens émerge sans avoir besoin d'être reconstruit.

#### Alternative aux boîtes et flèches...

Si on essaye de comprendre en quoi ces réflexions ont modifié la représentation de notre description d'un son selon la figure 1, nous avons tout d'abord transformé la flèche *son->représentation perceptive* en une double flèche *extérieur<->intérieur* qui représente une sorte d'état d'équilibre que nous entretenons avec notre environnement, puis nous avons questionné la flèche : *représentation perceptive -> représentation sémantique* et l'avons fait disparaître en fusionnant ces deux types de représentation pour considérer que ceci serait une sorte d'empreinte neuronale globale.

Il nous reste à nous attaquer à la flèche *représentation sémantique -> lexique* qui déjà a pris la forme *empreinte globale neuronale -> lexique*. La flèche *lexique -> verbalisation* ne sera pas directement questionnée ici, elle relève plutôt des travaux de psycholinguistique (pour une revue de ces travaux, voir Segui & Ferrand, 2000) .

## IV. Conclusion

Les verbalisations spontanées peuvent donc nous donner un bon panorama 1- de l'ensemble des termes qui viennent spontanément à l'esprit des gens pour décrire les sons ainsi que 2- de la manière dont les sujets résolvent la tâche demandée.

On retrouve notamment, à travers les verbalisations recueillies, les deux définitions du timbre : en tant que renvoi à la source ou comme décomposition analytique. En effet, il semble que la reconnaissance de la source soit immédiate, que les auditeurs aient tendance à rechercher un mode de production même quand le son est un hybride. De plus les adultes en manque du mot font le geste... qui pourrait produire le son. D'après la notion d'affordance de Gibson (1950), on peut d'ailleurs considérer qu'un objet contient les actions possibles sur lui. Toutes les descriptions de sources apparaissent ainsi comme des formes de descriptions d'actions ayant engendré le son. Ceci nous renvoie à la définition physique du son comme vibration issue d'une action sur un matériau.

Ce travail expérimental nous aura aussi permis de dégager deux résultats concernant le langage : 1- On parle avec des mots qu'on juge "pas appropriés" et 2- La communication est possible sans vocabulaire précis. Autrement dit, en contexte, on reconnaît le lien qu'un autre avait fait entre un son et un mot. Ainsi, nous pourrions dire, avec Schaeffer (1966), qu'"aussi bien, les mots que nous choisirons n'auront-ils pas plus d'importance que les symboles du solfège traditionnel : ce qui compte, c'est moins le contenu de chaque terme que le champ d'expérience qu'il délimite, par rapport aux autres termes. Ce vocabulaire est un mode de classement, dont le seul but est de diriger certains types de confrontation entre objets musicaux, et éventuellement d'en susciter d'autres. Ce n'est pas au niveau de la lecture ou de la dissension idéologique qu'on pourra juger de sa pertinence, mais au niveau de l'écoute. Les termes de tout solfège sont des chèques, dont notre expérience des sons fournit seule la provision" (p 508).

Mais l'ampleur du champ sémantique utilisé pour parler des sons nous montre aussi à quel point les différentes activités humaines sont difficiles à séparer. On ne peut donc plus concevoir les différentes activités de notre cerveau (et en particulier nos différents sens) comme indépendantes les unes des autres.

Ce travail nous a permis enfin de comprendre que le monde "réel" est justement ce pour quoi nos sens sont "accordés", doublé de notre façon matérielle de vivre (une "porte" n'a pas la même réalité pour nous, une fourmi et un éléphant), doublé d'une bonne couche de social, c'est-à-dire ce que les autres voient et ce qu'il est donc "possible" de voir sans passer pour un fou, marque de l'exclusion du groupe humain. A ce propos, maintenant que les "fous" commencent à parler (on fait écrire des autistes par clavier d'ordinateur interposé...), on commence à voir décrites d'autres parties de notre monde.

## V. Perspectives

Une nouvelle "jeune-pousse" (start-up) américaine ([www.dolphinsearch.com](http://www.dolphinsearch.com)) se sert de deux des résultats mis en valeur par ce travail pour constituer un moteur de recherche "intelligent", pour l'instant utilisé uniquement sur des réseaux intranet. Le fondateur s'est inspiré pour imaginer ce moteur de recherche du mode de reconnaissance des objets par le dauphin, qui utilise le spectre complet des sons réfléchis par les objets pour en déduire leur identité (c'est-à-dire le fait que le timbre contient l'identité de sa source). D'autre part, reconnaissant que le sens d'un mot est déterminé par tous ceux qui l'entourent dans un texte (son co-texte), ils déterminent le sens d'un mot dans son contexte par l'ensemble de son co-texte. Ceci permet donc une recherche par le sens et non par la forme de surface d'un mot. Le sens d'un mot serait donc assimilé à une forme plus complexe, multidimensionnelle, analogue du "timbre d'un objet" (comme raccourci du "timbre d'un son produit par un objet et y renvoyant"). L'ensemble des mots composant le co-texte d'un mot est considéré (et analysé) de la même façon que l'ensemble des fréquences composant le timbre d'un son. Si, comme le dauphin, on est capable de reconnaître un objet en analysant la forme globale du timbre qu'il renvoie, on devrait être capable de reconnaître le sens d'un mot dans un contexte donné en analysant la forme globale de son co-texte.

On peut ainsi envisager d'adapter les représentations sous forme d'arbres sémantiques utilisées pour modéliser le sens d'une situation, d'un objet ou d'une phrase (Tijus, 1996; Tijus & Moulin, 1997) à la modélisation du (ou de l'ensemble des) sens des mots. On aboutirait pour chacun d'eux à une sorte de schéma ou de noyau de sens. Ceci compléterait le travail effectué sur des verbes (Saunier et al., 1994) ou des "prépositions" (Victorri, 1999), dans une perspective proche de celle de Cadio et Nemo (1997) considérant que le noyau de sens de chaque mot repose sur des propriétés extrinsèques des objets auxquels ils renvoient (qui prennent donc en compte leur perception par un humain).



## Postface

"Et la morale de ceci est : Prenez soin du sens, les sons prendront soin d'eux-mêmes."

Lewis Carroll (1871 /1998)  
Alice aux pays des merveilles, p 132

"Quiconque saurait le secret usage des mots de tous les jours aurait un pouvoir illimité,  
et il ferait peur."

Jean Tardieux (1981)

## **Bibliographie**

*Il est sans importance pour chercher la vérité  
de savoir d'où provient une idée.*

Bouddha, Majjhima Nikaya.

Arnaud & Nicole (1970). La logique ou l'art de penser. Paris, Champs-Flammarion.

Asada, M. & Ohgushi, K. (1991). "Perceptual analyses of Ravel's *Bolero*." *Music perception* **8**(3): 241-250.

Ballas, J. (1993). "Common factors in the identification of an assortment of everyday sounds." *Journal of Experimental Psychology : Human Perception and Performance* **19**(2): 250-267.

Barbut, M. & Monjardet, B. (1970). Ordre et classification : algèbre et combinatoire. Paris, Hachette.

Baron-Cohen, S. (1995). Mindblindness : An essay on autism and theory of mind. Cambridge, Mass., MIT Press.

Bartlett, F. C. (1932). Remembering. Cambridge, Cambridge University Press.

Basso, P. (2000). "Théorie des processus de création d'objets à partir d'un milieu non-déterminé." *Intellectica* **29**.

Beaunis, B. (1918). "L'émotion musicale." *Revue Philosophique* **85**: 353-369.

Beguin, P. & Costermans, J. (1994). "Le traitement de l'information olfactive (revue critique)." *L'Année psychologique* **94**: 99-122.

Berger, K. W. (1964). "Some factors in the recognition of timbre." *Journal of the Acoustical Society of America* **36**(10): 1888-1891.

Bergson (1938). La Pensée et le Mouvant. Paris, PUF, collection Quadrige.

Bergson (1941). L'évolution créatrice. Paris, PUF., collection Quadrige.

Bergson, H. (1919). L'énergie Spirituelle. Paris, PUF (collection Quadrige 1999).

Berkeley (1991). Principes de la connaissance humaine. Paris, GF-Flammarion.

Berlin, B. & Kay, P. (1969). Basic color terms : their universality and evolution. Berkeley, University of California Press.

Berlioz, H. (1841 / 1994). De l'instrumentation. Paris, Le Castor Astral.

Berlioz, H. (1855 / 1987). Le chef d'orchestre : théorie de son art. Arles, Actes sud.

Berlioz, H. (1993). Traité d'instrumentation et d'orchestration. Paris, Editions Henry Lemoine.

Berthoz, A. (1997). Le sens du mouvement. Paris, Odile Jacob.

Bhushan, N. Rao, R. A. & Lohse, G. L. (1997). "The texture lexicon : understanding the categorisation of visual texture terms and their relationship to texture image." *Cognitive Science* **21**(2): 219-246.

Bianu, Z. (1996). Krisnamurti ou l'insoumission de l'esprit. Paris, Seuil.

Bismarck, G. v. (1974a). "Timbre of steady sounds : a factorial investigation of its verbal attributes." *Acustica* **30**: 146-158.

Bismarck, G. v. (1974b). "Sharpness as an attribute of the timbre of steady sounds." *Acustica* **30**: 159-172.

- Bower, G. H. & Holyoak, K. (1973). "Encoding and recognition memory for naturalistic sounds." *Journal of Experimental Psychology* **101**(2): 360-366.
- Boysson-Bardies (1999). La représentation des mots chez les bébés. Les langues des sourds / Les prépositions dans la rection des verbes / Le développement du langage avant deux ans. F. C. J.-E. Tyvaert. Reims, Presses Universitaires de Reims: 195-207.
- Bregman, A. S. & Ahad, P. (1995). Demonstrations of Auditory Scene Analysis. Montréal, Québec, Canada, McGill University.
- Britan, H. H. (1911). The philosophy of Music. New York.
- Cadio, P. (1998). L'articulation Sémantique / Pragmatique, Ecole de Sémantique des Langue, Asnelle.
- Cadio, P. & Nemo, F. (1997). "Pour une sémiogénèse du nom." *Langue Française : Aux sources de la polysémie nominale* **113**: 24-34.
- Cadio, P. & Nemo, F. (1997). "Propriétés extrinsèques en sémantique lexicale." *French Language Studies* **7**: 1-19.
- Cadoz, C. (1991). Timbre et causalité. Le timbre métaphore pour la composition. C. Bourgeois. Paris, Ircam.
- Callegari, P. (1994). Aspects conceptuels, expérimentaux et calculatoires de la qualité olfactive. Paris, DEA de Sciences cognitives, EHESS.
- Campbell, I. G. (1942). "Basal emotional patterns expressible in music." *The American journal of Psychology* **55**(1): 1-17.
- Carroll, L. (1871 /1998). Alice au pays des merveilles. Paris, Hachette Jeunesse.
- Casati, R. & Dokic, J. (1994). La philosophie du son. Nîmes, Jacqueline Chambon.
- Castellengo, M. (1994). La perception auditive des sons musicaux. Psychologie de la musique. A. Zenatti. Paris, PUF: 55-86.
- Cavallier, F. (1997). Le langage et la pensée. Paris, Ellipses.
- Caverni, J. P. (1988). La verbalisation comme source d'observables pour l'étude du fonctionnement cognitif. *Psychologie Cognitive : models et méthodes*. C. B. J.P. Caverni, P. Mendelsohn & G. Tiberghien. Grenoble, Presses Universitaires de Grenoble: 253-273.
- Céleste, B. Delalande, F. & Dumaurier, E. (1982). Observations de jeux sonores. L'enfant du sonore au musical. F. Delalande. Paris, INA / GRM / Buchet.Chastel: 180.
- Chion, M. (1982). La musique électro-acoustique. paris, PUF (Que Sais-je ?).
- Chion, M. & Reibel, G. (1976). Les musiques électro-acoustiques. Aix-en-Provence, INA/CRM/EdiSud.
- Chomsky, N. (1990). Le langage et la pensée. Paris, Petite Bibliothèque Payot.
- Clark, E. (1973). What's in a word ? On the child's acquisition of semantics in his first language. Cognitive Development and the Acquisition of Language. T. E. Moore. New-York, Academic.
- Clark, H. H. & Chase, W. G. (1972). "On the process of comparing sentences against pictures." *Cognitive psychology* **3**: 472-517.

- Clark, H. H. & Chase, W. G. (1974). "Perceptual coding strategies in the formation and verification of descriptions." *Memory & Cognition* **2**(1A): 101-111.
- Collins, A. M. & Quilian, M. R. (1969). "Retrieval time from semantic memory." *Journal of Verbal Learning and Verbal Behavior* **8**: 240-247.
- Condillac (1749). Essai sur l'origine des connaissances humaines.
- Cordier, F. (1986). "Les catégories naturelles : niveau de base et typicalité. Les approches développementales." *Revue Française de pédagogie* **77**.
- Cordier, F. (1987). Collections, catégories et acquisition du lexique. Dijon, CRDP.
- Cordier, F. & Tyvaert, J. E. (1999). "Les stratégies parentales de référenciation aux actions et événements." *Interactions & Cognition*.
- Crutcher, R. J. (1994). "Telling what we know : The use of verbal report methodologies in psychological research." *Psychological Science* **5**(5): 241-244.
- Cutting, J. E. (1982). "Plucks and bows are categorically perceived, sometimes." *Perception and Psychophysics* **31**(5): 462-476.
- Cutting, J. E. & Rosner, B. S. (1974). "Categories and boundaries in speech and music." *Perception and Psychophysics* **16**(3): 564-570.
- Cutting, J. E. & Rosner, B. S. (1976). "Discrimination functions predicted from categories in speech and music." *Perception and Psychophysics* **20**: 87-88.
- Cutting, J. E. Rosner, B. S. & Foard, C. F. (1976). "Perceptual categories for musiclike sounds : implications for theories of speech perception." *Quarterly Journal of Experimental Psychology* **28**: 361-378.
- Cyrułnik, B. (1989). Sous le signe du lien. Paris, Hachette.
- Cyrułnik, B. (1995). La naissance du sens. Paris, Hachette Littératures, collection Pluriel.
- Damasio, A. R. (1989). *Cognition* **33**: 25-62.
- Darwin, C. (1874). L'expression des émotions chez l'homme et les animaux. Paris.
- David, S. (1997). Représentations sensorielles et marques de la personne : contrastes entre olfaction et audition. Catégorisation et Cognition : de la perception au discours. D. Dubois. Paris, Kimé: 211-242.
- David, S. Dubois, D. Rouby, C. & Schaal, B. (1997). "L'expression des odeurs en français : analyse lexicale et représentation cognitive." *Intellectica* **24**(1): 51-83.
- De la Motte-Haber, H. (1994). Principales théories scientifiques en psychologie de la musique: les paradigmes. Psychologie de la musique. A. Zenatti. Paris, PUF: 27-54.
- Delalande, F. (1982). Vers une "psycho-musicologie". L'enfant du sonore au musical. F. Delalande. Paris, INA / GRM / Buchet.Chastel.
- Descartes, R. (1637). Discours de la méthode. Paris, GF-Flammarion (1992).
- Dessales, J.-L. (2000). Aux origines du langage, une histoire naturelle de la parole. Paris, Hermes.
- Devos, R. (1989). A plus d'un titre. Paris, Olivier Orban.
- Dokic, J. Droulez, J. & Bullot., N.-J. (2000). Comment et pourquoi percevons-nous des objets ?, MENRT 1 rue Descartes, 75005 Paris.
- Dolto, F. (1995). Tout est langage. Paris, Gallimard.

- Donnadieu, S. (1997). Représentation mentale du timbre des sons complexes et effet de contexte. Psychologie. Paris, Thèse de Psychologie de l'Université René Descartes, Paris V: 335.
- Dravnieks, A. (1982). "Odor quality : Semantically generated multidimensional profiles are stable." *Science* **218**: 799-801.
- Dravnieks, A. & Bock, F. C. (1978). "Comparison of odors directly and through profiling." *Chemical senses and flavour* **3**(2): 191-225.
- Dretske, F. (1969). Seeing and knowing. London, Routledge & Kegan.
- Droit, R.-P. (1998). La compagnie des Philosophes. Paris, Odile Jacob.
- Droit, R.-P. & Sperber, D. (1999). Des idées qui viennent. Paris, Odile Jacob.
- Dubois, D. (1986). "Normes de production d'exemplaires appartenant à 22 catégories sémantiques à partir d'une consigne "classique" et d'une consigne d'imagerie." .
- Dubois, D. (1994). "Identity and autonomy of psychology in cognitive sciences : some remarks from language processing and knowledge representation." *World Future* **42**: 71-78.
- Dubois, D. (1996). Matériels et consignes : un type de questionnement social dans la recherche expérimentale en psycholinguistique. Le questionnement social. J. Zapella. Paris, CLS: 89-98.
- Dubois, D., Ed. (1998). Catégorisation et Cognition : de la perception au discours. Paris, Kimé.
- Dubois, D. Resche-Rigon, P. & Ténin, A. (1997). Des couleurs et des formes : catégories perceptives ou constructions cognitives. Catégorisation et Cognition : de la perception au discours. D. Dubois. Paris, Kimé: 17-40.
- Dubois, D. & Rouby, C. (1997). "Une approche de l'olfaction : du linguistique au neuronal." *Intellectica* **24**.
- Dubois, D. Roudy, C. & Sicard, G. (1991). "Catégories sémantiques et sensorialités : de l'espace visuel à l'espace olfactif." *Enfance*. **1**: 141-151.
- Dumaurier, E. (1982). Le domaine sonore du très jeune enfant. L'enfant du sonore au musical. F. Delalande. Paris, INA / GRM / Buchet.Chastel.
- Dupuy-Engelhardt, H. (1990). La saisie de l'audible, Etude lexématique de l'allemand. Tübingen, Narr.
- Dupuy-Engelhardt, H. (1994). Le champ lexical de l'audible : étude confrontative Allemand - Français. Etudes lexicales. H. Dupuy-Engelhardt. Reims, CIRLEP - Université de Reims Champagne-Ardenne.
- Eco, U. (1994). La quête d'une langue parfaite dans l'histoire de la culture européenne. Paris, Leçon inaugurale du 2 octobre 1992 au Collège de France, Arlequin Vidéo.
- Eco, U. (1999). Kant et l'orithorynque. Paris, Grasset & Fasquelle.
- Edelman, G. M. (1992). Biologie de la conscience. Paris, Poches Odile Jacob.
- Ehrenfels, C. V. (1890). "Über Gestaltqualitäten." *Vierteljahrsschr. f. Wissensch. Philosophie* **3**: 249-299.
- Ellis (1973). . The psychology of learning and motivation. Bower. New York. **7**.
- Engen, T. (1987). "Remembering Odors and their Names." *American Scientist* **75**: 497-503.

- Engen, T. & Engen, E. (1997). "Relationship between development of odor perception and language." *Enfance. L'odorat chez l'enfant : perspectives croisées* **1**: 125-140.
- Enomoto, Y. & Yoshida, T. (1968). Descriptive adjectives on the listening of reproduced sounds. 6Th International Congress on Acoustics, Tokyo.
- Ericsson, K. A. & Simon, H. A. (1984). Protocol Analysis. Cambridge, MIT Press.
- Eustache, F. & Lambert, J. (1998). Introduction à la neuropsychologie, Ecole de sémantique des langues, Asnelle, ARC-CNRS.
- Faure, A. (1994). Analyse de protocoles verbaux obtenus en comparant des timbres musicaux, DEA de Sciences Cognitives, Université Pierre et Marie Curie (Paris VI).
- Faure, A. (1996). Analyse de protocoles verbaux obtenus en comparant des timbres musicaux. Deuxième Colloque Jeunes Chercheurs en Sciences Cognitives, Presqu'île de Giens (Vars).
- Faure, A. & McAdams, S. (1997). Comparaison de profils sémantiques et de l'espace perceptif des timbres musicaux. 4ème Congrès Français d'Acoustique, Marseille, France, Teknea.
- Faure, A. & McAdams, S. (1998). Du percept au concept : comment parle-t-on du timbre des instruments de musique ? 6ème colloque de l'école doctorale Neurobiologie & Comportement, Paris.
- Faure, A. & McAdams, S. (soumis). "Verbal descriptions of musical sounds and their recognition from verbal portraits." .
- Faure, A. McAdams, S. & Nosulenko, V. (1996). Verbal correlates of perceptual dimensions of timbre. 4Th International Conference on Music Perception and Cognition, Montréal.
- Faure, A. McAdams, S. Nosulenko, V. & Samoylenko, E. (en préparation). "Extraction of verbal correlates of perceptual dimensions of timbre from free verbalisations and semantic scales."
- Faure, A. McAdams, S. Poitrenaud, S. & Tijus, C. (en préparation). "Analyse du contenu de verbalisations sur des timbres : Comment parle-t-on des sons ?" .
- Filliozat, I. (1997). L'intelligence du coeur : rudiments de grammaire émotionnelle. Paris, Jean-Claude Lattès.
- Fodor, J. (1975). The Language of Thought. New York, Thomas Y. Crowell.
- Fodor, J. A. (1983). La modularité de l'esprit. Paris, Les éditions de minuit.
- Frize, N. (2000). Les références culturelles du bruit et de l'audition. Paris, Séminaire du Laboratoire d'Acoustique Musicale.
- Gabrielsson, A. & Sjögren, H. (1979). "Perceived sound quality of sound-reproducing systems." *Journal of Acoustical Society of America* **65**(4): 1019-1033.
- Gibran, K. (1923). Le Prophète. Paris, Gallimard.
- Gibson, J. (1950). The perception of the Visual World. Boston, Houghton-Mifflin.
- Grey, J. M. (1977). "Multidimensional perceptual scaling of musical timbres." *The Journal of the Acoustical Society of America* **61**(5): 1270-1277.
- Grey, J. M. & Gordon, J. W. (1978). "Perceptual effects of spectral modifications on musical timbres." *The Journal of the Acoustical Society of America* **63**(5): 1493-1500.
- Grice (1957). "Meaning." *Philosophical Review* **66**: 677-688.
- Grice (1979). "Logique et conversation." *Communication* **30**: 57-72.

- Gundlach, R. H. (1935). "Factors determining the characterization of musical phrases." *The American Journal of Psychology* **47**: 624-643.
- Guski, R. (1990). Concerning résidents' claim for quietness. Third Oldenburg Symposium on Psychological Acoustics, Oldenburg, Allemagne, A. Schick.
- Guyot, F. (1996). , Thèse d'acoustique de l'Université Paris VI.
- Guyot, F. Castellengo, M. & Fabre, B. (1998). Etude de la catégorisation d'un corpus de bruits domestiques. Catégorisation et Cognition : de la perception au discours. D. Dubois. Paris, Kimé: 41-58.
- Hamilton, E. (1998). Historique abrégé de la métaphore. Saint Denis, Mémoire de DEA de l'Université Paris 8-St Denis.
- Handel, S. (1989). Listening. Cambridge, Massachusetts, M.I.T. Press.
- Harnad, S., Ed. (1987). Categorical perception. Cambridge, Cambridge University Press.
- Hary, J. M. & Massaro, D. W. (1982). "Categorical results do not imply categorical perception." *Perception and Psychophysics* **32**(5): 409-418.
- Hausegger, F. V. (1885). Die Musik als Ausdruck. Vienne, Konegen.
- Hegel (1807). Phénoménologie de l'esprit. Paris, Aubier (1991).
- Heinlein, C. P. (1928). "The affective characters of the major and minor modes in music." *Journal of comparative Psychology* **8**(2): 101-142.
- Helmholtz, H. L. F. von. (1877). On the Sensation of Tone as the Physiological Basis for the Theory of Music. New York, Dover.
- Hevner, K. (1935). "The affective character of the major and minor modes in music." *The American Journal of Psychology* **47**: 103-118.
- Hevner, K. (1935). "Expression in music : a discussion of experimental studies and theories." *Psychological Review* **42**: 186-204.
- Hevner, K. (1936). "Experimental studies of the elements of expression in music." *The American Journal of Psychology* **48**: 246-268.
- Hevner, K. (1936). "An experimental study of the affective value of sounds in poetry." *The American Journal of Psychology* **48**: 419-435.
- Hevner, K. (1937). "The affective value of pitch and tempo in music." *The American Journal of Psychology* **49**: 621-630.
- Hofstadter, D. (1988). Ma Thémagie. Paris, InterEditions.
- Izmailov, C. A. & Sokolov, E. N. (1991). "Spherical model of color and brightness discrimination." *Psychological science* **2**(4): 249-259.
- Jean (I, 1-3). Les Quatre Evangiles. Paris, Galimard.
- Jimenez, M. (1997). La psychologie de la perception. France, Domino, Flammarion.
- Jusczyk, P. W. Rosner, B. S. Cutting, J. E. Foard, C. F. & Smith, L. B. (1977). "Categorical perception of nonspeech sounds by 2-month-old infants." *Perception and Psychophysics* **21**(1): 50-54.
- Kaplan, F. (1999). "Dynamiques de l'auto-organisation lexicale : simulations multi-agents et Têtes Parlantes." *InCognito* **15**: 3-23.

- Kaplan, F. (2000). L'émergence d'un lexique dans une population d'agents autonomes. Informatique. Paris, Thèse d'informatique de l'Université Jussieu, Paris VI.
- Karwoski, T. F. & Odbert, H. S. (1938). "Color-music." *Psychological Monographs* **50**(2): 1-59.
- Kendall, R. A. & Carterette, E. C. (1992a). "Verbal attributes of simultaneous wind instrument timbres. I- von Bismarck's adjectives." *Music Perception* **10**: 445-468.
- Kendall, R. A. & Carterette, E. C. (1992b). "Verbal attributes of simultaneous wind instrument timbres. II- Adjectives induced from Piston's *Orchestration*." *Music Perception* **10**: 469-502.
- Kendall, R. A. & Carterette, E. C. (1996). Difference thresholds for timbre related to spectral centroid. 4Th International Conference on Music Perception and Cognition, Montréal.
- Kerrick, J. S. Nagel, D. C. & Bennett, R. L. (1969). "Multiple ratings of sound stimuli." *The Journal of the Acoustical Society of America* **45**(4): 1014-1017.
- Kibbedi-Varga, A. (1996). Sémantique comparée des textes et des arts visuels, Séminaire de Sémantique des textes, Paris I - Sorbonne.
- Krimphoff, J. (1993). Analyse acoustique et perception du timbre, D.E.A. Acoustique Appliquée, Université du Maine, IRCAM, Paris.
- Krimphoff, J. McAdams, S. & Winsberg, S. (1994). "Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantifications psychophysiques." *Journal de Physique* **4**(C5): 625-628.
- Krishnamurti (1967). De la connaissance de soi (Inde, 1948-1950). Paris, Le courrier du livre.
- Krishnamurti (1994). De la nature et de l'environnement. Monaco, Editions du Rocher.
- Krishnamurti, J. & Bohm, D. (1975/1999). Les limites de la pensée. Paris, Stock.
- Kroll, J. F. & Stewart, E. (1994). "Category Inference in Translation and Picture Naming : Evidence for Asymmetric Connections Between Bilingual Memory Representation." *Journal of Memory & Language* **33**: 150.
- Krumhansl, C. L. (1989). Why is musical timbre so hard to understand ? Structure and perception of electroacoustic sound and music. S. Nielzen O. Olsson. Amsterdam, Elsevier: 43-53.
- Kurth, E. (1931). Musikpsychologie. Berlin.
- Lakoff, G. (1987). Women, Fire and Dangerous Things, What Categories Reveal about the Mind. Chicago, Chicago University Press.
- Lakoff, G. & Johnson, M. (1985). Les métaphores dans la vie quotidienne. Paris, Minuit.
- Land, D. G. Harper, R. & Griffiths, N. M. (1970). "An evaluation of the odor qualities of some stimuli proposed as standards for odor research." *The flavor industry* **1**(12): 842-846.
- Lestel, D. (1995). Paroles de singes. L'impossible dialogue homme-primate. Paris, Editions La Découverte.
- Li, X. Logan, R. & Pastore, R. (1991). "Perception of acoustic source characteristics : walking sounds." *Journal of the Acoustical Society of America* **90**(6): 3036-3049.
- Liberman, A. M. Harris, K. S. Hoffman, H. S. & Griffith, B. C. (1957). "The discrimination of speech sounds within and across phoneme boundaries." *Journal of Experimental Psychology* **54**(5): 358-368.



- Lichte, W. H. (1941). "Attributes of complex tones." *Journal of Experimental Psychology* **28**(6): 455-480.
- Light, L. L. & Carter-Sobell, L. (1970). "Effects of changed semantic context on recognition memory." *Journal of verbal learning and verbal behavior* **9**: 1-11.
- Loewenthal, K. (1967). "The development of codes in public and private language." *Psychonomic Science* **8**(10): 449-450.
- Ludwig, P. (1997). *Le langage*. Paris, GF-Flammarion.
- Luria, A. R. (1970). *Une prodigieuse mémoire*. Neuchâtel (Suisse), Delachaux et Niestlé S.A.
- Makyo (1994). *Grimion gant de cuir*. Grenoble, Gléna.
- Mamassian, P. (1996). Les contraintes de l'espace visuel, Séminaire du Laboratoire de Psychologie Expérimentale de Paris V.
- Martin, A. Wiggs, C. L. Ungerlaider, L. G. & Haxby, J. V. (1996). "Neural correlates of category-specific knowledge." *Nature* **379**(15): 649-652.
- Mattingly, I. G. Liberman, A. M. Syrdal, A. K. & Halwes, T. (1971). "Discrimination in speech and nonspeech modes." *Cognitive Psychology* **2**: 131-157.
- McAdams, S. (1993). La reconnaissance de sources et événements sonores. *Penser les sons : La psychologie cognitive de l'audition humaine*. S. McAdams & E. Bigand. Paris, PUF.
- McAdams, S. (1994). Audition non verbale : physiologie, perception et cognition. *Traité de Psychologie Expérimentale*. M. Richelle & J. Requin M. Robert. Paris, P.U.F.
- McAdams, S. & Misdariis, N. (1999). *Proceedings of Human Centred Processes '99*, ENST Bretagne, Brest, P. Lenca.
- McAdams, S. Winsberg, S. Donnadiou, S. De Soete, G. & Krimphoff, J. (1995). "Perceptual scaling of synthesized musical timbres : common dimensions, specificities, and latent subject classes." *Psychological Research* **58**: 177-192.
- Meini, C. (1996). *Le passage du percept au concept : la théorie de F. Dretske*. Deuxième Colloque Jeunes Chercheurs en Sciences Cognitives, Presqu'île de Giens (Vars).
- Melher, J. & Bertoncini, J. (1979). Infant's Perception of Speech and Other Acoustic Stimuli. *Psycholinguistic Series II*. J. Morton J. Marschall. London, Elek Books.
- Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Paris, Gallimard.
- Miller, G. A. (1962). *Psychology : The Science of Mental Life*. New-York, Harper and Row.
- Miller, J. D. Wier, C. C. Pastore, R. E. Kelly, W. J. & Dooling, R. J. (1976). "Discrimination and labeling of noise-buzz sequences with varying noise-lead times : an example of categorical perception." *Journal of the Acoustical Society of America* **60**(2): 410-417.
- Misdariis, N. Smith, B. Pressnitzer, D. Susini, P. & McAdams, S. (1998). *16th International Congress on Acoustics and 135th Meeting Acoustical Society of America*.
- Mosier, C. I. (1941). "A psychometric study of meaning." *The journal of social psychology* **13**: 123-140.
- Nahapétian, N. & Senk, P. (2000). "Le blues des Japonais à Paris." *Psychologies magazine* **188**: 64.

- Namba, S. Kuwano, S. Hatoh, T. & Kato, M. (1991). "Assessment of musical performance by using the method of continuous judgment by selected description." *Music perception* **8**(3): 251-276.
- Nelson, K. (1985). "Le développement de la représentation sémantique chez l'enfant." *Psychologie Française* **20**: 261-268.
- Newell, A. & Simon, H. A. (1972). *Human problem solving*, Englewood Cliffs, NJ : Prentice-Hall.
- Nielzén, S. & Olsson, O. (1993). "The influence of duration on verbal-attribute ratings of complex protomusical sounds." *Music Perception* **11**(1): 73-86.
- Ninio, J. (1989). *L'empreinte des sens : Perception, mémoire, langage*. Paris, Point.
- Nisbett, R. E. & Wilson, T. D. (1977). "Telling more than we can know : verbal reports on mental processes." *Psychological review* **84**(3): 231-262.
- Nosulenko, V. Parizet, E. & Samoylenko, E. (1998). "Application de la méthode d'analyse des verbalisations pour caractériser les bruits de véhicules." *Informations sur les sciences sociales* **37**(4): 593-611.
- Nosulenko, V. N. (1989). "The psychophysics of a complex signal : problems and perspectives." *Soviet Psychology* **27**(1): 62-78.
- Nosulenko, V. N. (1990). "A system approach to the study of auditory perception." : 555-565.
- Nosulenko, V. N. & Samoylenko, E. S. (1997). "Approche systémique de l'analyse des verbalisations dans le cadre de l'étude des processus perceptifs et cognitifs." *Information sur les Sciences Sociales* **36**(3): 223-261.
- Nyckees, V. (1997). "Catégories sémantiques et historicité des significations." *Histoire Epistémologie Langage* **19**(I): 97-119.
- Nyckees, V. (1997). "Pour une archéologie du sens figuré." *Langue Française* **113**: 49-65.
- O'Regan, J. K. (1992). "Solving the "Real" Mysteries of Visual Perception : The World as an Outside Memory." *Canadian Journal of Psychology* **46**(3): 461-488.
- Odbert, H. S. Karwoski, T. F. & Eckerson, A. B. (1942). "Studies in synesthetic thinking : I- Musical and verbal associations of color and mood." *The journal of General Psychology* **26**: 153-173.
- Osgood, C. E. (1952). "The nature and measurement of meaning." *Psychological Bulletin* **49**(3): 197-237.
- Osgood, C. E., Ed. (1979). *Focus on meaning*. The Hague, Mouton publishers.
- Osgood, C. E. & Suci, G. J. (1955). "Factor analysis of meaning." *Journal of experimental psychology* **50**(5): 325-338.
- Osgood, C. E. Suci, G. J. & Tannenbaum (1957). *The measurement of meaning*. Urbana, University of Illinois Press.
- Paris-Grignon, I. (1996). *Goût et Terroir*, Cité internationale universitaire de Paris.
- Parizet, E. & Nosulenko, V. (1999). "Multi-Dimensional Listening Test : Selection of Sound Descriptors and Design of the Experiment." *Noise Control Engineering Journal* **47**(6): 1-6.
- Payeur, C.-R. (1995). *Les chackras symbolisme et méditation*. Sherbrooke (Québec) Canada, Editions de l'aigle.

- Pecjak, V. (1970). "Verbal synesthesiae of colors, emotions and days of the week." *Journal of verbal learning and verbal behavior* **9**: 623-626.
- Petit, J.-L. (1998). Introduction à la phénoménologie, Séminaire de l'EHESS, Jean Petitot.
- Petitot-Cocorda, J. (1985). Les catastrophes de la parole. Paris.
- Piaget, J. (1945). La formation du symbole chez l'enfant. Neuchâtel, Delachaux et Niestlé.
- Pierce, J. (1984). Le son musical (musique, acoustique et informatique). Paris, Pour La Science, diffusion Belin.
- Pinker, S. (2000). Comment fonctionne l'esprit. Paris, Editions Odile Jacob.
- Piston, W. (1955). Orchestration. New York, W. W. Norton & Co.
- Plomp, R. (1970). Timbre as a multidimensionnal attribute of complex tones. Frequency Analysis and Periodicity Detection in Hearing. R. Plomp G. F. Smoorenburg. Sijthoff, Leiden.
- Poitrenaud, S. (1995). "The procope semantic network : an alternative to action grammars." *International Journal of Human-Computer Studies* **42**: 31-69.
- Poitrenaud, S. (1998). La représentation des PROCédures chez l'OPÉrateur : Description et mise en oeuvre des savoirs-faire. Psychologie des processus cognitifs. Saint-Denis, Thèse de Psychologie de l'Université de Paris VIII-Saint Denis: 265.
- Popescu-Belis, A. (1999). Modélisation multi-agent des échanges langagiers : application au problème de la référence et à son évaluation. Orsay, Thèse d'informatique de l'Université Paris XI.
- Prévert, J. (1949). L'orgue de barbarie. Paroles. Paris, Gallimard.
- Pudelko, B. (1999). La compréhension de la métaphore : catégorisation ou comparaison ? Une étude préliminaire. Saint Denis, Mémoire de DEA de l'Université de Paris 8-St Denis.
- Raiman, J. Bourdais, O. & Bourdais, J.-P. (1994). Ainsi vient la vie (de la conception à la naissance). Paris, Montparnasse Multimédia / Arborescence.
- Rastier, F. (1990). "La triade sémiotique, le trivium et la sémantique linguistique." *Actes Sémiotiques*.
- Récanati (1999). Le rôle du contexte dans l'interprétation des énoncés, Séminaire du Laboratoire de Psychologie Expérimentale, Paris V.
- Récanati, F. (1979). La transparence et l'énonciation. Pour introduire à la pragmatique. Paris, Seuil.
- Récanati, F. (1992). Contenu sémantique et contenu cognitif des énoncés. Introduction aux sciences cognitives. D. Andler. Paris, Gallimard.
- Remez, R. E. Cutting, J. E. & Studdert-Kennedy, M. (1980). "Cross-series adaptation using song and string." *Perception and Psychophysics* **27**(6): 524-530.
- Repp, B. H. (1987). "The sound of two hands clapping." *Journal of the acoustical Society of America* **81**: 1100-1109.
- Richard, J. F. (1992). Intériorité et comportement. Cours de psychologie. I- Origines et Bases. R. G. J. F. Richard. Paris, Dunod.
- Rigg, M. (1937). "An experiment to determine how accurately college students can interpret the intended meanings of musical composition." *The Journal of Experimental Psychology* **21**: 223-229.

- Risset, J.-C. (1994). Quelques aspects du timbre dans la musique contemporaine. Psychologie de la musique. A. Zenatti. Paris, PUF: 87-114.
- Rogoff, B. (1984). Thinking and Learning in Social Context. Everyday Cognition. Its development in Social Context. B. R. J. Lave. Cambridge, Harvard University Press.
- Rosch, E. Merris, C. B. Gray, W. D. Johnson, D. M. & Boyes-Braem, P. (1976). "Basic objects in natural categories." *Cognitive Psychology* **8**: 382-439.
- Rosen, S. & Howell, P. (1987). Auditory, articulatory, and learning explanations of categorical perception in speech. Categorical perception - The groundwork of cognition. S. Harnad. Cambridge, Cambridge University Press: 113-159.
- Rosen, S. M. & Howell, P. (1981). "Plucks and bows are not categorically perceived." *Perception and Psychophysics* **30**(2): 156-168.
- Roudy, C. Chevalier, G. Gautier, B. & Dubois, D. (1997). "Connaissance et reconnaissance d'une série olfactive chez l'enfant préscolaire." *Enfance*(1): 152-171.
- Sacks, O. (1988). L'homme qui prenait sa femme pour un chapeau. Paris, Editions du Seuil.
- Saldanha, E. L. & Corso, J. F. (1964). "Timbre cues and the identification of musical instruments." *Journal of the Acoustical Society of America* **36**: 2021-2026.
- Saldanha, H. M. & Rosenblum, L. D. (1993). "Visual influences on auditory pluck and bow judgments." *Perception and Psychophysics* **54**(3): 406-416.
- Samoylenko, E. McAdams, S. & Nosulenko, V. (1996). "Systematic analysis of verbalisations produced in comparing musical timbres." *International Journal of Psychology* **31**(6): 255-278.
- Samoylenko, E. S. (1986). Comparison operations in the cognitive-communicative tasks (Russian). Moscow, Institute of Psychology USSR Academy of Sciences.
- Samoylenko, E. S. (1987). "Comparison in the realization of cognitive-communicative tasks (Russian)." *Voprosi Psychologii* **3**: 128-132.
- Samoylenko, E. S. (1991). A scheme of analysis of verbal data. In : Projet de recherche franco-russe "Homme - environnement acoustique". Paris, Maison des Sciences de l'Homme.
- Sander, E. (2000). L'analogie, du naïf au créatif (Analogie et catégorisation). Paris, L'Harmattan.
- Saunier, E. Franckel, J. J. & Paillard, D. (1994). Mode de régulation de la variation sémantique d'une unité lexicale. Le cas du verbe "passer". La locution : entre lexicque, syntaxe et pragmatique. Klincksiek. Paris, INALF.
- Saussure (1955). Cours de linguistique générale. Paris, Payot.
- Schaal, B., Ed. (1997). Enfance : L'odorat chez l'enfant, perspectives croisées. Paris, PUF.
- Schaeffer, P. (1966). Traité des objets musicaux. Paris, Editions du seuil.
- Scherchen, H. (1953). La direction d'orchestre. Paris, Actes Sud.
- Schulte-Fortkamp, B. (1993). Judging noise in the laboratory. An analysis based on a new method. Contribution to Psychological Acoustics, Results of the sixth Oldenburg Symposium on Psychological Acoustics. Oldenburg, A. Schick: 487-504.

- Schulte-Fortkamp, B. & Weber, R. (1990). Subjective variables in the absolute judgement of loudness under laboratory condition. Results of the third Oldenburg Symposium on Psychological Acoustics. A. Schick. Oldenburg: 237-252.
- Schwartz, G. (1978). "Estimating the dimensions of a model." *Annals of Statistics* **6**: 461-464.
- Segui, J. & Ferrand, L. (2000). Leçons de parole. Paris, Odile Jacob.
- Sempé (1986). Quelques Romantiques. Paris, Editions Denoël.
- Siegel, J. A. & W., S. (1977). "Categorical perception of tonal intervals : musicians can't tell sharp from flat." *Perception and Psychophysics* **21**(5): 399-407.
- Smith, B. (1995). PsiExp: an Environment for Psychoacoustic Experimentation Using the IRCAM Musical Workstation. Society for Music Perception and Cognition Conference '95, University of California, Berkeley.
- Soja, N. N. Carey, S. & Spelke, E. S. (1991). "Ontological categories guide young children's induction of word meaning : object terms and substance terms." *Cognition* **38**: 179-211.
- Solomon, L. N. (1957). "Semantic approach to the perception of complex sounds." *The journal of the Acoustical Society of America* **30**(5): 421-425.
- Sopchak, A. L. (1955). "Individual differences in responses to different types of music, in relation to sex, mood and other variables." *Psychological monographs* **69**(11): 1-69.
- Sperber & Wilson (1987). "Précis of Relevance : Communication and Cognition." *Behavioral and Brain Sciences* **10**: 697-699.
- Sperber, D. & Wilson, D. (1986). La pertinence. Paris, Les éditions de Minuit.
- Steels (1998). Le langage naturel comme système adaptatif complexe, Ecole Normale Supérieure, rue d'Ulm, Séminaire du Département de Mathématiques et Informatique.
- Steels, L. & Kaplan, F. (1999). Bootstrapping Grounded Word Semantics. Linguistic evolution through language acquisition : formal and computational models. T. Briscoe. Cambridge, UK, Cambridge University Press.
- Stumpf, C. (1883 u. 1890). Tonpsychologie. Leipzig.
- Surrans, A. (1993). Le regard du musicien. Paris, Edition Plume.
- Szejer, M. & Stewart, R. (1994). Ces neuf mois-là. Paris, Robert Laffont.
- Tardieu, J. (1981). Jean Tardieu, un poète. Evreux, Gallimard.
- Taylor, L. J. Gandy, L. J. & Dark, G. (1974). "Linguistic description and auditory perception." *Perceptual and motor skills* **38**: 703-707.
- Tchaïkovski (1878). lettre à Tanéev, 27 mars 1878.
- Teplov, B. M. (1966). Psychologie des aptitudes musicales. Paris, PUF.
- Tijus, C. (1996). Assignment de signification et construction de la représentation. St. Denis, Mémoire d'Habilitation à diriger des recherches. Université de Paris 8- St Denis.
- Tijus, C.-A. (1994). Modèles connexionnistes de la Perception. Cours de psychologie, Tome 4 : Mesures et Analyses. R. G. J. F. Richard. Paris, Dunod.
- Tijus, C.-A. & Moulin, F. (1997). "L'assignation de signification étudiée à partir de textes d'histoires drôles." *L'Année psychologique* **97**: 33-75.

- Tucker, W. T. (1955). Experiments in aesthetic communications. Unpublished doctor's dissertation, University of Illinois.
- Tversky, A. (1977). "Features of Similarity." *Psychological review* **84**: 327-352.
- Tyvaert, J.-E. (2000). "L'être en langue : une analyse de l'opposition verbo-nominale en terme de capacité référentielle assortie de considérations épistémologiques portant sur les conditions d'une cognition véritablement naturelle." *Sémiotiques* **16**.
- Tyvaert, J.-E. (2000). Trois dilemmes philosophiques et leur résolution linguistique, Séminaire du Groupe STP (Du Sujet, Théorie et Praxis), EHESS.
- Urdapilleta, I. (1998). Catégorisation et implication de propriétés. Psychologie des processus cognitifs. Saint-Denis, Paris VIII-Saint-Denis: 241.
- Vaillant, P. (1993). Analyse de la stratégie de description verbale dans des tâches de description de paysages en présence de contraintes temporelles. Paris, DEA de Sciences Cognitives Université Paris-Sud Orsay.
- Van Derveer, N. J. (1979). Acoustic information for event perception. New York, Cornell University, Ithaca. Unpublished paper presented at the celebration in honor of Eleanor J. Gibson.
- Varela, F. Thompson, E. & Rosch, E. (1993). L'inscription corporelle de l'esprit. Paris.
- Varela, F. J. (1989). Autonomie et connaissance, essai sur le vivant. Paris, Seuil.
- Verlaine, P. (1938). "Art Poétique", Œuvres poétiques complètes, Y. G. Le Dantec éd., coll. La Pléiade, Paris, rééd. rév. J. Borel, Paris, 1965, 1977.
- Vian, B. (1991). Boris Vian et ses interprètes. Paris, PolyGram.
- Victorri (1999). Le processus de compréhension d'un énoncé, Séminaire du laboratoire Cognition Finalisée, Paris 8.
- Victorri, B. & Fuchs, C. (1996). La polysémie. Construction dynamique du sens. Paris, Hermes.
- Vygotski, L. (1934 / 1997). Pensée et Langage. Paris, La Dispute.
- Warren, W. H. & Verbrugge, R. R. (1984). "Auditory perception of breaking and bouncing events : a case study in ecological acoustics." *Journal of Experimental Psychology : Human Perception and Performance* **10**(5): 704-712.
- Weber, R. (1990). The continuous loudness judgement of temporally variable sounds with an "analog" category procedure. Results of the third Oldenburg Symposium on Psychological Acoustics. A. Schick. Oldenburg: 267-294.
- Wedin, L. (1972). "A multidimensional study of perceptual-emotional qualities in music." *Scand. J. Psychol.* **13**: 241-257.
- Wedin, L. & Goude, G. (1972). "Dimension analysis of the perception of instrumental timbre." *Scand. J. Psychol.* **13**: 228-240.
- Wertheimer, M. (1923). "Untersuchung zur Lehre von der Gestalt II." *Psychologische Forschung* **4**: 301-350.
- Wessel, D. L. Bristow, D. & Settel, Z. (1987). Control of phrasing and articulation in synthesis. International Computer Music Conference, San Francisco.
- Willems, E. (1985). L'oreille musicale. Fribourg / Suisse, Pro Musica.

- Wilson, T. D. (1994). "The proper protocol : validity and completeness of verbal reports." *Psychological Science* **5**(5): 249-252.
- Winsberg, S. & Carroll, J. D. (1989a). "A quasi non-metric method for multidimensionnal scaling via an extended Euclidean model." *Psychometrika* **54**: 217-229.
- Wright, P. (1971). "Linguistic description of auditory signals." *Journal of applied Psychology* **55**(3): 244-250.

## TABLE DES MATIERES

<b>Remerciements</b>	<b>4</b>
<b>Avant-Propos</b>	<b>6</b>
<b>Introduction</b>	<b>9</b>
<b>1- Parler des sons : un programme controversé</b>	<b>9</b>
La musique est de l'ordre de l'ineffable	9
Un nouveau travail sur le son nécessite de nouveaux outils	10
Liens entre langage, pensée et monde	11
<b>2- Définitions du timbre : un résumé de la problématique</b>	<b>11</b>
Démarche analytique vs globale	11
Un phénomène complexe...	13
<b>3- Les moyens : expérimentation et portée des conclusions</b>	<b>13</b>
Trois expériences	14
Douze sons pour représenter le timbre	15
Expérimentation écologique ?	16
<b>Partie I : Questions et inscription théorique</b>	<b>19</b>
<b>Chapitre 1. Être au monde</b>	<b>22</b>
I. Le monde existe-t-il en dehors de nous ?	22
I.1. Le monde existe	22
Evidence	23
Dualisme	23
Monde réel, mondes possibles	24
I.2. Tout n'est qu'illusion	24
La matière est "forme-pensée"	24
Vacuité ou néant	25
Monisme ?	25
I.3. Une illusion vitale	26
Nécessité	26
Une illusion partagée	26
L'action reste possible	27
Lignes de résistances	27
Voie moyenne	28
II. Catégorisation et représentation interne du monde ?	28
II.1. De la nécessité de catégoriser le monde ?	29
Continuum et segmentation	29
Voir le même à travers le différent	30
Economie cognitive	30
Catégories et activité humaine	31
II.2. De la nécessité des représentations mentales ?	31
Théorie de l'information	31
Le tout est plus que la somme des parties	33
Perception directe	33
Recherche d'indices	34
Dynamique, flux	34
II.3. Le langage est créateur	35
Perception pure ?	35
Nominalisme	36
Le langage crée les concepts	37
Le langage comme outil de discrétisation du monde	37
III. Résumé	38
<b>Chapitre 2. Parler</b>	<b>40</b>
I. Lien entre le langage et le monde	41
I.1. Sémantique : recherche du sens	42



Signe	42
Naturel / conventionnel ?	42
Référence	43
Compositionnalité	44
I.2. Polysémie	45
Emiettement des sens	46
Noyau de sens	46
Schéma de sens et construction dynamique du sens	47
Sens nouveaux et évolution du sens	48
I.3. Résumé	49
II. Langage comme moyen de communication	50
II.1. Communiquer	50
Pourquoi ?	50
Comment ?	51
Une nouvelle illusion ?	51
II.2. Parler pour l'autre	51
Parler pour soi / pour l'autre	52
Modélisation	52
II.3. Se faire comprendre	53
Maximes de Grice	53
L'interprétation pragmatique	54
Théorie de l'esprit	55
Tout est possible en contexte ?	55
II.4. Résumé	55
III. Lien entre le langage et la pensée	56
III.1. Le langage signe ou accomplissement de la pensée	56
Est-ce que le langage peut décrire la pensée ?	56
Pas de pensée sans mot	57
Le langage accomplit la pensée	57
Pensée pure ?	57
III.2. Et les animaux ?	58
Faut-il des mots pour penser ?	58
Dangers de l'anthropomorphisme	59
Si les lions pouvaient parler	59
Evidence d'une pensée sans langage	60
III.3. Résumé	60
IV. La vérité sort de la bouche des enfants	61
IV.1. Un être de langage	61
Prêter du sens aux productions	61
Regard de l'autre	62
IV.2. Pointer du doigt	63
Première fois	63
Co-occurrence et construction de la référence	63
La syntaxe prend le relais	64
IV.3. Explosion	64
Prosodie	65
Premier mot : "papa" ou "maman" ?	65
Son - onomatopée - mot	66
De plus en plus de mots	67
"Mots d'enfants" et polysémie	67
Inventions et non-mots	68
IV.4. Résumé	68
V. Conclusion	70

<b>Chapitre 3. Des sons</b>	<b>72</b>
I. Son, timbre, représentation perceptive	72
I.1. Acoustique	72
Son	73
Bruit - son complexe harmonique / inharmonique	74
Timbre	74
I.2. Perception d'un objet versus de qualités	75
I.3. Dans le domaine sonore	77
Ecouter, ouïr, entendre, comprendre	77
Identification - Qualification	78
Chez l'enfant	79
Position des compositeurs contemporains	80
I.4. Résumé	81
II. Catégorisation, reconnaissance, identification	81
II.1. Catégorisation	81
CNS, prototypes et air de famille	81
Drôles d'oiseaux	82
II.2. Reconnaissance et identification	83
Bruits	83
Influence du label sur la mémorisation	84
Instruments de musique	84
II.3. Perception catégorielle	85
Le timbre : temps d'attaque	86
Controverse	86
II.4. Résumé	87
III. Les verbalisations en psychologie	87
III.1. Méthodes extérieures / introspection	87
Introspection	88
Contre l'introspection	88
Sémantique différentielle	89
Les verbalisations libres remplacent l'introspection	89
Faut-il trancher ?	90
III.2. Modalités autres qu'auditives	90
En vision	90
Sur les odeurs	91
Sur le goût	92
III.3. Sur la musique	92
Majeur / mineur	92
Jugements continus en écoute d'une oeuvre	93
III.3. Sur les sons	94
Pas de vocabulaire spécifique	95
Linguistique	95
Sémantique différentielle	95
Bruits de voiture et d'aspirateurs	96
III.4. Sur le timbre	96
Sémantique différentielle	97
Verbalisations libres	98
III.5. Résumé	100
IV. Conclusion	100
<b>Hypothèses</b>	<b>101</b>

<b>Partie II : Expériences effectuées</b>	<b>102</b>
<b>Chapitre 4 : Présentation et Méthodologie des trois expériences réalisées</b>	<b>104</b>
I. Expérience 1 : Jugements de dissemblance entre paires de sons et verbalisations libres	104
I.1. Sujets	104
I.2. Stimuli	104
I.3. Procédure	105
I.4. Matériel	106
I.5. Principe et résultats de l'Analyse MDS	106
II. Expérience 2 : Échelles sémantiques	108
II.1. Présentation	109
II.1.1. Sémantique différentielle	109
II.1.2. Échelles unipolaires	109
II.1.3. 23 échelles choisies	109
II.1.4. Jugements de pertinence	110
II.1.5. Groupes de sujets	110
II.2. Sujets	111
II.3. Stimuli	111
II.4. Procédure	111
II.5. Matériel	112
II.6. Analyses préliminaires	112
II.6.1. Jugements bruts	112
Corrélations inter-sujets	112
Erreurs des sujets	114
Inversions d'échelles	114
II.6.2. Jugements corrigés	115
Comparaison des corrélations inter-sujets avant et après correction	115
Comparaison des corrélations inter-sujets entre les séances 1 et 2	116
Corrélations inter-sujets échelle par échelle	116
II.7. Utilisation des échelles sémantiques	117
II.7.1. Un exemple : l'échelle peu aigu - très aigu	118
II.7.2. Recherche systématique	119
II.7.3. Répartitions en fonction des sujets	121
II.7.4. Comportement de la moyenne	122
II.7.5. Différents calculs de moyennes	122
II.8. Analyse du nombre de re-écoute des sons	123
II.8.1. Une mesure des difficultés rencontrées ?	123
II.8.2. Description générale	123
II.8.3. Différences inter-séances	124
II.8.4. Différences inter-sujets	124
II.8.5. Différences inter-échelles	124
II.8.6. Différences inter-sons	126
II.8.7. Résumé	129
III. Expérience 3 : Reconnaissance de portraits verbaux	130
III.1. Construction de portraits des sons	130
III.2. Sujets	131
III.3. Stimuli	132
III.4. Procédure	132
III.5. Matériel	133
III.6. Analyses préliminaires : utilisation de la consigne	133
III.6.1. Nombre de sons associés en fonction du portrait proposé	133
Dépendance du type de portrait	133
Dépendance du sous-groupe de portrait	134
Dépendance du portrait	135
III.6.2. Nombre de sons associés en fonction des auditeurs	138
Dépendance de la compétence musicale des auditeurs	138
Dépendance de la personnalité des auditeurs	138
III.6.3. Résumé	139
<b>Chapitre 5 : Corrélats verbaux des dimensions perceptives</b>	<b>141</b>

I. Expérience 1 : recherche de corrélats dans des verbalisations libres	143
I.1. Choix du matériel analysé	143
I.2. Analyses	145
I.2.1. Corrélations de rang de Kendall	145
I.2.2. Régressions multiples	147
I.2.3. Représentation graphique	149
I.3. Résumé	153
I.4. Comparaison avec le vocabulaire de la littérature utilisant des échelles sémantiques	154
II. Expérience II : Recherche de corrélats parmi les échelles sémantiques	156
II.1. Analyses à partir des résultats moyennés sur des groupes de sujets	156
II.1.1 Différentes moyennes calculées	156
II.1.2. Corrélations avec les dimensions perceptives	159
II.1.3. Différences entre musiciens et non-musiciens	160
II.1.4. Différences inter-mode de calcul des moyennes	161
II.1.5. Résumé	162
II.2. Analyses sujet par sujet	162
II.2.1. Attributs verbaux de chaque dimension perceptive	164
II.2.2. Des attributs verbaux composites des dimensions perceptives ?	165
II.2.3. Seuils corrigés selon Bonféroni pour 23 analyses	167
II.2.4. Résumé	168
II.3. Régressions multiples incrémentales	169
II.3.1. Régressions	169
II.3.2. Différences musiciens - non-musiciens	173
II.4. Comparaison avec l'utilisation des échelles	175
II.4.1. Comparaison avec la notion de pertinence	175
II.4.2. Comparaison avec les corrélations inter-sujets	176
III. Comparaison entre les résultats des expériences 1 et 2	177
III.1. Comparaison des verbalisations spontanées et des positionnements sur les échelles	177
III.2. Profils sémantiques de chaque dimension perceptive	178
III.3. Profils perceptifs de chaque échelle sémantique	179
III.4. Comparaisons des conclusions des deux expériences	180
III.4.1. Corrélats verbaux des dimensions perceptives	180
III.4.2. Corrélats composites de plusieurs dimensions perceptives	183
IV. Analyse de profils sémantiques des sons obtenus à partir des échelles	184
IV.1. Calcul de distances entre les sons à partir des profils sémantiques	185
IV.2. Analyses multidimensionnelles	185
IV.3. Corrélations avec les dimensions perceptives, les paramètres acoustiques et les positions moyennes sur chacune des échelles sémantiques.	186
IV.4. Discussion	186
V. Conclusion	187
<b>Chapitre 6 : Description des sons</b>	<b>188</b>
I. Le logiciel STONE : un arbre hiérarchique de propriété	189
I.1. Formalisme : des objets, des propriétés	189
I.2. Construction de l'arborescence	190
II. Expérience 1 : Description de sons présentés par paires	193
II.1. Description générale	193
II.1.1. Répartition en fonction des participants	194
II.1.2. Répartition en fonction des sons	195
II.1.3. Résumé	196
II.2. Parler de ressemblances et dissemblances	196
II.3. Parler d'eux-mêmes	198
Répartition en fonction des participants	199
II.4. Décrire les sons	200
II.4.1. Déroulement temporel	202
Le début des sons	203
Le milieu des sons	205
La fin des sons	206
Autres catégories temporelles	209

Lien avec des mesures acoustiques	210
II.4.2. La forme des sons	210
II.4.3. Vocabulaire relevant des 5 sens	212
Répartition en fonction des participants	214
Audition	215
Toucher	215
Vision	215
Goût	216
Répartition en fonction des sons	216
Audition	216
Toucher	217
Vision	217
Goût	218
II.4.4. Émotions	218
II.4.5. Autres manières de décrire les sons	220
Qualités liées à la source	220
Divers	221
Énergie, Intensité	221
Noms de sons et imitation	222
II.5. Décrire une source sonore	222
II.5.1. Objets produisant des sons	224
Répartition en fonction des participants	226
Répartition en fonction des sons	227
II.5.2. Actions produisant des sons	229
II.5.3. Mécanisme produisant des sons	232
II.5.4. Description d'une source complexe	234
III. Expérience 3 : Description de sons en fonction de consignes différentes	235
III.1. Comparaison des 4 consignes	236
III.1.1. Longueur des descriptions	236
Mesurée en nombre de caractères	237
Mesurée en nombre d'unités verbales	237
Consigne	237
Sons	238
Ordre de description	238
Locuteurs	238
Conclusion	238
III.1.2. Locutions utilisées comme portrait	239
Noms d'instruments de musique ou de familles d'instruments	239
Noms de mode d'excitation, de matériaux et d'association des deux	240
Adjectifs	241
Conclusion	242
III.2. Comparaison de la consigne "Instruments" avec l'expérience 1	242
III.2.1. Ce qui a produit le son	243
III.2.2. Concerne le sujet	246
Difficultés	246
Marques de la personne	247
III.2.3. Concerne le son	248
III.2.4. Comparaison avec un autre son	248
III.3. Des gestes quand on ne trouve pas comment dire	248
III.3.1. Quels gestes ?	249
III.3.2. Les gestes dépendent-ils des sons ?	249
III.3.3. Les gestes dépendent-ils de la consigne ?	250
III.3.4. Conclusion	251
III.4. Quelques remarques faites par les sujets	251
III.4.1. Sur les sons en général	251
III.4.2. Sur la difficulté de la tâche	252
III.4.3. Sur la description de qualités / identification de la source	252
III.4.4. Sur les liens langage-perception	253
IV. Discussion	253
IV.1. Apports de l'expérience 1	253

IV.1.1. Des verbalisations libres comme données ?	253
IV.1.2. Les locuteurs parlent du son	254
IV.1.3. Les locuteurs parlent de la source	255
IV.1.4. Parlez-moi des sons, je vous dirai qui vous êtes...	256
IV.2. Comparaison avec l'approche de Samoylenko et Nosulenko	257
IV.2.1. Le mode d'élaboration et le principe sont très différents.	257
IV.2.2. Comparaison des deux arborescences.	258
Niveau logique	258
Niveau de référence au stimulus	259
Niveau sémantique	259
IV.3. Apport de l'expérience 3	260
IV.4. Conclusion	261
<b>Chapitre 7. Pertinence des échelles et sens des portraits</b>	<b>263</b>
I. Expérience 2 : pertinence des échelles et commentaires des sujets	264
I.1. Jugements de pertinence	264
I.1.1. Pertinence globale brute	265
Jugements de l'ensemble des sujets	266
Comparaison entre différents groupes de sujets	266
I.1.2. Pertinence son par son	267
Corrélations inter-sujets sur les jugements de pertinence	267
Description générale	267
Différences musiciens / non-musiciens	271
Différences entre les sons	271
Comparaison avec les jugements globaux	272
I.1.3. Lien entre le jugement de pertinence globale et l'utilisation spontanée (annexe A.7.4).	272
I.1.4. Lien avec l'utilisation resserrée ou dilatée des échelles	274
I.2. Commentaires et préférences des sujets	276
I.2.1. Échelles préférées par les sujets	276
I.2.2. Commentaires sur les échelles	278
I.2.3. Quand les sujets expliquent ce qu'ils ont l'impression de faire.	279
I.3. Conclusion : comparaison entre ces différentes mesures	280
II. Expérience 3 : sens des portraits	287
II.1. Jugement de sens des portraits	287
II.1.1. Jugement de sens en fonction du portrait proposé	287
Dépendance du type de portrait	287
Dépendance du sous-groupe de portrait	288
Dépendance du portrait	288
II.1.2. Jugement de sens en fonction de la compétence musicale des auditeurs	289
Types de portraits	289
Sous-groupes de portraits	290
II.2. Comparaisons avec les autres données	291
II.2.1. Lien avec les jugements de pertinence de l'expérience 2	291
II.2.2. Lien avec le nombre de productions spontanées	293
II.3. Résumé	299
III. Conclusion	300

<b>Chapitre 8. Données d'association de sons aux portraits</b>	<b>302</b>
I. Cohérence des associations entre les participants	303
I.1. Cohérence d'association en fonction des types de portraits (%RC et %CC)	303
I.2. Cohérence d'association en fonction de la compétence des sujets (%RC et %CC)	307
I.3. Conclusion	311
II. Sons choisis pour chaque portrait	313
II.1. Dispersion des réponses sur l'ensemble des timbres proposés	313
II.2. Portraits décrivant le son	314
II.2.1. Qualités simples testées dans l'expérience 2	314
Différence entre portraits "positifs" et "négatifs"	316
Différence entre musiciens et non-musiciens	317
II.2.2. Autres qualités simples	317
Différence entre musiciens et non-musiciens	318
II.2.3. Association de deux qualités	318
Différence entre musiciens et non-musiciens	319
II.2.4. Association de 3 qualités	320
Différence entre musiciens et non-musiciens	321
Analyse de la diagonale	321
II.3. Portraits décrivant une source sonore possible	321
II.3.1. Noms d'action	321
II.3.2. Noms de matériaux	323
II.3.3. Noms de sources sonores	323
II.3.4. Familles d'instruments	325
II.2.5. Noms d'instruments	327
II.2.6. Noms d'instruments visés par le concepteur	328
II.4. Comparaison de portraits proches dans des contextes différents	329
II.4.1. Qualités triples / doubles / simples	329
II.4.2. Sources / action + matériau	330
II.4.3. Matériau / Famille	331
II.4.4. Instruments / instruments visés	332
II.4.5. Source sonore / qualité sonore	333
II.5. Meilleurs portraits pour chacun des 12 sons	334
II.6. Résumé	336
III. Comparaison avec les verbalisations spontanées	338
III.1. Portraits décrivant le son	340
III.1.1. Qualités simples testées dans l'expérience 2	340
Comparaison Spontanée E1 - Portraits E3	340
Comparaison Spontanée E3 - Portraits E3	342
III.1.2. Autres qualités simple	343
Comparaison Spontanée E1 - Portraits E3	343
Comparaison Spontanée E3 - Portraits E3	344
III.1.3. Association de deux qualités	345
Comparaison Spontanée E1 - Portraits E3	345
Comparaison Spontanée E3 - Portraits E3	345
III.2. Portraits décrivant une source sonore possible	346
III.2.1. Noms d'action	346
Comparaison Spontanée E1 - Portraits E3	346
Comparaison Spontanée E3 - Portraits E3	347
III.2.2. Noms de matériaux	348
Comparaison Spontanée E1 - Portraits E3	348
Comparaison Spontanée E3 - Portraits E3	349
III.2.3. Noms de sources sonores	350
Comparaison Spontanée E1 - Portraits E3	350
Comparaison Spontanée E3 - Portraits E3	351
III.2.4. Familles d'instruments	352
Comparaison Spontanée E1 - Portraits E3	352
Comparaison Spontanée E3 - Portraits E3	353
III.2.5. Noms d'instruments	353

Comparaison Spontanée E1 - Portraits E3	353
Comparaison Spontanée E3 - Portraits E3	353
III.2.6. Noms d'instruments visés par le concepteur	354
III.3. Résumé	354
IV. Conclusion	356

## **Partie III : Discussion** **357**

### **I. Les sons pour nous** **357**

I.1. Un son perçu comme une source sonore	357
Niveaux de catégorisation	358
Réminiscence d'une fusion Perception / Motricité	359
Le son est avant tout signe ?	360
Un objet de mémoire	361
I.2. Qualités des sons	362
Une forme temporelle	362
Correspondance avec les descriptions acoustiques	362
Rôle de l'attaque ?	363
I.3. Parler du son / parler de la source	363
Influence des compétences musicales	364
Axes de classification	365
Littérature française	366
Conclusion	366

### **II. Le langage pour décrire nos perceptions** **367**

II.1. Un vocabulaire inadapté ?	367
Un vocabulaire non pertinent ?	367
Impossibilité de décrire une perception ?	368
Utilisation commune de ce vocabulaire	368
II.2. Sans vocabulaire spécifique, on se comprend	368
Le sens des mots est ajusté dans le contexte ?	369
Empathie ?	369
Mots, non-mots et néologismes	370
Faire résonner du sens chez l'autre	370

### **III. Notre perception du monde** **371**

III.1. Métaphores	371
Vocabulaire des musiciens	372
Couleurs sonores	372
Lumières, goût et impressions tactiles	373
Hauteur sonore	373
III.2. Perception	375
Métaphores ou synesthésie	375
Des modalités sensorielles indépendantes	375
Sons de parole - Reste de l'audition	376
Perception pure ?	376
III.3. Mémoire	377
Une mosaïque pour mémoire	378
De la mémoire en toutes choses	378
Vibrones, licornes, et autres chimères	379
III.4. Vers un modèle du lien percept-concept (ou du sens ?)	379
Théorie de l'information	379
Mosaïque perceptive	380
Entrées - Sorties ?	380
Unité du sens	381
Emergence d'une forme cérébrale ?	382
Alternative aux boîtes et flèches...	382

### **IV. Conclusion** **383**

### **V. Perspectives** **384**



<b>Postface</b>	<b>385</b>
<b>Bibliographie</b>	<b>386</b>
<b>Table des matières</b>	<b>400</b>
<b>Annexes et CD</b>	<b>410</b>

**RESUME** : Des sons aux mots, comment parle-t-on du timbre musical ?

Il s'agit de comprendre, à travers l'étude du timbre musical (une qualité du son qui permet d'identifier la source sonore et de distinguer deux sons de même hauteur, même intensité et même durée mais produits par des instruments différents) comment s'organisent les liens entre nos perceptions et les concepts qu'on y associe, révélés par des verbalisations.

Il semble difficile de décrire avec des mots nos perceptions sonores et il n'apparaît pas de vocabulaire spécifique du timbre musical. Par contre, il semble très facile de reconnaître un son, dans un ensemble de 12 sons, à partir d'une des descriptions produites par d'autres personnes. Plusieurs types de portraits verbaux ont été testés lors de cette reconnaissance. Les portraits décrivant un objet ou une action ayant pu produire le son semblent beaucoup plus faciles à reconnaître que les descriptions de qualités des sons entendus. Alors qu'il nous est difficile de parler de nos perceptions sonores, l'association qu'un autre auditeur a pu faire entre un son entendu et une description produite est facile à comprendre. La communication du contenu de nos perceptions sonores est donc possible même en l'absence d'un vocabulaire spécifique. Ceci nous oblige à considérer le langage non pas dans sa composante référentielle, mais dans son aspect collectif et intersubjectif.

**SUMMARY** : From sounds to words, how do we talk about musical tone ?

Analyzing how we describe sounds allows a better understanding of what they mean to us and how perceptual representations are enriched by semantic representations. In spite of the lack of a specific sound-related vocabulary, listeners asked to describe the similarities and differences between musical instrument sounds are quite prolific in their verbal productions. They tend to describe either the action or vibrating object at the origin of the sound, its temporal form or a variety of perceptual attributes. In the latter case, many terms from other sensory domains (visual, tactile, smell) are used. Ratings of sounds on semantic scales using descriptors of perceptual attributes extracted from these free verbalizations correspond closely to freely verbalized relations. In spite of the paucity of sound-related vocabulary and the semantic ambiguities of the words used, listeners are able to recognize auditorally a described sound. Verbal communication in poorly lexicalized domains is thus strongly constrained by context.

**DISCIPLINE** : Sciences Cognitives.

**MOTS CLEFS** : perception auditive, timbre, dénomination, verbalisations libres, échelles sémantiques, liens entre percepts et concepts.