HAL
open science

Caractérisation acoustique et perceptive du mouvement évoqué par les sons pour le contrôle de la synthèse

Adrien Merer

To cite this version:

Adrien Merer. Caractérisation acoustique et perceptive du mouvement évoqué par les sons pour le contrôle de la synthèse. Acoustique [physics.class-ph]. Université de Provence - Aix-Marseille I, 2011. Français. NNT : . tel-00602787

HAL Id: tel-00602787

https://theses.hal.science/tel-00602787v1

Submitted on 23 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CNRS - Laboratoire de Mécanique et d'Acoustique (UPR7051)
Université de Provence - Aix-Marseille 1
Ecole doctorale 353 - Sciences pour l'ingénieur : Mécanique, Physique, Micro et Nanoélectronique

Thèse
Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITE DE PROVENCE - AIX-MARSEILLE I
Discipline : ACOUSTIQUE

Caractérisation acoustique et perceptive du mouvement évoqué par les sons pour le contrôle de la synthèse

---

Adrien Merer

Présentée et soutenue publiquement le 6 juin 2011

Jury:

Sølvi Ystad	CNRS - LMA	(Directrice de thèse)
Richard Kronland-Martinet	CNRS - LMA	(Directeur de thèse)
Mitsuko Aramaki	CNRS - INCM	(Examinatrice)
Rolf Inge Godøy	Universitetet i Oslo	(Examinateur)
Marcelo Wanderley	McGill University	(Rapporteur)
Kristoffer Jensen	Aalborg Universitet	(Rapporteur)

Remerciements

Cette thèse a démarrée grâce à l'impulsion de Pascal Gobin du conservatoire de Marseille et de mes trois encadrants Sølvi Ystad, Richard Kronland-Martinet et Mitsuko Aramaki.

Je tiens à remercier toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce travail par leurs conseils, leurs remarques ou en ayant participé aux différents tests perceptifs que j'ai menés durant la thèse.
Je ne prendrais pas ici le risque de ne pas être exhaustif tant les rencontres ont été nombreuses et riches durant la thèse, mais je souhaite que tous ceux dont l'expérience, les réflexions, le travail et la motivation ont conduit à ce document, soient assurés de la gratitude que je leur porte.
Je pense avoir eu de la chance en ayant la possibilité de faire une thèse dans d'aussi bonnes conditions et je souhaite à d'autres d'avoir la même chance. En particulier, il me semble de plus en plus rare de pouvoir profiter d'une aussi grande liberté de thématique et d'organisation du travail sans être trop confronté à des contraintes économiques et politiques. J'ai également pu profiter d'un environnement de recherche où l'aspect humain n'est pas laissé de côté et où il a toujours été agréable de venir travailler.
Je tiens donc à remercier tous ceux qui ont oeuvré à réunir de telles conditions et j'espère que les changements qui s'opèrent actuellement dans le monde de la recherche ne vont pas supprimer les éléments qui y sont nécessaires.
Enfin je souhaite remercier tous ceux avec qui, en parallèle de la thèse, j'ai partagé des moments musicaux, artistiques, festifs ou simplement vivants et qui ont donc contribué à faire que le bilan de ces trois ans soit à ce point positif.

Table des matières

Introduction **5**
Contrôle de processus de synthèse sonore . . . 5
Sémiotique des sons . . . 6
Enjeux et méthodologie . . . 8
Organisation du document . . . 9

1 Notions de mouvement **12**
1.1 Mouvement évoqué par les sons . . . 12
1.1.1 Points de départ . . . 12
1.1.2 Vocabulaire et définitions . . . 13
1.1.3 Motivations . . . 15
1.2 Le mouvement musical . . . 16
1.2.1 Modèles basés sur des mouvements corporels . . . 16
1.2.2 Modèles mécaniques . . . 17
1.3 Mouvements corporels, gestes . . . 18
1.3.1 Geste et contraintes corporelles . . . 18
1.3.2 Geste instrumental . . . 18
1.3.3 Gestes et description du son . . . 19
1.4 Acoustique du mouvement . . . 20
1.4.1 Le mouvement physique . . . 20
1.4.2 Analyse de la dynamique des signaux . . . 21
1.4.3 Transformation et synthèse . . . 22
1.5 Perception du mouvement . . . 23
1.5.1 Psychoacoustique . . . 23
1.5.2 Aspects cognitifs . . . 24
1.6 Conclusion . . . 26

2 Le contrôle de synthétiseurs **27**
2.1 Contrôle “intuitif” . . . 28
2.2 Contrôle gestuel et synthèse de sons musicaux . . . 29
2.3 Approches par algorithmes évolutionnistes . . . 30
2.4 Timbre et contrôle . . . 32
2.4.1 Contrôle et espaces de timbre . . . 32
2.4.2 Timbre et contrôle intuitif . . . 33
2.5 Apport des musiques électroacoustiques . . . 34
2.5.1 Musique électronique . . . 34
2.5.2 Musique concrète . . . 34
2.6 Conclusion . . . 36

3 Approche par catégorisation **37**
3.1 Quels stimuli et quel protocole . . . 38
3.1.1 Quels stimuli ? . . . 38
3.1.2 Quel protocole ? . . . 40

3.2 La catégorisation . . . 41
3.2.1 Théorie de la catégorisation . . . 41
3.2.2 Catégorisation de sons . . . 41
3.3 Test 1 : Classification libre . . . 42
3.3.1 Sujets . . . 42
3.3.2 Stimuli . . . 42
3.3.3 Protocole . . . 43
3.3.4 Résultats . . . 43
3.4 Test 2 : Classification contrainte . . . 48
3.4.1 Sujets . . . 48
3.4.2 Protocole . . . 48
3.4.3 Résultats . . . 48
3.4.4 Discussion . . . 49
3.5 Analyses des signaux . . . 50
3.5.1 Descripteurs de timbres “classiques” . . . 50
3.5.2 Descripteurs spécifiques . . . 52
3.5.3 Sélection de descripteurs . . . 53
3.5.4 Résultats . . . 54
3.6 Conclusion . . . 57

4 Approche par caractérisation graphique 58
4.1 Dessiner les sons, perception multimodale . . . 58
4.1.1 Dessiner les sons . . . 59
4.1.2 Interactions son-image . . . 61
4.2 Interface de caractérisation graphique . . . 62
4.2.1 Choix des paramètres de contrôle . . . 62
4.2.2 Implémentation . . . 62
4.3 Constitution et caractérisation d’un corpus de sons abstraits . . . 66
4.3.1 Présélection des sons . . . 66
4.3.2 Corpus final . . . 69
4.4 Protocole . . . 69
4.4.1 Sujets . . . 69
4.4.2 Conditions . . . 69
4.4.3 Test 3 : Caractérisation du corpus . . . 70
4.4.4 Test 4 : Caractérisation graphique . . . 70
4.5 Résultats . . . 74
4.5.1 Test 3 . . . 74
4.5.2 Test 4 . . . 76
4.5.3 Analyse des trajectoires . . . 76
4.5.4 Résumé . . . 84
4.6 Conclusion . . . 85

5 Analyse par synthèse de mouvements 86
5.1 Analyse des signaux . . . 86
5.1.1 Analyse temps-fréquence . . . 87
5.1.2 Test d’instationnarité . . . 89
5.2 Moteur de synthèse temps réel . . . 89
5.2.1 Stratégie de contrôle . . . 90
5.2.2 Calibration . . . 92
5.3 Test de validation . . . 92
5.3.1 Stimuli . . . 93
5.3.2 Sujets . . . 94

5.3.3 Protocole . . . 94
5.3.4 Résultats . . . 95
5.4 Vers un contrôle générique du mouvement évoqué par les sons . . . 100
5.5 Conclusion . . . 101

6 Discussion et perspectives **102**
6.1 Retour sur les sons abstraits . . . 102
6.1.1 Utilisation des sons abstraits en psychiatrie . . . 102
6.1.2 Le naturel . . . 103
6.1.3 Sur l'utilisation de sons abstraits . . . 104
6.2 Vers une méthodologie générale pour le contrôle intuitif de la synthèse . . . 106
6.3 Applications et perspectives . . . 107
6.3.1 Applications à court terme . . . 107
6.3.2 Applications musicales . . . 108
6.3.3 Perspectives . . . 109

Conclusion **111**

Annexes **113**
A Consignes et questionnaire des tests du chapitre 3 . . . 113
B Consignes et questionnaires des tests du chapitre 4 . . . 115
B 1 Test 3 . . . 115
B 2 Test 4 . . . 119
C Mots utilisés pour décrire des sons abstraits dans le Test 3 . . . 120
D Liste des publications de l'auteur . . . 123

Bibliographie **124**

Introduction

Contrôle de processus de synthèse sonore

Depuis les premiers synthétiseurs électroniques et les premiers sons générés par micro-ordinateur [Mathews, 63], la synthèse sonore numérique a toujours été un domaine en rapide évolution. Il est maintenant possible de reproduire par synthèse, n'importe quel son et ce, de façon très convaincante. Les performances de calcul des ordinateurs ne sont plus un frein à la complexité des programmes dont beaucoup peuvent être exécutés en temps réel. On peut alors se poser la question de savoir pour quelles problèmes il est encore possible d'apporter une contribution. On se doit d'abord de préciser que les méthodes de synthèse sont généralement séparées en deux classes, la première étant la synthèse par modèle physique qui reproduit le fonctionnement d'un système et la seconde par modèle de signal qui reproduit des aspects perçus. Pour ce qui est des modèles physiques, il semble qu'il existe des problèmes ouverts pour ce qui est de reproduire des subtilités de jeu d'un instrument acoustique [Serra, 07] ou d'identifier ce qui différencie un bon instrument de musique d'un instrument médiocre [Bensa *et al.*, 04]. Pour ce qui concerne les modèles de signaux, [Serra, 07] parle d'avancements possible avec la prise en compte simultané du contrôle et de la méthode de synthèse dans la conception d'un synthétiseur. D'ailleurs, aussi bien pour les modèles physiques que pour les modèles de signaux, il semble que la prise en compte du rôle central qu'occupe l'utilisateur et par la même le contrôle qu'il exerce sont des problèmes clés qui n'ont pas encore obtenus de réponses définitives.

Dans [Hunt *et al.*, 02], les auteurs partent du constat que dans un instrument de musique électronique, la production du son et la façon dont cette production est contrôlée sont totalement découplés alors qu'ils sont inséparables dans un instrument acoustique. Les auteurs décrivent une expérience mettant en jeu un simple oscillateur dont on peut contrôler l'intensité et la fréquence ainsi que deux curseurs linéaires. L'expérience montre clairement que pour obtenir un système qui puisse s'apparenter à un instrument de musique, il ne s'agit pas simplement de lier chaque curseur à un paramètre de synthèse mais par exemple de faire en sorte que le son s'arrête quand les curseurs ne bougent pas et utiliser les curseurs pour contrôler la hauteur sur une large plage pour l'un et plus finement pour l'autre. Si on met momentanément de côté l'aspect gestuel, il parait donc évident que les contrôles offerts à l'utilisateur et la façon dont ils sont liés aux paramètres de synthèse sont d'une importance capitale dans l'intérêt final qu'aura un synthétiseur.

Comme le fait remarquer [Serra, 07], les plus grand défis futurs seront probablement dus à l'utilisation de sons diffusés par haut-parleurs et qui occupent une place chaque jour plus importante dans notre vie quotidienne. Il s'agit en fait de maîtriser les informations qui sont transmises à l'aide des sons dans un contexte où le son n'est plus directement lié à une source, c'est-à-dire à une cause mécanique clairement identifiée.
A l'heure actuelle les sons diffusés par haut-parleur sont majoritairement des sons pré-enregistrés quand il ne s'agit pas de sons produits par de la synthèse analogique rudimentaire (type buzzer ou carillons). Même si en théorie, les sons enregistrés offrent infiniment moins de possibilités en termes d'adaptation au contexte d'utilisation, il existe nombre de banques de sons très documentées et contenant des descriptions sémantiques relativement pertinentes par rapport aux besoins des utilisateurs. Il apparaît donc clairement que des synthétiseurs doivent fournir des descriptions du son produit pour qu'ils puissent être utilisés de façon comparable à ce qui est fait avec les banques

de sons.
Actuellement, si on recense les domaines qui font appel à la synthèse, on constate qu'il s'agit principalement du monde de la musique et plus précisément de certains genres de musique (dites musiques "électroniques"). Pour expliquer la relativement faible part des domaines autres que musicaux où la synthèse est utilisée, on peut avancer l'hypothèse que les synthétiseurs ne fournissent pas aux utilisateurs des possibilités de contrôle adaptées à leurs besoins ou en tout cas qu'elle n'exploite pas suffisamment les possibilités de contrôle qui sont théoriquement infinies.

Concevoir un contrôle pertinent est un problème qui pose nécessairement la question de ce qui est perçu par l'auditeur et plus précisément des informations que retire l'auditeur de l'écoute des sons en regard des modifications qu'il a effectué sur les paramètres de contrôle. Ce problème se pose aussi bien dans le cas de modèles de signaux que de modèles physiques même si dans le second cas, notre expérience quotidienne des lois de la physique nous aide à plus facilement appréhender le lien entre les modifications effectuées sur un système et ses conséquences sur le son produit.
Ce problème de faire le lien entre une modification des paramètres de contrôle et les conséquences sur le sonore doit, pour concevoir un algorithme de synthèse, être abordé dans le sens inverse, c'est-à-dire en faisant le lien entre le son et les informations qui en sont déduites par un auditeur ; problème qui peut être qualifié de sémiotique des sons (voir par exemple [Jekosch, 05]).

Sémiotique des sons et synthèse

De nombreuses études s'intéressent à un même grand problème, sous des noms différents : "reconnaissance de sources" [McAdams, 93], "attribution de sens" [Jekosch, 05], "identification" [Ballas, 93] [Gygi *et al.*, 04], [Lutfi, 07], recherche d'invariants [Paavilainen *et al.*, 99], sémioacoustique [Vogel, 99]. Plus généralement, il s'agit de tout ce qui concerne la recherche de corrélats acoustiques à des dimensions perceptives du son, c'est-à-dire de comprendre sur quels éléments structurels se base le système auditif pour récupérer une information donnée. On utilise dans ce qui suit le terme de "sémiotique" des sons qui à l'origine, est un terme issu des sciences du langage et de la philosophie avec également une définition très précise en médecine où il désigne la recherche d'indices caractérisant une pathologie. C'est donc un terme très connoté qui définit un système où toute l'attention se porte sur le signifiant (le signe).

Historiquement, la recherche en synthèse sonore est très liée à cette problématique notamment au travers d'études qui ont permis, à partir de signaux très simples, de reproduire des caractéristiques pertinentes du point de vue de la perception (notamment avec les travaux de [Risset et Wessel, 99] et l'approche de l'analyse par synthèse). L'approche que nous souhaitons suivre se situe donc dans cette même logique mais en prenant le problème à l'inverse puisqu'il s'agit d'identifier des attributs pour les utiliser *a posteriori* comme moyen de fournir un contrôle intuitif pour la synthèse. En d'autre termes, il s'agit de s'appuyer sur les travaux, qui sur la base de tests perceptifs, ont mis en évidence des attributs des sons susceptibles de fournir à l'auditeur une information donnée, pour concevoir des synthétiseurs.
La littérature en sémiotique des sons est vaste et les approches sont souvent très différentes, suivant qu'elles s'intéressent à des sons musicaux, à des sons d'environnement ou à la parole. Il est également évident que suivant le type d'attributs étudiés, les mécanismes mis en jeu et les méthodes pour étudier ces mécanismes vont être différents. De ce fait, on distinguera deux types d'attributs que l'on qualifiera d'attributs **objectifs** pour ce qui concerne des caractéristiques ayant une définition physique et les attributs **subjectifs** pour ce qui concerne les informations qui n'ont pas de définition physique et qui en ce sens, varient suivant les auditeurs.
Les attributs objectifs englobent tout ce qui est lié aux caractéristiques de la source sonore comme la taille, le matériau, la géométrie, le type d'interaction ayant donné naissance au son (impact, frottement, *etc.*) et les grandeurs dynamiques qui y sont liées comme par exemple la force d'impact

ou la vitesse de déplacement de la source. A ce stade, il faut préciser que même si ces d'attributs ont une origine physique quantifiable, leur perception n'en reste pas moins limitée et souvent déformée. A titre d'exemple, [Houix, 03] a montré que les auditeurs n'étaient pas capable de distinguer des barres de différentes longueur et différentes sections sur la seule base de sons produits en impactant la barre à différentes positions. De même, [Aramaki *et al.*, 11] ont montré que le matériau perçu n'est pas nécessairement le matériau réellement à l'origine du son (notamment les impacts sur des vitres en verre qui sont associés à du bois ou du plastique à cause d'une confusion entre l'objet "le verre" et le matériau).
Parmi les attributs subjectifs, on peut citer tout ce qui fait appel à un jugement qualitatif de la part de l'auditeur (qualité d'un produit, le caractère agréable, émotions suscitées...) c'est-à-dire relevant du domaine du "ressenti". Si les attributs objectifs ne sont pas forcément clairement identifiés par des auditeurs, les attributs subjectifs peuvent tout à fait être stables pour un groupe d'individus donné. Par exemple, la notion de qualité/solidité semble perçue de manière consensuelle dans les bruits de fermeture de portières automobiles et a pu être relié à des caractéristiques précises du signal comme l'amplitude des impacts dus aux crochets de la serrure [Bezat, 07].

Dans cette thèse, le parti pris a été de s'intéresser à un attribut à la fois objectif et subjectif : le mouvement évoqué par les sons. Cette notion sera précisée par la suite mais on peut dors et déjà dire qu'il s'agit des informations liées aux variations dans le son qui peuvent être décrites par les auditeurs en termes similaires de ceux utilisés pour décrire le déplacement d'un objet dans l'espace. On s'intéresse exclusivement a des sons monophoniques pour mieux se distinguer des problèmes liées à la position spatiale de sources sonores qui est en partie traduite par des indices binauraux. L'intérêt de travailler avec un attribut à la fois objectif et subjectif est de pouvoir envisager de transposer nos résultats à d'autres problématiques sans se limiter à l'une ou l'autre des catégories tout en ayant des bases physiques sur lesquelles appuyer nos réflexions.

Invariants

La sémiotique des sons, qu'elle soit ou non liée à la synthèse, peut être abordée en considérant l'existence d'invariants du signal. Pour illustrer la notion d'invariant, on peut prendre exemple dans le domaine de la parole où l'on trouve des travaux qui sont parmi les premiers à utiliser cette notion (voir par exemple [Gracco et Abbs, 86]). En effet, des chercheurs se sont posés la question d'identifier dans le son, ce qui caractérise une lettre de l'alphabet donnée et donc quels aspects varient d'un individu à l'autre et quels autres aspects sont nécessaires pour qu'une lettre soit reconnue.
Trois approches fondamentalement différentes sont possibles pour ce problème. La première est de considérer que certains aspects du signal notamment la structure formantique des signaux de parole, sont globalement invariants et que c'est sur ces invariants que l'on se base pour identifier/différencier les informations (théorie portée entre autres par [Stevens, 00]). Par ailleurs, il existe une approche qui met l'invariant au niveau du geste phonatoire même si ces gestes ont des conséquences sonores indiscutables [Liberman et Mattingly, 85]. Enfin, une autre théorie propose de mettre de côté la notion d'invariance pour préférer la notion de variabilité adaptative en considérant que des éléments (du signal) sont utilisés en fonction de leur disponibilité et de leur capacité à discriminer les informations qu'ils apportent à l'auditeur, tout en étant interchangeables et pouvant être remis en question à tout moment [Lindblom, 87].
Dans le cas du problème de la synthèse, il n'y a pas lieu de se poser la question de la nature des invariants puisqu'il n'est pas possible de se baser sur d'autres aspects que ceux qui concernent directement le signal. En revanche, il sera intéressant, en se basant sur ces différentes théories, de se poser la question des origines potentiellement liées à des aspects gestuels ainsi que la façon dont le cerveau utilise les caractéristiques du signal qui peuvent être à l'origine d'une information véhiculée par le son.

Vers le sens des sons

La sémiotique des sons est une thématique qui a fait récemment l'objet d'un intérêt particulier dans le domaine de la synthèse et plus globalement de la réalité virtuelle. Ainsi le projet européen "Sound to sense, sense to sound"[1] a regroupé 11 laboratoires autour de ce thème et a fourni un gros travail de synthèse de l'état de l'art mis en jeu [Leman et Cirotteau, 05].
A plus petite échelle, le projet ANR "Vers le sens des sons"[2] s'est également intéressé à cette problématique en regroupant des équipes issues de la psychoacoustique, des sciences cognitives et de la synthèse sonore.
Cette thèse s'inscrit directement dans le cadre de ce projet de recherche.

Objectifs et méthodologie

Avant de préciser les objectifs de la thèse, il est nécessaire de revenir sur les différents points de départ qui y ont conduits. Initialement, cette thèse était intitulée "Vers une approche acoustique et cognitive de la sémiotique des objets sonores" en référence à une étude [Ystad *et al.*, 05] qui a fortement motivée ce travail. Cet intitulé traduit la volonté de placer l'auditeur et l'utilisateur au centre du travail de recherche en synthèse sonore. Il s'agit également d'intégrer à cette recherche, des résultats et des problématiques qui concernent la perception sonore c'est-à-dire qui sont du domaine de la psychologie expérimentale et des sciences cognitives.
De cette volonté découle un fort intérêt pour l'approche du son suivie par les compositeurs/chercheurs en musiques électroacoustiques notamment au travers des sons utilisés dans ces musiques. En particulier, ils posent la question des informations qu'il est possible de retirer lors de l'écoute d'"objets sonores" c'est-à-dire de sons pris dans des contextes différents de ceux dans lequel ils ont été crées. Il y a donc également au départ de cette thèse, la volonté de mettre en relation les problématiques et les résultats issus de cette musique avec la recherche en synthèse sonore.
Des deux précédents points de départ découle un troisième qui concerne la volonté d'aborder la problématique du mouvement, problématique incontournable en musique et qui, dans certains contextes, dépasse largement le cadre défini classiquement en physique et en synthèse sonore.

L'objectif autour duquel s'articule cette thèse est donc de concevoir un synthétiseur permettant de contrôler des aspects liés au mouvement évoqué par les sons. En pratique, il s'agit de déterminer des paramètres du signal sonore, relativement à la notion d'invariant, qui sont susceptibles de provoquer chez des auditeurs une sensation de mouvement, pour pouvoir utiliser ces paramètres dans un synthétiseur. Au delà de ces objectifs, le but a plus long terme est de définir une méthodologie générale permettant de fournir un contrôle intuitif pour un système de synthèse sonore.
Il s'agit également de caractériser les informations relatives à des mouvements "évoqués" par les sons pour l'analyse et la synthèse sonore. Enfin on souhaite mettre en relation des domaines de recherche complémentaires et notamment les connaissances issues de la musique électroacoustique au service d'une "sémiotique des sons" tournée vers la synthèse sonore.

Démarche

Pour atteindre nos objectifs, il a été nécessaire de tester une approche originale qui ne se base pas sur des connaissances *a priori* du problème pour proposer un synthétiseur, et nous permet donc d'envisager plus sereinement de pouvoir la transposer à d'autres problématiques. La démarche que nous proposons de suivre est composée de trois étapes principales (*cf.* figure 1).

- La première étape est une étape d'évaluation perceptive visant à regrouper des sons qui ont une même caractéristique perçue ou qui représentent différents degrés (ou états) d'une même

1. $S2S^2$, Projet coordonné par l'université de Florence en Italie et terminé en 2005
2. "SenSons", ANR jeunes chercheur de S. Ystad terminé en 2009 `http://www.sensons.cnrs-mrs.fr/`

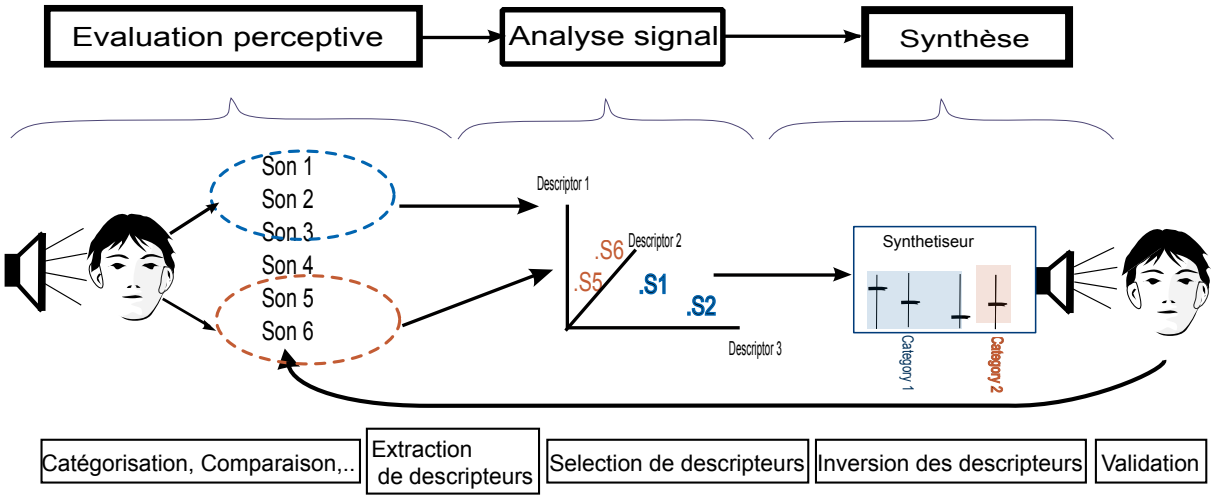


Figure 1 – Synoptique de la méthode générale proposée

caractéristique, sur une échelle perceptive. Cette étape pourra être l'occasion de définir les caractéristiques d'intérêt c'est-à-dire celles sur lesquelles plusieurs auditeurs s'accordent dans leur jugement et qui seront à termes, les paramètres de contrôle proposés à l'utilisateur.
- La seconde étape consiste à analyser les sons au regard de ces propriétés perçues pour identifier les caractéristiques des signaux qui en sont responsables. On cherchera donc des caractéristiques communes à des sons qui auront été jugé similaires ainsi que des propriétés qui les distinguent d'autres sons ayant été jugés différents.
- La troisième étape consiste à contrôler ces caractéristiques des sons, dans des modules de transformation/synthèse sonore. Enfin, il est bien évidemment nécessaire d'évaluer le dispositif de transformation/synthèse pour vérifier qu'il reproduit bien les propriétés perçues pour lesquelles il a été crée.

Aspects méthodologiques

La recherche en synthèse sonore conduit généralement à la création de méthodes, d'algorithmes et plus généralement d'outils qui peuvent être distribués voire commercialisés. Un des principaux enjeux est donc d'identifier et de garder à l'esprit les applications potentielles (voir les applications visées) de l'outil développé qui vont avoir un impact important sur le travail réalisé. Il semble que les contraintes liées aux objectifs quand il s'agit d'outils pour la création (musique, design sonore) sont moins importantes dans le sens où il existe dans ces contextes une place pour l'imprévu. On se place ici dans une démarche où nous souhaitons définir une méthode générale et pas seulement un outil spécifique qui concerne le mouvement. De plus il s'agit de pouvoir à la fois aborder des problèmes précis ayant des applications industrielles ainsi que des problèmes plus abstraits ayant trait au musical.

Cette thèse s'inscrit dans un mode de financement et une logique offrant une grande liberté dans l'organisation de la recherche. Nous avons donc pris le parti de se laisser la possibilité d'aborder des problèmes qui ne rentrent pas directement dans la ligne de conduite dictée par les objectifs énoncés précédemment et de se permettre à tout moment d'explorer des voies de recherche qui paraissent intéressantes. Le présent document ne décrit donc pas l'ensemble des travaux qui ont été menés durant les trois ans dédiés au travail de thèse et se limite aux aspects les plus aboutis.

Organisation du document

L'organisation de ce document suit la démarche classique qui consiste à présenter l'état de l'art pour se situer dans le contexte actuel pour ensuite présenter successivement les différentes étapes du travail effectué. L'étude bibliographique qui est faite ici est relativement large et fait appel à de nombreux domaines de recherche, elle sera donc découpée en deux parties, l'une consacrée au mouvement (chapitre 1) et l'autre au contrôle de processus de synthèse sonore (chapitre 2). Elle sera aussi l'occasion pour nous de mieux définir notre travail, certains points de départ de la thèse étant plus des idées que des faits et observations précises. En particulier, un temps important est consacré pour définir la notion de mouvement évoqué, notion clé de cette thèse.
Nous avons ensuite suivi la démarche globale proposée, en se focalisant sur la première étape d'évaluation perceptive de sons, et ce, au travers de 2 séries d'expériences. La première série aborde l'évaluation perceptive avec une approche par catégorisation, où il était question de regrouper des sons qui évoquent des mouvements similaires, ce qui a conduit à l'identification des grands types de mouvements (chapitre 3). Cette première étude a été suivi d'analyses des sons qui ont permis de préciser les contraintes inhérentes à notre démarche et incité à mettre en place une deuxième série d'expériences.
La deuxième série d'expériences utilise le dessin pour permettre au sujet décrire les informations liées au mouvement qui sont perçues lors de l'écoute de sons abstraits (chapitre 4). Plus précisément, le test se base sur une interface graphique permettant de dessiner des trajectoires à l'aide d'un nombre restreint de paramètres. Les résultats obtenus ont ensuite été vérifiés par une dernière expérience s'appuyant sur des sons de synthèse (chapitre 5).
Ces études perceptives représentent l'essentiel du travail réalisé au cours de la thèse.
Cette thèse a été l'occasion de mener de nombreuses réflexions, notamment concernant la sémiotique des sons, le dernier chapitre (chapitre 6) est donc en partie consacré à des discussions autour de ces aspects. Enfin le dernier chapitre s'intéresse également à des applications qui ont été effectuées et d'autres qui sont encore à l'état de projet, ce qui constitue un ensemble de perspectives relativement vaste.

Vocabulaire

Certains termes fréquemment utilisés dans ce document pouvant être ambigus il s'agit ici de leur donner une définition qui auront cours dans tout ce qui suivra.

Paramètres de contrôle : Paramètres mis à disposition de l'utilisateur d'un synthétiseur. Ils doivent souvent être distingués des *paramètres de synthèse* qui sont les paramètres qui peuvent être modifiés dans l'algorithme de synthèse et qui ne sont pas nécessairement accessibles aux utilisateurs.

Mapping : Bien qu'il s'agisse d'un terme anglais, il n'existe pas de traduction d'utilisation pratique, il sera donc utilisé dans ce document pour désigner la mise en relation entre différents paramètres d'un synthétiseur et principalement les paramètres de contrôle avec les paramètres de synthèse.

Attribut perçu : Tout type d'information perçue par un auditeur lors de l'écoute d'un son. En ce sens, la perception comprend l'ensemble des traitements effectués par système le auditif et le cerveau [3].

Synthétiseur Les limites sont assez floues entre ce qui relève de la synthèse et ce qui relève de la transformation (par exemple, la synthèse granulaire se base sur des échantillons sonores). Lorsque que le terme synthétiseur est employé, nous sous entendons un outil permettant de générer et/ou transformer des sons. Il va sans dire que nous nous intéressons exclusivement à la synthèse sonore numérique.

Avertissement

Dans la version numérique de ce document, le texte affiché en bleu correspond à des liens hypertextes qui renvoient à des pages du document ou a des cibles externes. En particulier, dans la bibliographie, ces liens donnent accès à des sites qui permettent d'obtenir des versions numériques des documents cités.

Des exemples sonores et des vidéos accompagnant ce document sont disponible à l'adresse http://www.lma.cnrs-mrs.fr/~kronland/TheseAMerer.

3. Il est difficile de définir et distinguer ce qui est du domaine de la perception et ce qui est du domaine la cognition comme en témoigne les nombreux échanges qui ont eut lieu à ce sujet sur la liste de diffusion "Auditory list" http://www.auditory.org/mhonarc/2004/msg00323.html. Il semblait donc nécessaire de préciser la définition que nous utilisons.

Chapitre 1

Notions de mouvement

1.1 Mouvement évoqué par les sons **12**
1.1.1 Points de départ 12
1.1.2 Vocabulaire et définitions 13
1.1.3 Motivations 15
1.2 Le mouvement musical **16**
1.2.1 Modèles basés sur des mouvements corporels 16
1.2.2 Modèles mécaniques 17
1.3 Mouvements corporels, gestes **18**
1.3.1 Geste et contraintes corporelles 18
1.3.2 Geste instrumental 18
1.3.3 Gestes et description du son 19
1.4 Acoustique du mouvement **20**
1.4.1 Le mouvement physique 20
1.4.2 Analyse de la dynamique des signaux 21
1.4.3 Transformation et synthèse 22
1.5 Perception du mouvement **23**
1.5.1 Psychoacoustique 23
1.5.2 Aspects cognitifs 24
1.6 Conclusion **26**

Ce chapitre a pour objectif de mieux définir la thématique ou plutôt l'attribut des sons qui est étudié tout au long de cette thèse. En effet, même s'il s'agit clairement de travailler sur les notions de mouvement/dynamique des sons sans considérer les aspects liés à la spatialisation, il semble qu'il soit nécessaire de les préciser. Il est apparu lors des différentes présentations publiques de ce travail, que ces notions paraissent évidentes pour tout le monde mais sont interprétées de façon très différentes suivant le contexte (physique, musique,...) ce qui conduit à de nombreux malentendus. On verra que la notion de mouvement renvoie aussi bien au déplacement d'un objet dans l'espace qu'à des évolutions temporelles ou encore à aspects des sons qui sont plus de l'ordre de la métaphore. Dans un premier temps, il s'agira de clarifier les idées qui nous ont conduites à aborder cette thématique pour ensuite fixer le vocabulaire à utiliser.
Ce chapitre a également pour but de recenser les travaux qui s'intéressent à ces notions ou à des notions proches et qui doivent être pris en compte avant toute recherche. Seront donc abordés ici les aspects liés à la musique, à la physique des sources mobiles et à l'analyse de signaux non-stationnaires ainsi qu'aux mouvements corporels et à la perception auditive du mouvement.

1.1 Mouvement évoqué par les sons

1.1.1 Points de départ

A l'origine de ce travail, se trouve une observation faite sur une classe particulière de sons qui ne peuvent être décrits à l'aide de la source qui les a produits et qui seront appelés dans la suite *sons abstraits*. Le vocabulaire qui est "naturellement" utilisé pour décrire ce type de sons est sans nul doute un point de départ important dans la volonté de s'intéresser aux notions de

mouvement/dynamique.
En effet, dans une étude préliminaire de [Schön *et al.*, 10], les auteurs ont demandé aux sujets d'associer le plus rapidement possible un mot à des sons abstraits (7 sujets ont évalué une base de 70 sons). Parmi les mots obtenus, on trouve la liste suivante :
frottement, recul, arrêt, rapide, vibration, chute, rebonds, secousses, rapprochement, accélération, rotation, freinage, rebondissement, mouvement, saute, flexion, envolée, glisse, frappe, démarrage, tombe, *etc.*.
Ces termes font référence à des aspects dynamiques et/ou sont des métaphores décrivant des phénomènes qui ont une structure temporelle marquée. Il semble donc que cet aspect "dynamique" soit une des caractéristiques importantes pour décrire les sons sans faire directement référence à l'événement qui est à l'origine du son. On peut même avancer l'hypothèse que parmi les informations véhiculées par les sons (en tout cas ce type de sons) celles qui sont du domaine de la dynamique ou du mouvement ont une place non négligeable.

Parmi les points de départ de la thèse se trouvent également les réflexions menées au sein du MIM [1] et qui ont donné lieu aux Unités Sémiotiques Temporelles (UST)[Fremiot *et al.*, 96]. Les UST mettent en avant la gestion du temps comme élément structurel essentiel d'une composition. De plus, cette gestion du temps s'appuie sur des variations, ou la continuation d'un élément et par là des représentations liées au mouvement (*sic*, p.23). Concrètement, les "figures" ou "unités" sémiotiques répertoriées font en grande partie appel à des notions qui sont de l'ordre du mouvement et de la dynamique. Parmi les 20 UST référencées on a : "chute", "trajectoire inexorable", "contracté-étendu", "élan", "freinage", "qui avance", "qui tourne", "qui veut démarrer". La métaphore du mouvement semble donc être un moyen de décrire et surtout de segmenter la musique en éléments qui ont du sens.

Ces deux brefs exemples montrent que la notion de mouvement/dynamique est fondamentale pour la description ou l'analyse de sons. On a également pu constater à travers les termes utilisés, que certains font directement référence au mouvement physique ou à ses caractéristiques (vitesse, trajectoire,...) tandis que d'autres concernent le geste ou les mouvements du corps humain. Il est fait appel à la notion de mouvement pour décrire des évolutions temporelles qui ne sont pas nécessairement les conséquences du déplacement d'un corps dans un environnement et des interactions avec cet environnement.
A ce stade, on peut avancer une première tentative de définition :
On entend par mouvement/dynamique, l'évolution temporelle d'un ou plusieurs éléments qu'on peut décrire en termes :

- D'échelle de temps ou de durée globale des variations (notion de segmentation)
- De forme globale de ces variations
- De vitesse/accélération de ces variations

C'est sur la base de ces éléments qu'il s'agit de fixer le vocabulaire.

1.1.2 Vocabulaire et définitions

Jusqu'ici, nous avons employé les termes mouvement et dynamique et il est nécessaire de justifier ce choix de vocabulaire par rapport à d'autres termes qui seraient susceptible de décrire la notion qui nous intéresse. La liste suivante regroupe des synonymes qui, associés au sonore, pourraient correspondre à notre pré-définition : activité, changement, cinématique, circulation, dynamisme, développement, flux, fluctuation, marche, mobilité, progression, trajet, variation.
Il existe également des attributs liés à la manière dont le son est produit (et à la source sonore) qui sont également à considérer : action, geste, déplacement.

1. Laboratoire Musique et Informatique de Marseille, groupe de recherche composé d'artistes, compositeurs et scientifiques http://www.labo-mim.org/

Bien que la rigueur nous invite à évaluer chacun des termes proposés ci-dessus, il a été nécessaire de fixer des critères de sélection et par là même d'en écarter certains. En premier lieu, le terme utilisé s'inscrit nécessairement dans une littérature qui ne peut être ignorée ce qui exclu toute forme de néologisme ou de termes qui n'ont jamais été utilisés en lien avec le sonore. Ensuite ce terme doit être suffisamment explicite pour ne pas risquer de confusions et de débats sans objets. Enfin en toute logique, ce terme doit s'approcher au plus près des notions et des idées qui nous préoccupent. Nous avons donc restreint la discussion sur le caractère explicite et sur la précision, à des termes qui répondent au premier critère de pertinence au regard de la littérature : déplacement, cinématique, dynamique et mouvement.

Déplacement

En mathématique, on a une définition précise : un déplacement est une isométrie affine directe. Un vecteur suffit pour déterminer entièrement cette transformation qui conserve les dimensions.
Dans un contexte plus général, la notion de déplacement renvoie implicitement à la dualité actif/passif dans le sens où l'on trouve dans les diverses définitions des termes comme "actions de..." ou "faire changer". En anglais, "displacement" renvoie à la notion d'activité et également à des aspects plus péjoratifs : "when people are forced to leave the place where they normally live"[2]. Enfin, aussi bien en français qu'en anglais, le déplacement renvoie au principe d'Archimède et plus précisément à la force définie par ce principe.
De par ces fortes connotations en science et cette notion de passivité, le terme déplacement est donc trop restrictif pour l'usage que nous souhaitons en faire.

Cinématique

La cinématique est la branche de la mécanique qui étudie le mouvement des corps. Le terme cinématique est donc lié à des problèmes de géométrie et de mécanique élémentaire, problèmes dont les liens avec le son ne sont pas évidents. On pourrait parler de "cinématique des sons" ce qui reviendrai à travailler dans un domaine qui s'intéresse au mouvement des sources acoustiques. Il est donc difficile d'utiliser le terme de cinématique en tant que concept ou comme "attribut" des sons, autrement dit comme information perçue en écoutant un son puisqu'il s'agit plutôt d'un domaine d'étude.

Dynamique

En physique, la *dynamique* concerne tout ce qui est relatif aux forces et par le principe fondamental de la dynamique, au déplacement. De manière plus générale, la *dynamique* consiste à considérer des phénomènes dans leur évolution, dans le temps.
Quand on l'applique au sonore, la *dynamique* renvoie presque systématiquement aux variations de l'intensité. En audio, les traitements de *dynamique* sont ceux qui vont modifier le rapport entre les instants de faible amplitude et ceux de forte amplitude (porte de bruit, compresseur, limiteurs, expandeurs). La *dynamique* d'un système de traitement audio correspond au rapport entre le signal admissible de plus grande intensité et le bruit propre du système. Ainsi en musique, on a les exemples suivants :
"Si les moyens purement musicaux d'effectuer cette rentrée sont innombrables, il n'y a que deux procédés dynamiques : le crescendo et le diminuendo" (Dupré, Improvis. orgue, 1925, p. 113)[3] ou encore "Écart de niveau sonore entre les passages les plus forts et les plus faibles d'une œuvre musicale." (Schaeffer, Rech. mus. concr., 1952, p. 187)[3]. On voit donc que dans le contexte général qui nous intéresse, cette confusion possible entre une dynamique générale et la restriction aux

2. Cambridge Advanced Learner's Dictionary - http://dictionary.cambridge.org/dictionary/british/displacement
3. Cité par http://www.cnrtl.fr/definition/dynamique

variations d'intensité nous empêche d'utiliser ce terme par la suite. De plus, *dynamique* dans le langage courant renvoie des notions comme l'énergie, l'efficacité voir l'enthousiasme, notions qui sont relativement éloignées des questions d'évolution et de variations qui nous intéressent.

Mouvement

Le terme *mouvement* renvoie à des significations relativement différentes mais qui toutes véhiculent des idées intéressantes pour notre problème. Le premier sens du mot mouvement est un changement de position (spatiale). Plus généralement les différentes définitions renvoient aux notions de changement, de variations et de fluctuations.
Dans différents contextes, on retrouve la notion d'**ensemble** ou de réunion d'éléments. Par exemple en danse, un mouvement est une suite de gestes et de déplacements alors que dans un contexte politique ou social, on parle de mouvement pour désigner un regroupement de personnes ayant des objectifs communs.
En musique, et plus précisément en musique dite "classique", un mouvement est une **partie** de l'œuvre . La notion de mouvement est donc par ce sens liée au problème de segmentation et d'unités signifiantes.
En anglais, il existe deux termes concurrents et souvent interchangeables pour traduire le mot mouvement : "motion" et "movement". Il semble que la distinction aussi fine soit-elle concerne le caractère actif (movement) ou passif (motion), ce qui se traduit dans l'expression "motion without movement" c'est-à-dire "agitation sans déplacement". Dans la littérature concernant la musique, c'est principalement le terme "motion" qui est utilisé pour parler des évolutions qui nous intéressent puisque "movement" dans ce contexte prend le même sens qu'en français (partie de l'œuvre musicale). D'autre part "motion" dans sa forme verbale signifie faire un signe ou un geste, définition relativement intéressante pour nous, tandis que 'movement" peut également renvoyer à un élément des mécaniques d'horloges et de montres.

En résumé, le terme "mouvement" renvoie non seulement à la notion de déplacement mais aussi aux notions d'ensembles et de parties. Toutes ces significations me semblent concourir à préciser la notion que nous avons tenté de décrire précédemment et le terme de *mouvement* sera donc utilisé par la suite. En revanche pour lever toute ambiguïté, on ajoutera qu'il s'agit "d'évocation" pour ne pas se limiter au déplacement d'objets. Dans ce document on s'intéresse donc au *mouvement évoqué par les sons* même si pour des questions de commodité, on omettra une partie de l'expression et utilisera le terme "mouvement" seul.

1.1.3 Motivations

Comme discuté en introduction, cette notion de mouvement a l'avantage de ramener à la fois à des notions objectives et à des notions beaucoup plus abstraites. On aura donc des bases physiques pour aborder un problème qui est plus de l'ordre du "ressenti" ou de la perception.
S'intéresser à cette problématique sera aussi l'occasion (et même l'obligation) de considérer le signal sonore dans ses aspects non-stationnaires ce qui dans le contexte actuel de recherche semble tout à fait pertinent (*c.f.* section 1.4). Enfin, on a vu les utilisations de la notion de mouvement comme outil de description mais il existe un nombre important d'applications potentielles, aussi bien musicales qu'industrielles qui motivent à s'intéresser à cette problématique. Comme applications industrielles, il pourra s'agir de situations où le son est utilisé pour donner des informations à un utilisateur l'information en question pouvant être "je me déplace dans la direction X et avec une vitesse Y" ou "il est nécessaire de se déplacer". Dans un contexte musical, il s'agira de permettre à l'utilisateur de spécifier une trajectoire et de générer un son en accord avec cette trajectoire.

Dans les sections suivantes, on se propose de présenter comment la notion de mouvement a été abordée dans les domaines divers comme la musique, les sciences humaines, l'acoustique ou les

neurosciences.

1.2 Le mouvement musical

A en juger par le grand nombre d'études rencontrées (dont seul un petit nombre est mentionné ici), un important intérêt s'est porté sur la compréhension des liens semble-t-il indéniables qui existent entre musique et mouvement. D'après [Todd, 95] l'idée de "relations intimes" entre mouvement musical et mouvement physique est très ancienne. Il rapporte entre autres les écrits d'Aristoxène qui parle de mouvement mélodique et de mouvement rythmique et trouve même un exemple encore plus ancien dans l'Egypte antique avec l'expression pour dire "chanter" qui signifie aussi "mouvement des bras". On trouve également des références à un mouvement musical et notamment mélodique dans [Helmholtz, 75, (1875)]. Même si tous les auteurs s'accordent quant à l'évidence de l'existence de liens entre musique et mouvement, toute étude bibliographique se trouve très rapidement confrontée à deux problèmes majeurs. Le premier problème que l'on rencontre est lié au vocabulaire, la notion de mouvement étant souvent utilisée sans définition précise. Par exemple M.A. Harley dans un article concernant l'analyse de compositions de Y. Xenakis, précise dans le titre "Spatial sound movement[..]" [Harley, 94] et utilise ce terme aussi bien pour des aspects de spatialisation que pour les gestes des musiciens et les structures musicales. Le second problème est dû au fait que les objectifs sont souvent très différents, certaines études s'intéressant à l'interprétation musicale c'est-à-dire aux éléments qui varient entre différentes versions d'une même œuvre écrite tandis que d'autres s'intéressent plus globalement à l'analyse de musique et à la recherche de critères définissant la forme ou la structure d'une œuvre musicale. On trouve enfin des études qui sont directement centrées sur la compréhension des liens entre musique et mouvement et dont l'intérêt va se porter sur des aspects de la musique très différents (profil mélodique, variations de tempo au cours ou en fin d'une œuvre, *etc.*) et également sur des définitions et des caractéristiques du mouvement souvent très différentes. Dans ces études, les auteurs recherchent des liens de nature très différente allant de la recherche de lois d'évolution de paramètres musicaux à des comparaisons purement métaphoriques. [Todd, 93] parle d'allusions aux mouvements dans l'interprétation musicale qui sont à la fois "métaphoriques" dans ce qui concerne le tempo et "psychobiologique" dans ce qui concerne le geste et la perception des variations. Enfin, toutes ces études concernent des domaines très différents allant de la musicologie, la perception de la musique ou la composition assistée par ordinateur. On peut résumer ces problèmes par une question posée par [Honing, 03] : "Y a-t-il une vraie expérience perceptive du mouvement lors de l'écoute de musique ou est-ce simplement métaphorique à cause du lien avec le mouvement physique ou humain ?" [4]
Par conséquent, il est fort difficile d'organiser une étude bibliographique sur ce sujet et de mettre en regard les différentes études qui la constitue, les choix effectués étant dans tous les cas discutables. On peut cependant organiser les différents travaux en fonction du type de mouvement qu'elles visent ce qui les rassemble en deux groupes : celles qui font le parallèle entre musique et mouvements du corps et celles qui décrivent la musique avec des modèles mécaniques.

1.2.1 Modèles basés sur des mouvements corporels

Dans [Shove et Repp, 95] les auteurs rapportent différents travaux pionniers qui ont comparé à l'aide d'études empiriques des gestes à des aspects caractéristiques de compositeurs (ces travaux seront décrits plus en détail section 4.1.1). Dans [Friberg et Sundberg, 99] les auteurs ont étudié les liens entre l'arrêt d'une course (effectuée par des danseurs) et le *ritardando*. Le ralentissement de la vitesse des coureurs se fait suivant une loi en racine carrée du temps ce qui correspond au tempo moyen durant le ritardendo.
[Friberg *et al.*, 00] ont utilisé des profils de la force d'un pied appuyant sur une plaque comme

4. traduction de "Is there a true perceptual experience of movement when listening to music, or is it merely a metaphorical one owing to associations with physical or human motion ?"

enveloppe temporelle d'un son (avec un contenu spectral d'orgue) et ont demandé à des auditeurs de décrire les sons ainsi obtenus d'abord de manière libre puis sur la base d'échelles liées au mouvement. Les résultats indiquent que les sons peuvent être catégorisés en termes de mouvement mais que c'est plus le rythme de la marche que le profil de force qui est responsable de l'évocation de mouvements.
[MacDougall et Moore, 05] ont étudié la fréquence des mouvements de la tête durant différentes situations de la vie quotidienne et trouvent une large majorité d'occurrences autour de 2Hz. Cette fréquence correspond à celles observées dans diverses études concernant les tempos spontanément donnés par des sujets en tapant du doigt ou en réglant un métronome.
[Eitan et Granot, 06] ont demandé à des sujets d'associer des mouvements corporels à de brèves phrases musicales. Les caractéristiques de ces mouvements "imaginés" étant collectés à l'aide d'un questionnaire qui s'intéressait entre autres au type de mouvement (courir, marcher, tomber, glisser), à la direction du mouvement, et à l'énergie globale. Les paramètres musicaux étudiés étaient les profils mélodiques, rythmiques et dynamiques, avec des variations régulières de ces paramètres. Les résultats montrent un certain nombre d'effets significatifs comme par exemple le profil mélodique qui est relié à la vitesse et l'énergie ainsi qu'a des déplacements dans les trois directions spatiales et principalement dans le sens inverse (augmentation de la hauteur des notes liée à des déplacements corporels vers le bas).
Neil Todd, un des auteurs qui s'est beaucoup intéressé aux comparaisons entre mouvement et musique, a proposé dans [Todd *et al.*, 02] un modèle du contrôle moteur en réponse à la musique (danse suivant un rythme) après avoir beaucoup travaillé sur des modèles "mécaniques" des variations de tempo [Todd, 92, Todd, 95].

1.2.2 Modèles mécaniques

[Kronman et Sundberg, 87] ont proposé un modèle de *ritardando* basé sur l'équation du mouvement pour une décélération constante (jusqu'à atteindre une vitesse nulle). Ce modèle est ensuite comparé aux variations de tempo mesurées sur des enregistrements de différents concertistes. Le problème étant d'adapter les paramètres du modèle (masse/amortissement ou accélération) pour qu'ils correspondent aux ralentissements observés [Honing, 03].
[Feldman *et al.*, 92] ont étudié à la fois les *accelerando* et les *ritardando* et ont proposé un modèle "mécanique" en considérant la position sur la partition comme correspondant à la position d'un solide soumis à une force dépendant du temps. L'analyse à partir d'enregistrements du répertoire leur permet de conclure sur des variations cubiques du tempo et par conséquent une force quadratique dans le modèle. Les auteurs en déduisent que le lien entre les variations de tempo et le mouvement physique n'est pas "juste une métaphore".
Dans une approche différente, [Palmer *et al.*, 07] ont étudié les contraintes "biomécaniques" du corps humain comme par exemple le couplage entre les doigts du pianiste et soutiennent l'hypothèse que le parallèle entre mouvement physique et musique est dû à ces contraintes.
[Johnson et Larson, 03] soutiennent l'idée que le mouvement musical n'est que métaphorique. Ces métaphores ont pour objet le mouvement physique qui se traduit notamment dans les profils mélodiques (appelés dans l'article le mouvement tonal). Pour les auteurs, la métaphore du mouvement est la conséquence de l'expérience quotidienne que l'on fait des lois de la physique (notamment la gravité) et qui permet de communiquer des idées complexes et parfois floues de ce que l'on perçoit de la musique.

Les critiques portées sur la plupart des études présentées visent principalement la méthodologie qui consiste à développer un modèle mathématique et à y faire correspondre des observations qui du fait de la paramétrisation du modèle peuvent presque toujours s'adapter. La musique instrumentale est générée par un ensemble de gestes et le son produit reflète en partie ces gestes comme le font remarquer [Shove et Repp, 95]. La musique (notamment la musique populaire) est souvent le vecteur de la danse c'est-à-dire de mouvements du corps. De ces deux constats, on entrevoit quelques éclaircissements dans cette profusion de travaux et d'articles. Ainsi, il semble que les

liens entre musique et mouvement se trouvent aussi bien dans la manière dont la musique est construite et évolue au cours du temps que dans les évocations qu'elle génère chez les auditeurs. A ma connaissance, aucune étude ne s'intéresse aux battements de cœur ou à la respiration comme support d'un mouvement musical ce qui est étonnant vu la grande inventivité dont ont fait preuve certains auteurs. Il faut tout de même mentionner que ces recherches ont conduit à mettre en place des méthodes thérapeutiques pour les troubles de la motricité à partir de musique et notamment des aspects rythmiques de la musique (*cf.* [Thaut et Abiru, 10] pour une synthèse de ces recherches).

1.3 Mouvements corporels, gestes

Comme la plupart des sujets traités dans ce chapitre, l'étude des mouvements corporels fait l'objet de travaux qui constituent des domaines de recherches à part entière (on parle souvent de sciences du mouvement).
Parmi les aspects du geste qui sont importants pour notre problématique, deux seront discutés dans d'autres parties de ce document à savoir les liens forts entre geste et perception (section 1.5) et le contrôle gestuel de la synthèse sonore (section 2.2). Dans un premier temps, on s'intéresse donc principalement aux résultats d'études qui ont une portée générale sur la notion de mouvement corporel et la compréhension que l'on en a, notamment au regard des limites imposées par notre corps. Comme on l'a vu dans la section 1.2, les liens entre gestes et son dépassent largement le geste instrumental dans le sens où le mouvement musical est souvent associé aux mouvements corporels qui peuvent donc servir de référence dans la description de phénomènes sonores. Dans un second temps, il sera donc question du mouvement comme support de description des sons.

1.3.1 Geste et contraintes corporelles

Depuis les travaux de Jules Marey (1884) et les premiers chromatographes (*cf.* figure 1.1), de très nombreux travaux ont été menés pour caractériser les mouvements du corps (*cf.* [Blake et Shiffrar, 07] pour un historique de la capture de mouvement), principalement dans le domaine de la vision. On sait désormais que très peu d'informations (quelques points blancs sur un écran) sont suffisantes pour reconnaître le mouvement d'un corps humain. En revanche, cette reconnaissance est très sensible aux relations de phases entre les différents points mobiles ainsi qu'à l'orientation (le mouvement de quelqu'un qui marcherait au plafond ne serait pas reconnu comme tel).

[Wexler et Klam, 01] ont étudié l'impact du contrôle moteur dans une tâche de prédiction de la position de cibles mobiles. Les auteurs ont montré que l'estimation de la position est meilleure quand les sujets sont actifs, c'est-à-dire quand le déplacement de la cible était dirigé par l'action de leur main, que quand ils étaient simples spectateurs. Cette étude s'intéressait également au mouvement des yeux et a montré que le suivi de la cible avec les yeux était susceptible de compenser les pertes de performances dans le cas où aucun geste de la main n'accompagne le mouvement.
L'analyse des mécanismes moteurs fournit des résultats qui dépassent largement leur domaine d'origine comme c'est le cas de [Viviani et Cenzato, 85] où les auteurs travaillaient sur l'écriture et ont défini des règles qui permettent de segmenter tout type de gestes. Les auteurs ont entre autres mis en évidence une loi de puissance 2/3 entre la vitesse et la courbure qui s'applique à d'autres gestes et dont on voit aisément l'utilisation potentielle dans les modèles de synthèse sonore. De plus cette relation met en évidence les interactions entre processus moteurs et perception [Viviani et Stucchi, 92].

1.3.2 Geste instrumental

Le geste instrumental est souvent très complexe et demande un apprentissage sur le long terme. Cet apprentissage à entre autres pour but d'améliorer la précision avec laquelle les gestes sont réalisés. Dans sa thèse, [Rasamimanana, 08] a mis en évidence que les musiciens tendent à réduire

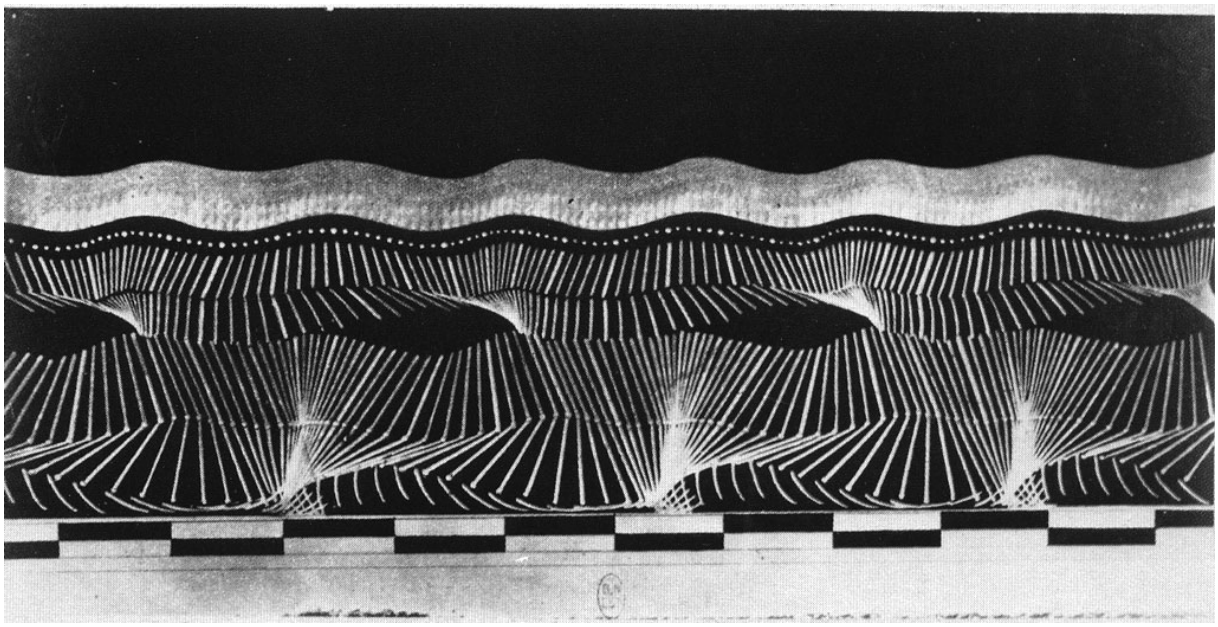


Figure 1.1 – Chronomatographe de Marey réalisé en 1883. Un homme vêtu d'une combinaison noire avec des rayures blanches sur les bras et les jambes est photographié à intervalle régulier permettant pour la première fois de visualiser le mouvement. *Reproduit à partir de* `http://goldberg.berkeley.edu/courses/S06/IEOR-QE-S06/images.html`

ce qu'il appelle le "coût de performance", c'est-à-dire l'effort nécessaire à la production du son. Les mouvements de l'archet et les différents modes de jeu se différencient à l'aide de seulement deux paramètres qui sont l'accélération minimum et maximum. Dans le cas des instruments à corde frottée qui mettent en jeu des liens forts entre gestes du musicien et son produit (notamment entre la vitesse de l'archet et l'intensité), il est intéressant de comprendre ce qui est inhérent aux contraintes dues à la physique de l'instrument et celles qui sont dues aux contraintes biomécaniques du musicien notamment quand le tempo augmente. En musique, tous les gestes ne sont pas destinés à la production du son mais certains sont voués à véhiculer d'autres informations qui sont souvent plus du domaine de l'émotion [Wanderley, 99]. Un vocabulaire a même été posé pour distinguer ces gestes, on parle de gestes *accompagnateur* et de gestes de *production sonore*. Dans une revue des différents critères de classification des gestes et en prenant l'approche descriptive [Ramstein, 91] indique qu'il existe trois types de critères qui sont la cinématique, le spatial et le fréquentiel qui tous trois décrivent les limites des gestes en termes de vitesse d'exécution, d'espace balayé et de périodicité. A noter que dans la réalisation de gestes complexes et particulièrement les gestes du musicien, les différences physiologiques ont une importance notable. Ainsi, [Cossette *et al.*, 00] ont montré que dans le contrôle de la pression dans la bouche par des flûtistes professionnels, différentes stratégies sont utilisées pour utiliser les muscles du conduit vocal mais conduisent toute à une même pression.

1.3.3 Gestes et description du son

Le problème de l'analyse de gestes captés (en anglais motion capture) est très proche de celui de l'analyse de contenus audio comme l'a montré [Müller, 07] qui regroupe dans un même ouvrage une revue des techniques d'analyse automatique de contenus aussi bien pour le son que pour des mouvements captés. L'auteur montre entre autres que certaines stratégies sont identiques notamment en ce qui concerne la recherche d'attributs qui évoluent au cours du temps et qui sont souvent invariants suivant certaines déformations. Dans les deux cas, les modèles intègrent des contraintes comme l'impossibilité "physique" de réaliser certains gestes ou la prise en compte des

lois d'harmonies en musique. Enfin, les deux domaines partagent plusieurs applications comme celles liées à la réalité virtuelle et au contrôle gestuel de la musique.
Le geste peut servir de support pour une description des sons comme c'est le cas dans [Smalley, 97] qui base son système de description "spectromorphologique" sur le geste instrumental et l'ensemble de lois/contraintes y afférent. De même [Godøy, 06] étend la classification de Schaeffer en ajoutant une dimension gestuelle. Il s'appuie sur des types de gestes qui sont potentiellement à l'origine des sons pour proposer des objets "sonores-gestuels" comme base d'une description qui peut s'appliquer à n'importe quel type de sons.
Dans [Caramiaux *et al.*, 09], les auteurs ont demandé aux sujets de réaliser des gestes en lien avec différents types de sons qui leur étaient présentés. Les mouvements obtenus étaient décrits par 7 variables (position, vitesse, accélération normale/tangentielle, courbure, rayon et torsion) et les analyses ont mis en évidence des liens geste-son comme par exemple entre sonie et position d'une part et entre vitesse/accélération et acuité d'autre part pour des sons de flûte.
[Lindblom et Sundberg, 71] ont étudié les liens entre les mouvements des lèvres et les transformations sonores produites. Même si ces travaux visaient principalement l'amélioration des modèles phonatoires, ils illustrent précisément la relation entre les gestes phonatoires et les limites physiologiques (comme l'ouverture de la bouche) avec les conséquences sur les sons produits (décalage des formants).

[Vines *et al.*, 04], ont montré qu'il existe de fortes similitudes dans les informations qui découlent des données visuelles (des gestes du musicien) et les informations qui découlent du son produit en ce qui concerne la notion de tension musicale et pour l'analyse du phrasé du musicien. Les auteurs ont demandé aux auditeurs de bouger un curseur pour suivre ces deux aspects lors de la présentation soit d'une vidéo sans son soit du son seul et ont montré par des méthodes d'analyse fonctionnelle des données, que les informations perçues sont très proches.

Les gestes sont également utilisés pour décrire le timbre d'instruments de musique comme l'ont montré [Traube et Depalle, 04] dans le cas de la guitare où les musiciens font le parallèle entre geste phonatoire et timbre. Enfin, [Van Nort, 09] s'appuie sur une description du sonore via le geste pour proposer des stratégies de contrôle du son qui soient pertinentes (ce qui nous ramène au problème de production du son).

1.4 Acoustique du mouvement

1.4.1 Le mouvement physique

La notion de mouvement est intrinsèque au son, celui-ci étant généré par le déplacement d'un corps se propageant à une vitesse dépendant du milieu. La modélisation des effets d'un mouvement quelconque sur la génération et la réception dépend de beaucoup de paramètres et il existe peu de lois générales. Dans un premier temps, on peut distinguer deux cas :
- Le cas de la source en mouvement (ex : voiture qui roule) avec pour conséquence sur les sons des modifications sur la propagation du son
- Le cas du son produit par un corps en interaction avec le milieu dans le lequel il se déplace (ex : bille qui roule) avec un lien important (et souvent complexe) entre les caractéristiques du déplacement de l'objet et le son généré.

Pour le premier cas, on se doit ici de mentionner le célèbre effet Doppler découvert par Christian Doppler en 1842 et dont on doit les premières observations expérimentales à John Scott Russel en 1848 [Andrade, 59]. L'effet Doppler pour une source monochromatique dans le cas où les vitesses de la source V_s et de l'auditeur V_a sont faibles par rapport à la vitesse du son [5] c, la fréquence de la source vue par l'auditeur est multipliée par la quantité $1 - \frac{(V_s - V_a)}{c}$.
L'effet d'une source (ou d'un observateur) mobile modifie également l'intensité. En plus de la

5. Approximation non nécessaire pour obtenir une expression simple de l'effet mais qui correspond mieux aux cas qui nous intéressent.

variation de la pression proportionnelle à l'inverse de la distance, la pression (vue par l'observateur) est multipliée par un facteur $(1 - \frac{(V_s - V_a)}{c})^{-2}$.
Le cas limite de l'objet qui se déplace à des vitesses supérieures ou égales à la vitesse du son (dans le milieu considéré) est celui de la création d'ondes de choc. Le front d'onde crée par le déplacement de l'objet est rattrapé voire dépassé par la source sonore ce qui crée une accumulation et des discontinuités de la pression. Dans le cas d'un avion se déplaçant à une vitesse supérieure à la vitesse du son et pour un auditeur au sol, celui voit une onde en "N" dont le maximum d'énergie se situe autour de 2 à 3Hz. A noter qu'aucun son n'est perçu avant le passage de l'avion.
Dans le cas d'une source omnidirectionnelle en champ libre et si on néglige les effets dus à l'absorption de l'air, la pression varie suivant l'inverse de la distance entre source et auditeur, (l'intensité varie donc en $1/r^2$). Plus généralement, on aura donc toujours une variation de l'intensité avec la distance source auditeur mais qui dépend de la directivité de la source (qui très souvent a une dépendance fréquentielle). Enfin la propagation dans un milieu non libre implique des réflexions qui s'ajoutent au signal atteignant l'auditeur, le déplacement impliquant des variations souvent très complexes des caractéristiques de ces réflexions. En fonction des échelles de temps mis en jeu, on parlera alors de filtrage, de réverbération ou d'échos.
Toutes ces transformations sont prises en compte dans les modèles de synthèse proposés par [Chowning, 71] et [Moore, 83].

Le cas où la production du son est due au déplacement d'un corps est encore plus complexe. Il faut dans un premier temps distinguer l'état des milieux qui interagissent pour générer le son (solide-solide, fluide-solide,...). Ces différentes interactions font l'objet de disciplines à part entière, l'aéroacoustique pour les interactions fluide-fluide ou fluide solide, la vibroacoustique pour les interactions solide-solide. Il est donc totalement vain de chercher des lois générales qui relient des caractéristiques du son et des caractéristiques du mouvement à l'origine du son. En revanche, quelques cas particuliers ont fait l'objet de recherches appliquées à la synthèse sonore comme c'est le cas pour l'objet qui roule sur une surface solide. [Stoelinga *et al.*, 03] ont observé que les réflexions sur les bords de la surface de roulement produisent un "filtrage en peigne" dont l'écart entre les pics varie avec le déplacement de l'excitateur. Pour la synthèse d'objets qui roulent (et par conséquent dont la position par rapport au bord évolue) les auteurs préconisent donc l'ajout d'un filtrage dont les fréquences des zéros f_n sont reliées aux distances par rapport aux bords $L_i(t)$ par la relation : $f_n = (\frac{(2n+1)c}{2L_i(t)})^2$ avec $n = \{1, 2, 3, ...\}$ et c est la vitesse de propagation dans la direction du bord (et qui peut dépendre de la fréquence). Dans le cas de rebonds, on a un lien simple entre décroissance d'amplitude et intervalle entre les impacts d'une part et caractéristiques du mouvement (vitesse initiale, angle,...) [Rath et Fontana, 03]. Par exemple, dans le cas idéal d'une chute non amortie le temps entre deux impacts t_{impact} est proportionnel à la vitesse initiale v_0 : $t_{impact} = \frac{2v_0}{g}$, où g est l'attraction de la terre.
En ce qui concerne les interactions fluides solides, les problèmes sont équivalents si l'une ou l'autre des parties est stationnaire. La production du son est principalement due aux phénomènes de turbulence, les travaux de Lighthill, [Lighthill, 52, Lighthill, 54] a montré comment la pression acoustique augmente en fonction de la vitesse (pour des géométries particulières). Dans le cas d'un objet cylindrique de dimension caractéristique, D et à des faibles nombre de Reynolds, la création de tourbillons se fait à une fréquence f qui est proportionnelle à la vitesse de l'écoulement V, le nombre de Strouhal fD/V étant constant ($\simeq 0,2$). Même s'il existe des relations entre mouvement (du fluide ou de l'objet) et paramètres des sons, on a souvent à faire avec de fortes non-linéarités (comme celles exploitées dans les instruments de musique) qui rendent ces relations parfois très complexes.

1.4.2 Analyse de la dynamique des signaux

Descripteurs de timbre dynamiques

Depuis les études sur la notion de timbre (depuis [Helmholtz, 75, (1875)]) et la recherche de corrélats acoustiques des dimensions perceptives du timbre (par exemple [Berger, 64, Grey et Gordon, 78, McAdams *et al.*, 95]), un grand nombre de *descripteurs de timbre* ont été proposés pour caractériser les signaux sonores. Beaucoup de travaux ont également été mené dans le domaine du "Music Information Retrieval" (MIR) pour tout ce qui concerne l'analyse appliquée à la transcription l'indexation automatique. Suivant les objectifs, les besoins en termes de description du signal sont très différents, par exemple la reconnaissance d'instruments de musique solistes peut parfois atteindre des très bonnes performances avec seulement quelques coefficients cepstraux [Krishna et Sreenivas, 04] alors que des approches beaucoup plus complexes peuvent être nécessaire pour des enregistrements polyphoniques [Aucouturier *et al.*, 07]. On trouve même dans [Pachet et Roy, 07] un système de génération automatique de descripteurs qui conduit à 10^20 descripteurs différents. MPEG-7 [6] [Kim *et al.*, 05] propose maintenant un recensement et une normalisation du calcul de ces descripteurs qui permet de travailler sur des bases communes. Même si il existe quelques descripteurs qui s'intéresse à des variations au cours du temps (comme le flux spectral [Dixon, 06]), le problème principal qui se pose pour l'analyse des aspects dynamiques des sons concerne la difficulté de travailler avec des profils temporels de descripteurs [Bloit *et al.*, 09a].

Mesure d'instationnarité

L'analyse temps fréquence est particulièrement adaptée pour la description des signaux dans leur évolution. C'est sur cette base que [Xiao *et al.*, 07b, Xiao *et al.*, 07a] ont proposé une mesure d'instationnarité des signaux. L'idée est de comparer chaque trame d'une analyse temps fréquence au spectre "moyen". En pratique, il s'agit de calculer la distribution statistique obtenue par tirage aléatoire lorsque l'on considère le spectre du signal avec des phases aléatoires (on obtient alors des substituts "stationnarisés"). La méthode répétée avec des fenêtres d'analyses de tailles différentes permet non seulement de tester l'hypothèse nulle de stationnarité du signal et de quantifier la non-stationnarité, mais en plus de déterminer la ou les échelles d'observation où celle-ci est maximale. Différentes mesures de distances entre les spectres locaux et le spectre moyen peuvent être utilisées pour modifier l'importance relative des variations de faibles et de hautes amplitudes. La non-stationnarité (dans les cas où l'hypothèse nulle est rejetée) est fonction du rapport entre la moyenne du carré des écarts entre spectre moyen et spectre locaux avec la moyenne cette même quantité obtenue par tirage aléatoire avec les substituts :

$$INS = \sqrt{\frac{\frac{1}{N}\sum_1^N \left(d(S_n,\overline{S}) - \frac{1}{N}\sum_1^N d(S_n,\overline{S})\right)^2}{\frac{1}{J}\sum_j \frac{1}{N}\sum_1^N \left(d(S_{n,j},\overline{S}) - \frac{1}{N}\sum_1^N d(S_{n,j},\overline{S})\right)^2}} \tag{1.1}$$

où S_n est le spectre local à l'instant n (temps discret de durée N), $\overline{S}$ le spectre moyen, $d(,)$ la mesure de distance et l'indice j correspond aux substituts. A noter que ces travaux font suite à de plus anciens exploitant les mêmes idées de comparaisons de spectre à court terme [Martin et Flandrin, 85] [Laurent *et al.*, 98].

1.4.3 Transformation et synthèse

Dans ce domaine, le mouvement est souvent lié à la notion de spatialisation, il n'en reste pas moins qu'une partie de ses travaux sont applicables à des sons monophoniques.

6. Motion Picture Expert Group qui définie les différentes normes en matière audiovisuel, la norme MPEG-7 s'intéresse plus particulière au à l'analyse des médias sur support numérique

Un des pionniers dans ce domaine est sans conteste [Chowning, 71] qui a proposé un système de spatialisation sur quatre haut-parleurs prenant en compte à la fois l'amplitude de chaque enceinte mais aussi, la réverbération et l'effet doppler (voir également [Kronland-Martinet et Voinier, 08]). Concernant la réverbération, Chowning considère à la fois des effets locaux (premières réflexions) et des effets globaux, idée reprise et développée par [Moore, 89] et [Yadegari *et al.*, 02]. La plupart des travaux en synthèse qui s'intéressent au mouvement se focalisent sur la spatialisation et le plus souvent en utilisant les fonctions de transferts de la tête (HRTF) [Evans, 98], [Tsakostas, 07] mais également avec d'autres techniques de spatialisation comme la synthèse de champ (WFS) [Kimura *et al.*, 05]. Pour la spatialisation HRTF, outre les problèmes pratique de coût de calcul, il s'agit de s'affranchir des problèmes dus au fait qu'il s'agit d'une mesure discrète des modifications de l'onde incidente sur chacune des deux oreilles en fonction de l'angle d'incidence, notamment par interpolation des fonctions de transfert [Cheng et Wakefield, 01]. Certains auteurs [Jenison *et al.*, 98] [Iwaya et Suzuki, 07] ont proposé d'améliorer ces systèmes de transformation/synthèse en intégrant les variations de hauteurs dues à l'effet Doppler.
En ce qui concerne la synthèse de sons produits par le mouvement d'objets en interaction (frottement, roulement, rebond,...) une vaste littérature existe avec des approches par modèle physique et par modèles de signaux (voir par exemple [Rath *et al.*, 03] pour une revue de ces méthodes). Les modèles physiques sont particulièrement adaptés quand il s'agit de générer simultanément son et image [Cadoz *et al.*, 82], [O'Brien *et al.*, 01], [van den Doel *et al.*, 01] et où des notions comme le réalisme ou le naturel entre en compte dans la synthèse. Enfin, à la frontière entre mouvement physique et geste, [Gibet, 87] s'est intéressée à la synthèse de sons musicaux en modélisant le geste des instrumentistes et l'interaction avec son instrument.

1.5 Perception du mouvement

Le son nous renseigne en permanence sur ce qui se passe tout autour de nous et en particulier à 180 ° derrière notre tête. Les positions mais plus encore les changements de positions des éléments environnants sont des informations qui touchent à la sécurité voire à la survie des espèces animales. Il n'est donc pas étonnant que la perception sonore du mouvement soit un domaine aussi important mais aussi complexe à étudier.

1.5.1 Psychoacoustique

Discrimination de positions et de vitesses de sources en mouvement

Les études psychoacoustiques sur la perception du mouvement de sources mobiles sont très liées à celles qui concernent la capacité du système auditif à localiser une source fixe dans l'espace. Les mouvements étudiés sont tous situés dans le plan horizontal, principalement dans le cas d'un mouvement angulaire (*i.e* avec une distance source auditeur constante). On trouve par exemple des travaux qui concernent le plus petit mouvement d'angle audible [7] qui font suite aux études sur la discrimination de positions (plus petit angle audible). Ces seuils dépendent beaucoup des conditions expérimentales notamment du type de stimuli utilisés, allant de environ 1 ° pour des bruits large bande [Harris et Sergeant, 71] et augmentant pour des bruits bande étroite ou des sons purs. Les performances dépendent également du système de simulation du mouvement : sur des systèmes allant de 2 à n haut-parleurs disposés en cercle, au casque (panning d'amplitude), ou en utilisant des HRTF (personnalisées ou non). D'autres études se sont intéressées à la précision dans la discrimination de vitesses angulaires [Grantham, 86] sur la base de sons purs (500Hz) diffusés en stéréo (avec contrôle du panning) ainsi que [Carlile et Best, 02] avec des bouffées de bruit en utilisant des HRTF. Dans le même registre, [Agaeva, 04] a étudié la discrimination de vitesse dans le cas de trajectoires dans le plan vertical (par des haut-parleurs placés sur un arc de cercle). La

7. En anglais MAMA pour Minimum Audible Movement Angle

grande majorité de ces études obtiennent des seuils qui augmentent avec la vitesse de référence, allant de 2 deg/s pour une vitesse de référence de 15 deg/s jusqu'à environ 20 deg/s pour une référence a 150 deg/s.
[Rosenblum *et al.*, 87] puis [Lutfi et Wang, 99] et [Kaczmarek, 05] se sont intéressés aux importances relatives de différents indices acoustiques (différences de temps d'arrivée sur les deux oreilles, intensité et transposition doppler) pour différencier des mouvements avec des vitesses, accélérations et positions finales différentes. Les résultats indiquent que ces indices sont tous susceptibles d'être utilisés avec notamment une préférence pour l'effet doppler quand la vitesse est suffisamment élevée pour que la variation de hauteur soit perceptible. Bien entendu, il existe des variations de performances en fonction de la vitesse de référence et du contenu spectral des stimuli utilisés dans les expériences.

Dans la plupart des études aussi bien concernant les discriminations de vitesse que d'angle, les auteurs obtiennent une variabilité inter-individus non négligeable qui met évidence des différences de stratégies et/ou de performances des auditeurs. Ceci montre également la complexité du problème étudié avec les problèmes méthodologiques qui en découlent. Par exemple, [McBeath et Neuhoff, 02] ont montré que des variations d'intensité peuvent induire des variations de la hauteur perçue qui sont importantes ce qui pose problème quand on cherche à découpler les effets de différents indices acoustiques (ici doppler et intensité).
Une des questions sous-jacente à ces études est de savoir si on perçoit directement le mouvement ou s'il est déduit de la position. Il n'y a pas de réel consensus pour ce problème, notamment à cause du fait qu'il met en jeu des processus cognitifs complexes et dont certains aspects sont développés dans la section suivante.

Attributs physique des sources en mouvement

Dans [Houben *et al.*, 04], les auteurs s'intéressent à la vitesse mais également à la taille perçue dans le son produit par des objets qui roulent sur la base de sons enregistrés. Les résultats obtenus montrent que des sons issus de sources dont les vitesses diffèrent de plus de 30% seront perçues comme étant différents. En revanche, les performances chutent nettement quand taille et vitesse varient simultanément ce qui s'explique en partie par le fait que ces variables impliquent des modifications qui touchent des mêmes caractéristiques du signal et qui créent des indéterminations. Dans un tout autre domaine, [Li *et al.*, 91] se sont intéressés à la perception des bruits de pas et plus précisément au genre du marcheur. Les résultats varient fortement d'un stimulus à l'autre (certains obtiennent 94% de bonnes réponses tandis que d'autres se retrouvent sous la barre des 50%). Ces performances sont corrélées aux physiques (taille, poids) des marcheurs à l'origine des stimuli et par là même aux propriétés acoustiques des sons générés (répartition spectrale de l'énergie et loi d'amortissement).

1.5.2 Aspects cognitifs

Les études concernant le mouvement et l'audition en neurosciences se sont beaucoup intéressées à l'identification des zones du cortex activées (lobes pariétales inférieurs, planum temporale, aires pré-moteurs,*etc.*) [Griffiths *et al.*, 98] [Griffiths *et al.*, 00] qui sont similaires pour des mouvements de sources dans le plan horizontal et dans le plan vertical [Pavani *et al.*, 02]. A noter que ces résultas se retrouvent en utilisant différentes méthodes d'imagerie cérébrale (tomographie par émission de positron, résonance magnétique) [Warren *et al.*, 02]. [Beer et Röder, 04] ont également montré l'influence de la tâche et précisément de l'attention portée par les sujets sur la perception du mouvement aussi bien au niveau auditif que visuel. Par contre, les réponses cérébrales sont différentes suivant la méthode de simulation de mouvement [Getzmann et Lewald, 09]. Les auteurs ont ainsi montré que les résultats les plus nets en termes de temps de réponse et d'amplitude

des potentiels évoqués [8] sont obtenus dans le cas où le mouvement est simulé en utilisant un ensemble de haut-parleurs plutôt que la spatialisation au casque par HRTF ou en manipulant ITD et ILD [9]. Enfin, il est relativement connu [Salmelin *et al.*, 95, Brovelli *et al.*, 02] que des sollicitations motrices conduisent à l'émergence d'une composante autour de 20Hz (composante gamma) dans les potentiels évoqués, mais à notre connaissance aucune mise en évidence n'a été établie avec des stimulations sonores. Pour notre problème il est relativement difficile d'identifier les implications de la plupart des résultats obtenus dans le domaine des sciences cognitives. En revanche, l'identification de corrélats électrophysiologiques ouvre des pistes méthodologiques nouvelles pour le problème du contrôle de la synthèse, permettant d'envisager des mesures objectives de la perception du mouvement.

Liens action/perception

Un des résultats très importants en sciences cognitives est celui obtenu par [Pellegrino *et al.*, 92] qui ont mis en évidence l'activation de neurones de façon similaire quand des gestes sont effectués et quand on perçoit une personne tierce en train d'effectuer ces gestes (on parle de neurones miroirs). Ce résultat bien connu, au moins pour ce qui est des stimuli visuels, met en évidence des liens qui existent entre perception et action. Dans le domaine du sonore, plusieurs études ont montré des résultats similaires notamment pour des claquements de langue et de doigts [Hauk *et al.*, 06]. Les auteurs ont montré que ces sons se démarquent, au niveau électrophysiologique, de stimuli aux propriétés acoustiques similaires, mais qui ne peuvent pas être associés à des gestes.
On peut également voir dans ce lien entre motricité et perception, une explication des différences entre individus ayant un développement et des capacités moteurs différentes [Aglioti et Pazzaglia, 10]. Ces résultats ont donc poussé certains chercheurs à concevoir une théorie de la perception qui met au centre les aspects liés à la motricité (théorie motrice de la perception) [Pellegrino *et al.*, 92].
Il existe également plusieurs études parmi lesquelles [Farrer *et al.*, 03] qui montrent que des gestes effectués simultanément à une stimulation sensorielle modifient la perception de cette dernière. En particulier, [Lackner et Shenker, 85] ont montré que la localisation d'une source simulée par ITD et ILD dépendait de la position des membres et de la représentation de cette position que se font les sujets. Ainsi en stimulant avec des vibrations, certains muscles du bras, il est possible d'en modifier la perception de la position, et ces modifications influent sur la localisation qui est faite d'une source sonore attachée au bras. Cette expérience est à rapprocher de l'effet "ventriloquist" [10] puisque dans le processus d'intégration multimodale la localisation de la source est principalement déduite des informations provenant de l'une des modalités concernées.
Par ailleurs, quand on s'intéresse à la localisation d'une source sur la base des différences inter-aurales, il existe une ambiguïté potentielle qui peut être levée grâce à des mouvements de la tête [Wallach, 40] mais pas par des déplacements de la source [Wightman et Kistler, 99]. Encore une fois il s'agit d'un lien entre motricité et perception auditive d'aspects liés au mouvement.

Persistance et mouvement induit

Dans le domaine de la vision aussi bien que dans le domaine du sonore, quand nous sommes soumis à un stimulus qui se déplace dans une direction donnée (toujours la même) pendant suffisamment longtemps et qu'on nous présente ensuite un stimulus stationnaire, il est perçu comme se déplaçant dans la direction opposée à celle du premier stimulus [Grantham, 89]. Il s'agit donc d'un effet de persistance [11] qui modifie ce qui est perçu. L'effet est plus marqué pour des sons de

8. Réponse électrique mesurée sur le cuir chevelu, voir notamment [Mauguière et Fischer, 07] pour une revue du problème.
9. Pour Inter-aural Time Difference and Inter-aural Level Differences qui sont respectivement les différences de Temps d'arrivée et d'intensité de l'onde sonore entre les deux oreilles.
10. effet d'interaction entre visuel et auditif dans la localisation de sources qui sera décrit plus en détail section 4.1.2
11. en anglais le terme utilisé est motion "aftereffect"

mêmes caractéristiques et d'autant plus s'ils sont riches spectralement [Grantham, 98]. Cet effet est également dépendant de la trajectoire dans le sens où il est surtout observé pour des mouvements dans le plan horizontal et avec une distance source auditeur constante [Neelon et Jenison, 03].
Dans le même registre mais avec beaucoup moins d'évidence dans le domaine du sonore, on trouve le phénomène d'illusions de mouvement propre. L'exemple classique est celui des trains à l'arrêt dans une gare : si un train autre que celui dans lequel on se trouve, démarre, on a l'impression que c'est le nôtre qui démarre. Plusieurs expériences ont réussi à produire un effet similaire en utilisant uniquement des sons et de la spatialisation sur haut-parleur, la mesure de mouvement "induit" étant effectué grâce à des mesures de mouvement de la tête ou du thorax (*cf.* [Sakamoto *et al.*, 97] pour une revue de ces différentes études).
On a donc ici deux exemples "d'illusions" auditives qui sont directement liées au mouvement et dont il faudra nécessairement tenir compte dans la suite et notamment dans le cas de tests d'écoute où beaucoup de sons sont écoutés les uns à la suite des autres pour les effets de persistance.

Sources s'approchant de l'auditeur

Il existe une littérature particulière pour le cas de la source qui s'approche [12] de l'auditeur en opposition au cas où elle s'éloigne. Au départ, [Neuhoff, 98] a montré que dans une tâche d'estimation de variations de sonie, les auditeurs avaient tendance à attribuer plus de changement quand la sonie augmentait que quand elle diminuait pour des variations identiques. Ensuite, il a été montré que les auditeurs sous-estiment systématiquement le temps d'arrivée quand une source s'approche d'eux, et pensent que la source arrive plus tôt que la réalité ce qui n'est pas le cas pour des sources qui s'éloignent [Neuhoff, 01]. Ces dissymétries se retrouvent également dans des mesures objectives comme celle de l'activité cardiaque (par électromyographie) qui s'accélère significativement lors de l'écoute de sources qui s'approchent [Bach *et al.*, 09] [Tajadura-Jimenez *et al.*, 10]. Les auteurs parlent d'impact émotionnel de ce type de sons et s'entendent pour dire qu'il s'agit là d'une fonction de détection de dangers potentiels. A noter que cet effet d'alerte semble plus fort chez les femmes que chez les hommes [Neuhoff *et al.*, 09]. Il va sans dire que le cas des sources qui s'approchent devra être pris en compte dans ce qui suit ne serait-ce qu'au travers de sons dont la sonie augmente.

1.6 Conclusion

Les travaux de recherche s'intéressant aux mouvements sont extrêmement nombreux et demandent souvent de faire des passerelles entre des domaines très différents. On ne peut parler d'une seule notion de mouvement mais bien de mouvements qui suivant le contexte prennent une signification différente à la fois déplacement d'un objet, geste ou simple évolution au cours du temps.
Nous avons choisi de nous intéresser à un mouvement qui permet avant tout de décrire les sons dans leur évolution quelle que soit l'interprétation qui en ai faite, le mouvement *évoqué* par les sons. Le choix de cette problématique implique de s'intéresser au signal sonore avec des méthodes particulières et qui sont tout à fait à l'ordre du jour dans le contexte actuel de recherche. La perception et particulièrement la perception sonore du mouvement est un problème complexe dont beaucoup de phénomènes sont encore mal compris. Enfin, même si le problème est complexe, il semble évident que le mouvement est un aspect qu'il est très intéressant d'étudier dans le contexte de la synthèse sonore, pour donner des contrôles pertinents aux utilisateurs.

12. en anglais "Looming sounds" ce qui signifie apparaître mais aussi menaçant

Chapitre 2

Le contrôle de synthétiseurs

2.1 Contrôle "intuitif" **28**
2.2 Contrôle gestuel et synthèse de sons musicaux **29**
2.3 Approches par algorithmes évolutionnistes **30**
2.4 Timbre et contrôle **32**
2.4.1 Contrôle et espaces de timbre 32
2.4.2 Timbre et contrôle intuitif 33
2.5 Apport des musiques électroacoustiques **34**
2.5.1 Musique électronique 34
2.5.2 Musique concrète 34
2.6 Conclusion **36**

Le contrôle en synthèse sonore, est l'ensemble des modifications qu'un utilisateur va pouvoir effectuer sur le son produit. Lorsque l'on définit une méthode de synthèse, on définit donc un ensemble de modifications possibles qui seront accessibles à travers des paramètres que l'on appellera ici *paramètres de synthèse*. Dans la plupart des cas, ces paramètres ne peuvent être utilisés directement, pour des raisons très différentes. Si le nombre de paramètres est trop élevé (plusieurs centaines en synthèse additive) il devient très difficile de les manipuler directement. Un paramètre qui n'est défini que de façon mathématique peut très bien apporter des modifications sur le signal qui ne seront pas audibles ou qui seront trop complexes pour être comprises et reproduites par un utilisateur. Plusieurs paramètres dont les variations conjointes produisent des effets non linéaires sur le signal pourront également être très difficiles à manipuler. Même dans les cas où les paramètres correspondent à des grandeurs physiques, celles-ci peuvent être difficiles à mesurer ou ne pas être d'intérêt pour l'utilisateur. Enfin, dans le cas de la synthèse permettant d'accéder à des sons qui "n'existent pas dans la nature", on sera très souvent désemparé pour définir les contrôles qui sont intéressants, et en donner des définitions compréhensibles par les utilisateurs. Dans ce qui suit, on s'intéresse plus particulièrement au contrôle pour la synthèse par modèle de signal qui seule est envisageable dans le cas d'attributs subjectifs qui ne peuvent être décrits avec des lois physiques.
Le problème du contrôle est donc de définir les relations adéquates entre des paramètres de synthèse et les manipulations qui seront permises à l'utilisateur, pour que le synthétiseur ainsi constitué trouve tout son intérêt. Le terme anglais **mapping** qui désigne cette problématique signifie"cartographie" quand il est employé seul, ce qui illustre l'idée de réseau et de relation complexes entre les différents niveaux de paramètres. Si on considère le schéma de mapping le plus complet possible, on peut le décomposer en 4 étapes successives (*c.f.* figure 4.6). La première étape de *description du geste* consiste à recueillir la volonté de l'utilisateur au travers d'interfaces homme-machine. Il s'agit souvent de transformer des données issues de capteurs en données pertinentes pour décrire le geste, notamment au travers de filtrage des données ou par le calcul de variables latentes (par exemple la vitesse et l'accélération déduite de la position). La deuxième étape d'*utilisation du geste* cherche à relier ces paramètres gestuels à des paramètres de contrôle "haut-niveau" c'est-à-dire mettant en jeu des relations complexes d'un point de vue signal et qui sont perceptivement intéressantes. La troisième et la quatrième étapes consistent à définir ces paramètres "haut-niveau", éventuellement au travers de contrôles intermédiaires qui sont définis d'un point de vue "signal" et qui peuvent aider à définir les relations entre paramètres "haut-niveau" et paramètres de synthèse. Par exemple,

si les contrôles intermédiaires sont des descripteurs de timbre comme la rugosité et le centre de gravité spectral, l'étape trois consistera à définir le lien entre des paramètres haut niveau et ces descripteurs, l'étape 4 consistera trouver une façon de faire évoluer ces descripteurs en agissant sur les paramètres de synthèse.

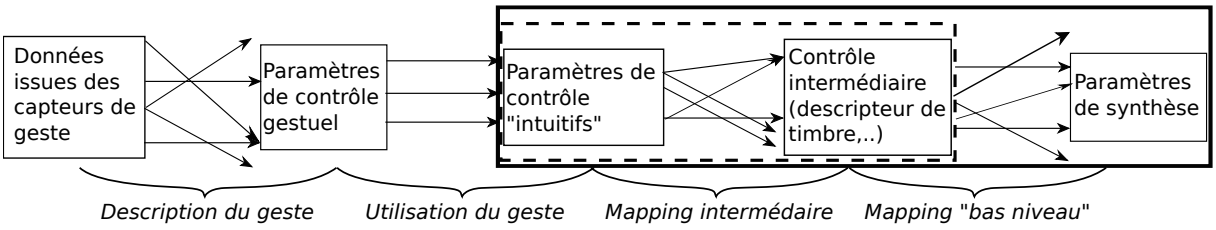


Figure 2.1 – Différentes étapes de mapping entre les contrôles gestuels et les paramètres de synthèse. Notre étude se focalise sur le lien entre paramètres "intuitifs" et paramètres de synthèse (encadré en noir). En pointillé l'étape de mapping qui est dépendante de la stratégie adoptée.

Notre travail se situe donc au niveau de la troisième et de la quatrième étape, c'est-à-dire dans la création de contrôles intuitifs, le contrôle gestuel ne sera pas directement abordé même si on sera forcément confronté à l'utilisation d'interfaces homme-machine qui sont classiquement sur les ordinateurs (clavier, souris). Les travaux qui concernent le contrôle gestuel ont néanmoins beaucoup d'intérêt pour notre problématique puisqu'ils incluent nécessairement l'étape qui nous intéresse. En outre, la synthèse étant principalement utilisée en musique, des avancés notables ont eu lieu avec la création de nouveaux instruments de musique (lutheries "numériques") qui utilisent presque toujours des dispositifs de contrôle gestuel avancés.

Nous verrons ensuite que de nouvelles approches moins directement liées à la musique ont vu le jour, soit en utilisant des méthodes d'apprentissage soit en s'appuyant sur des études perceptives.

Le type d'attributs perçus que l'on veut pouvoir contrôler sont responsables d'une grande part des choix qui sont faits pour aboutir à la stratégie de contrôle. Un intérêt particulier sera donc porté sur la distinction entre les choix qui sont spécifiques à une problématique donnée et ceux qui sont plus généralisables.

Le contrôle de la synthèse a souvent besoin de vocabulaire pour définir les paramètres accessibles aux utilisateurs, notamment dans le cas où on ne peut pas réduire le son à sa cause et c'est un aspect qui a beaucoup été étudié par les compositeurs de musiques électroacoustiques. Historiquement, ce domaine est très lié à la synthèse sonore et a beaucoup oeuvré pour en améliorer le contrôle. Enfin, les musiques électroacoustiques utilisent une très large variété de sons ce qui lui donne une certaine légitimité pour ce qui est d'aborder le problème du contrôle de la synthèse dans un cadre suffisamment général. Ce chapitre se termine donc par un aperçu rapide de travaux menés autour des musiques électroacoustiques et qui sont d'intérêt pour notre problématique. Avant toute chose, il s'agit de préciser ce qu'on entend par contrôle intuitif, c'est donc l'objet de la section qui suit.

2.1 Contrôle "intuitif"

Les interfaces graphiques de synthétiseurs "logiciels" et les sérigraphies sur les synthétiseurs "matériels" se limitent souvent à retranscrire le nom des paramètres mathématiques qui sont contrôlés. Quand il ne s'agit pas d'essayer de reproduire des sons d'instruments existants, il semble très difficile de décrire les sons obtenus et il est souvent très difficile d'organiser les réglages proposés (presets). Dans le logiciel de synthèse Massive [1], les presets sont aussi bien organisés par timbre décrit avec les adjectifs comme "Bright", "dark", "warm", "cold", "fat", "thin", que par méthode de synthèse et par genre musical. On voit donc la volonté qu'il y a, de permettre à l'utilisateur de

1. de la société Native Instrument http://www.native-instruments.com

trouver facilement les sonorités qu'il recherche même si les solutions apportées sont mises en place "à la main" par les musiciens chargés de créer les presets des synthétiseurs.

Dans la littérature on trouve des termes comme "contrôle haut-niveau" [Agon *et al.*, 00] ou "contrôle perceptif" [Arfib *et al.*, 02] pour désigner des approches du contrôle qui visent la facilité d'utilisation. [Wessel et Wright, 02] parlent de contrôle "intime" et même s'ils se situent plus dans la problématique du contrôle gestuel, on comprend bien l'idée de liens subtils entre l'intention du musicien et le son produit que véhicule cette expression. L'aspect qui nous semble être le plus important dans le contrôle est la facilité à utiliser un synthétiseur, on s'intéresse donc ici plus particulièrement au contrôle "intuitif" [Vertegaal et Bonis, 94], pour lequel on souhaite les caractéristiques suivantes :

- Le résultat sonore obtenu après manipulation d'un paramètre peut être facilement anticipé par l'utilisateur
- Aucune connaissance *a priori* de la méthode de synthèse n'est nécessaire pour prévoir le son produit à la suite d'une manipulation
- Les paramètres de contrôle sont adaptés aux besoins de l'utilisateur
- La méthode produit des sons de façon reproductible

Le contrôle intuitif n'implique pas nécessairement des transformations simples du signal ni des mapping "one to one". Au contraire, il pourra être nécessaire de réaliser des relations non-linéaires et mettant en jeu un grand nombre de paramètres du signal pour donner à l'utilisateur un contrôle très simple qui se définit en un mot. Par exemple, les systèmes de ralentissement sans modification de la hauteur, dont le contrôle donné à l'utilisateur se résume à "plus rapide" ou "plus lent", impliquent au niveau du signal, de mettre en œuvre des modifications complexes, surtout si on souhaite un rendu sonore de qualité et si on veut permettre d'agir sur une large plage de valeurs.
Un contrôle intuitif doit avoir une définition claire et sans ambiguïté et à ce titre, l'utilisation d'un mot ou de deux adjectifs antagonistes risque de ne pas répondre à cette contrainte. D'ailleurs, la description de sons avec des mots a été étudiée au travers de la méthode du différentiel sémantique (les sons sont évalués sur des échelles comme "clair-sombre" ou "agréable-désagréable"). Il ressort d'études perceptives utilisant cette méthode qu'il existe parfois des problèmes avec les définitions qui peuvent être interprétées différemment par les sujets [Al-Hindawe, 96].
Le contrôle "intuitif" pose donc clairement le problème de comment définir les caractéristiques sonores (qui souvent évoluent dans le temps) sans ambiguïté et sans risque que plusieurs utilisateurs en fassent des interprétations (au sens de traduction) différentes. Les définitions purement verbales risquent donc dans certains cas, de ne pas être très adaptées et d'autres solutions qui se basent sur des métaphores, des représentations graphiques et même exemples sonores, doivent être envisagées avec dans tous les cas un fort questionnement sur ce qui est finalement perçu par les utilisateurs et les auditeurs en général.

2.2 Contrôle gestuel et synthèse de sons musicaux

Les instruments acoustiques forment, avec l'instrumentiste, un ensemble quasiment indissociable dans la production sonore [Hunt *et al.*, 00]. En effet les relations entre gestes et sons produits sont très complexes, le geste doit souvent être très précis pour reproduire un son donné. En outre, les gestes du musicien, comme ceux réalisés avec un archet, font varier simultanément un grand nombre de paramètres acoustiques. Dans le cas d'instruments "numériques", tous les éléments cités précédemment sont séparables et interchangeables. Par exemple, on peut utiliser des dispositifs de capture de geste très différents pour contrôler un seul et même paramètre de synthèse. En revanche, pour concevoir des instruments qui aient un intérêt musical, il s'agit de créer des relations subtiles entre gestes et son produit. Le mot clé qui résume assez bien la problématique du contrôle gestuel (musical) est l'**expressivité**. En effet, l'enjeu est de donner à l'interprète des possibilités de contrôles suffisantes pour lui permettre de transmettre son intention musicale (cette dernière notion est très importante quand on s'intéresse à l'interprétation [Barthet, 08]).

Ainsi [Choi *et al.*, 95] proposent un dispositif qui permet de se déplacer parmi des paramètres de synthèse projetés dans un espace à trois dimensions (à partir de différents types de contrôleurs gestuels). Appliqué à un système de synthèse vocal, différentes configurations correspondant à des voyelles fournissent des points de repère dans l'espace, les positions intermédiaires étant des interpolations entre ces points de repère. De même, [Mulder *et al.*, 97] ont mis en œuvre un dispositif qui permet de "sculpter" le son en reliant le mouvement des mains à des caractéristiques spectrales et temporelles des sons. Il est intéressant de noter que les auteurs insistent sur l'importance de trouver au niveau des gestes, des caractéristiques qui soient pertinentes et "intuitives" ce qu'ils traduisent en considérant la forme et l'orientation des mains. Certains travaux se sont donc attachés à concevoir des dispositifs qui vont permettre d'obtenir une palette de sons très différents sans donner de définition explicite aux paramètres de contrôle [Valle et Lombardo, 03]. Par exemple, le dispositif pourra être constitué d'un curseur ne comportant pas de label et qui permet de modifier simultanément un très grand nombre de caractéristiques du signal.
D'ailleurs dans ce contexte, il semble que la conception du dispositif fasse partie intégrante du processus de création, les contraintes imposées par le dispositif étant une partie importante de qui a été crée [Gobin *et al.*, 04]. Le contrôle gestuel a également été abordé d'un point de vue purement artistique, dans le but de créer des situations qui offrent aux auditeurs des expériences nouvelles du sonore, à travers des dispositifs interactifs. Par exemple, [Paine, 04] utilise la reconnaissance de geste par caméra vidéo, pour créer des installations où le mouvement de l'auditeur contrôle différentes transformations sonores avec comme but principal, l'engagement (au sens corporel) de l'auditeur.

Le problème du contrôle gestuel est intimement lié aux notions de mouvement et de sémiotique. Le geste est un mouvement du corps, mouvement qui, dans les instruments, se traduit par une modification du son, la sémiotique se trouvant donc dans le lien entre geste et son produit. Pour [Winkler, 95], la création de nouvelles lutheries correspond à "rendre le mouvement musical". Beaucoup de travail a été effectué dans ce domaine [Bowler *et al.*, 90], [Wanderley et Battier, 00], [Wanderley, 02], [Wanderley et Depalle, 04] avec notamment une volonté de formalisation du problème.
Généralement le mapping est défini entre N paramètres de contrôles et M paramètres de synthèse. On parle de mapping "one-to-one" quand un seul paramètre de contrôle est lié à un seul paramètre de synthèse. De même on a les mapping "one-to-many" et "many-to-one" pour les cas ou plusieurs paramètres (de contrôle ou de synthèse) sont reliés à un seul (d'après [Hunt et Wanderley, 02]). Enfin une distinction est faite entre les mapping "explicites" pour lesquels une relation algébrique permet de relier les deux types de paramètres, et les mapping "implicites" pour lesquels la relation est de type "boite noire" notamment via l'utilisation de systèmes d'apprentissage (réseau de neurones, machine à vecteur de support, *etc.*).

Dans le but de pouvoir tester facilement différentes relations entre paramètres de contrôle et paramètres de synthèse, certains chercheurs comme [Bevilacqua *et al.*, 05], [Steiner, 06] ou [Van Nort et Wanderley, 06], [Malloch *et al.*, 08] ont même proposé des "boîtes à outils" qui fournissent des mises en œuvres de différentes méthodes d'apprentissage et de calcul matriciel dédiées au mapping.
En résumé, le contrôle gestuel est un vaste domaine allant de l'analyse fine des gestes de musiciens aux installations artistiques interactives, dont on retiendra l'importance de la notion de mouvement ou au moins d'évolution temporelle. Dans plusieurs de ces travaux (notamment [Choi *et al.*, 95]) on retrouve le terme "intuitif", notion qui semble avoir de manière générale, une grande importance en contrôle.

2.3 Approches par algorithmes évolutionnistes

Les algorithmes génétiques sont des méthodes de calcul numérique qui se basent sur un ensemble de solutions possibles et qui doivent leur nom au fait qu'ils s'inspirent de la théorie de l'évolution pour

améliorer ces ensembles de solutions à chaque itération. En synthèse sonore, ce type d'algorithme a par exemple été utilisé pour résoudre des problèmes d'optimisations pour régler les paramètres de modèles de synthèse par modulation de fréquence [Horner *et al.*, 93]. Ce type d'algorithme dans leur version dite "interactive" permet à l'utilisateur d'ajouter de l'information (perceptive) au système au cours de son exécution. On trouve donc de très nombreuses applications (*Cf.* [Takagi, 01] pour un aperçu des applications possibles dans différents domaines) notamment pour réaliser des mapping complexes intégrant des relations non-linéaires entre paramètres de contrôle et paramètres de synthèse dont le nombre peut être important.

[Takala *et al.*, 93] ont utilisé un algorithme génétique pour permettre aux utilisateurs de naviguer entre différents sons obtenus par synthèse additive et représentés par des "arbres de timbre", chaque nœud de ces arbres correspondant à des opérations arithmétiques. L'algorithme fournit des sons nouveaux qui sont évalués par l'utilisateur et dont le jugement est utilisé pour les futures évolutions, permettant ainsi d'obtenir des sons complexes adaptés aux besoins. De plus, des modules de synthèse d'image sont connectés aux modules de synthèse sonore et utilisent une approche évolutionniste similaire pour contrôler la dynamique des animations.

L'utilisation d'algorithmes génétiques en synthèse a souvent pour objectif de fournir un mapping qui soit pertinent d'un point de vue perceptif. [Johnson, 99] donne l'exemple d'un musicien à qui l'on demanderait de jouer un son mélancolique pour qui il s'agit d'une requête tout à fait légitime et facilement réalisable. En revanche au niveau informatique, il s'agit d'une requête difficilement réalisable ce qui pousse à chercher des méthodes pouvant y répondre. L'approche évolutionniste consiste à ajouter au corpus de sons, un son qui a été jugé mélancolique afin que les sons descendants héritent de ce caractère. Ainsi, [Johnson et Gounaropoulos, 06] ont proposé une interface de contrôle basé sur des descriptions sémantiques du timbre en se basant sur le vocabulaire utilisé par les musiciens pour décrire le timbre (chaud, brillant, métallique,..). Ainsi leur système utilise à la fois un dispositif d'apprentissage par réseau de neurones pour proposer une évaluation d'un son qu'on lui donne en entrée à partir d'une banque de sons annotés. Ensuite, un algorithme génétique permet d'y apporter les modifications souhaitées toujours sur la base des adjectifs/adverbes décrivant le timbre.

[Dahlstedt, 01] a proposé un système de mapping automatique qui s'intéresse uniquement au contrôle dans le sens où il ne manipule que des paramètres qui seront ensuite transmis à un synthétiseur via MIDI. Chaque son est évalué par l'utilisateur et le système propose de nouveaux sons en fonction de ses réponses sans évaluation possible du signal audio par l'algorithme. De même [Mandelis, 03] fait évoluer des paramètres de contrôle qui sont transmis à des synthétiseurs externes avec pour objectif premier d'obtenir des sons "nouveaux".

Dans une logique similaire mais avec des objectifs différents, [Garcia, 01, Garcia, 02] a proposé une méthode pour concevoir automatiquement des algorithmes de synthèse sonore constitués de modules simples comme des oscillateurs, et des opérateurs mathématiques élémentaires. La méthode vise à trouver la "meilleure formule" pour reproduire un son cible en minimisant la distance entre le son généré par l'algorithme et le son cible (plusieurs fonctions coût sont testées). Deux algorithmes génétiques sont donc nécessaires puisque chaque "formule" utilise différents paramètres en entrée. Même si cette approche n'a pas directement l'objectif d'aller vers un contrôle intuitif, on voit aisément l'intérêt de la méthode dans une version où seraient pris en compte des critères perceptifs.

Avec une approche dite "interactive", le programme ESSynth [Manzolli *et al.*, 01] (et développé dans [Caetano *et al.*, 05]) cherche à créer des sons nouveaux à partir de sons cibles (définis par l'utilisateur). L'approche est donc une approche similaire à celle de l'analyse-synthèse [Risset et Wessel, 99] et sur la base d'une définition algorithmique, il s'agit en fait de créer des variantes de l'ensemble de sons initialement définis. Cette approche est très intéressante car les paramètres de contrôle sont en partie définis par les sons que l'utilisateur choisit en entrée de l'algorithme et qui influencent beaucoup les timbres qui seront obtenus. En revanche, les modifications qui peuvent être apportées au signal n'ont pas de définitions claires.

Egalement basé sur l'apport d'informations par l'utilisateur, McDermott et ses collaborateurs ont mis en place un dispositif qui permet de naviguer parmi les paramètres d'un synthétiseur à l'aide d'un seul curseur linéaire [McDermott *et al.*, 05, McDermott *et al.*, 07, McDermott *et al.*, 08].

Le principe est de donner accès à l'utilisateur à trois sons situés aux deux extrémités et au centre du curseur puis de générer tous les intermédiaires. La position choisie par l'utilisateur est ensuite utilisée comme entrée pour générer 2 nouveaux sons qui prennent en compte les évolutions successives. Les auteurs ont montré grâce à des tests d'écoute que cette approche permet de reproduire un son cible relativement facilement et en particulier dans le cas de non-experts par rapport à la même tâche effectuée en donnant accès à l'ensemble des paramètres de synthèse. Le contrôle est donc, dans une certaine mesure, défini à partir de sons, et l'utilisateur procède à un "morphing" entre ces sons de base pour naviguer dans l'espace des paramètres de contrôle.

Même si ces approches sont incontestablement efficaces pour naviguer parmi des espaces de sons complexes, elles ont le désavantage de fonctionner en "boîte noire" et de ne pas permettre de mieux comprendre les liens entre signal et perception. Il est dans tous les cas nécessaire de définir dans quelles plages de valeurs chaque paramètre de synthèse va varier ainsi que la loi de contrôle pour obtenir une variation perçue qui soit linéaire.

Dans certains contextes, il semble également que ces approches ont une complexité de mise œuvre qui les rendent difficilement utilisables (temps d'apprentissage, complexité algorithmique,...).

2.4 Timbre et contrôle

La notion de timbre est, suivant le point de vue qu'on adopte, très liée à la notion d'invariant (notion discuté dans l'introduction) dans le sens où elle décrit ce qui caractérise le son produit par un objet ou un ensemble d'objets [2]. Cette notion a motivé un très grand nombre d'études visant à donner au timbre, une définition objective de façon similaire à celle qui existent pour la hauteur et l'intensité notamment en utilisant la synthèse sonore. On peut d'ailleurs considérer que les paramètres d'entrée d'un synthétiseur agissent nécessairement sur le timbre. Toutes ces remarques montrent que le timbre est donc une notion incontournable quand on s'intéresse au contrôle et il n'est pas étonnant que beaucoup d'exemples de recherches sur le contrôle s'inspirent de travaux sur le timbre.

2.4.1 Contrôle et espaces de timbre

La synthèse additive offre des possibilités infinies pour générer tout type de sons mais a très vite posée le problème du contrôle [Schindler, 84], dans le sens où un grand nombre de composantes spectrales qui varient en fréquence et en amplitude au cours du temps doivent être gérées simultanément. Un des premiers à avoir l'idée d'utiliser les résultats obtenus dans des études sur le timbre pour contrôler des processus de synthèse est certainement D.Wessel. Dans [Wessel, 79], l'objectif était de contrôler un système de synthèse additive (à 25 composantes) à l'aide d'un nombre réduit de paramètres. En se basant sur un test de dissemblance et des analyses multidimensionnelles correspondantes, il a donc obtenu deux dimensions pour naviguer entre les différents timbres, la première étant liée à la répartition spectrale d'énergie, la seconde à la "nature de l'attaque".
Les espaces de timbre comme celui proposé par [McAdams *et al.*, 95] ont été mis en place dans le cadre de recherches en psychoacoustique, ainsi, [Gounaropoulos et Johnson, 06] affirment que ce ne sont pas des représentations adaptées pour la synthèse même si elles peuvent être très pertinentes pour des tâches de catégorisation/classification. En effet, la réduction d'un espace de timbre à un petit nombre de dimensions implique une perte d'information qui ne permet pas directement de définir des valeurs pour les paramètres de synthèse.
En outre, il existe d'autres modèles de timbre qui ont été crées pour des problématiques liées à la synthèse sonore [Desainte Catherine et Marchand, 99] [Jensen, 99] et qui sont donc bien adaptés à la mise en place d'une stratégie de contrôle.
Le problème qui se pose avec les espaces de timbre réduits peut se résoudre lorsque l'on fait appel

2. On épargne ici le lecteur des définitions fournies par les agences de normalisation [ANSI, 60, AFNOR, 77] dont beaucoup s'accordent à dire que leur utilité est très réduite.

à des méthodes d'apprentissage qui à partir de larges bases de données, rajoutent l'information manquante.
En se basant sur le calcul temps réel de descripteurs de timbre, [Jehan et Schoner, 01b] [Jehan et Schoner, 01a] contrôlent un modèle de signal de type sinusoïde+bruit. Leur système permet par exemple de contrôler des sons de saxophone à l'aide d'un micro et de sa propre voix. Pour ce faire, ils utilisent un système d'apprentissage supervisé qui réalise un mapping non linéaire entre le vecteur de descripteurs de timbre et le modèle additif.

Egalement dans la logique du contrôle "intuitif", [Groux et Verschure, 08] proposent de contrôler un moteur de synthèse additive à l'aide d'une analyse en composantes principales (ACP) de l'évolution des composantes spectrales (l'ACP prend donc en entrée des fonctions qui dépendent du temps) et un apprentissage basé sur des machines à vecteur de support. Dans un premier temps, cette méthode est appliquée au contrôle de la sonie et de la hauteur (tout en préservant le timbre) pour un saxophone solo mais est suffisamment générale pour être appliquée à d'autres paramètres perceptifs.

De la même manière, [Hoffman et Cook, 06, Hoffman et Cook, 07] ont proposé une généralisation de ces approches sous le nom de "Feature based synthesis" (FBS). La FBS consiste à résoudre le problème inverse suivant : Trouver une combinaison u' des paramètres du modèle de synthèse tel que $F(u') = v$ où F est un estimateur de la caractéristique à contrôler (descripteur de timbre) et v une valeur cible de cette caractéristique. En pratique le problème est souvent mal posé et les applications ne sont pas toujours très convaincantes même si cette approche reste très prometteuse.

Il reste néanmoins un problème ouvert quant à l'information qui n'est pas décrite par les descripteurs de timbre, mais qu'il est nécessaire d'ajouter au système pour pouvoir synthétiser des sons. Par exemple, il peut être nécessaire de définir un contenu spectral qui servira de base au système de transformation ou de générer des textures dont certaines caractéristiques évoluent aléatoirement au cours du temps. La question est de savoir dans quelles mesures des aspects que l'on définit indépendamment du contrôle visé ont une influence sur ce contrôle. Même dans les cas d'outils qui se limitent à de la transformation, il est délicat d'évaluer l'influence des caractéristiques du signal de départ sur les transformations qui sont effectuées.

2.4.2 Timbre et contrôle intuitif

Un des exemples les plus parlants et des plus accessibles de contrôle intuitif est celui développé dans [Gounaropoulos et Johnson, 06] où les auteurs ont proposé un contrôle du timbre défini à partir d'adverbes et d'adjectifs qu'utilisent naturellement les musiciens pour décrire le son (clair, chaud, dur, métallique,...). L'approche se base sur un test d'écoute où les sujets ont jugé des sons sur des échelles définies par ces adjectifs ainsi qu'un système d'apprentissage (réseau de neurones). En prenant un son en entrée, l'algorithme fournit une estimation du jugement pour chaque échelle et permet de modifier le son en agissant sur ces mêmes échelles. Cette approche est très puissante, mais demande encore à être développée pour obtenir des résultats sonores convaincants.
Un travail similaire a été mené par [Howard *et al.*, 07] où les auteurs se sont concentrés sur la partie qui concerne les tests perceptifs, l'implémentation étant réalisée par le contrôle de descripteurs de timbre dans un modèle de synthèse additive. Dans le cas spécifique des sons d'impact et plus précisément du matériau perçu (bois, métal, verre) [Aramaki *et al.*, 11] ont proposé une approche qui a permis d'aboutir à un contrôle offrant à l'utilisateur, la possibilité de passer continûment d'un matériau à un autre à l'aide d'un unique paramètre de contrôle. En résumé, la méthode se base sur un test d'écoute durant lequel les sujets devaient choisir entre 3 catégories de matériaux lors de l'écoute de sons obtenus par morphing entre des re-synthèses d'enregistrements d'impact sur différents matériaux. L'analyse acoustique a ensuite permis d'identifier plusieurs descripteurs permettant de discriminer les sons suivant le matériau perçu ainsi qu'une stratégie pour contrôler ces descripteurs. Pour ce problème comme pour d'autres problématiques qui concernent les sons d'environnement, la physique permet souvent de trouver des caractéristiques des signaux qui sont importantes d'un point de vue perceptif. Par exemple dans le cas de structures résonnantes, l'analyse modale met en évidence des relations entre la taille de l'objet et la fréquence fondamentale, ainsi

qu'entre l'excitation (notamment le point d'excitation) et la répartition spectrale de l'énergie. Enfin pour les sons d'environnement au sens de [Gaver, 93b], il existe différentes approches qui fournissent des sons de synthèse adaptés à une description sémantique en lien avec des applications multimédias (jeu vidéos, réalité virtuelle). En particulier, [Verron *et al.*, 10] ont mis en œuvre une stratégie de contrôle qui permet par exemple à l'utilisateur, de contrôler l'intensité ou le ruissellement pour des sons de pluie, la froidure pour des sons de vents, *etc.*. Un contrôle original a été mis en place qui permet de modifier l'élargissement de sources spatialisées. Des tests d'écoute ont été nécessaires pour déterminer la capacité des auditeurs à discriminer différentes largeurs de sources et donc le nombres de sources virtuelles nécessaires.
L'utilisation de test d'écoute est globalement dédiée à des aspects très précis du contrôle mais semble dans beaucoup de cas, incontournable.

Sans avoir recours à des tests d'écoute pour savoir ce qui est perçu par les auditeurs, les compositeurs de musiques électroacoustiques sont en permanence dans le questionnement sur le sens véhiculé par les sons qu'ils utilisent et sur la manière de traduire des idées musicales en combinaisons et transformations sonores. Les recherches qui ont eu lieu autour de ces musiques sont donc incontournables quand on s'intéresse au contrôle de la synthèse, c'est pourquoi, la section qui suit donne des bases pour mieux comprendre les liens et les apports potentielles à notre problématique.

2.5 Apport des musiques électroacoustiques

Au début des années 50, apparaissent presque simultanément deux nouveaux genres de musique qui vont bouleverser la manière d'envisager la composition : la musique concrète et la musique électronique. La musique concrète s'élabore directement en manipulant de la matière sonore (enregistrée) en opposition aux musiques composées "sur papier" et qui deviennent des sons seulement quand elles sont jouées par des instrumentistes. La musique électronique quant à elle, est une musique composée uniquement avec des sons de synthèse et dans un premier temps uniquement obtenus par synthèse additive avec des oscillateurs analogiques. Ces deux musiques sont les premières, qui pour exister, sont tributaires de toute une chaîne électroacoustique, dont le haut-parleur est le dernier maillon. De plus, elles opèrent directement sur le son et s'affranchissent de ce qui peut être considéré comme une barrière entre le compositeur et le public : l'interprète. Même si ces deux musiques ont souvent été opposées, leurs évolutions et l'apparition d'autres musiques (notamment la musique pour bande) font qu'elles peuvent raisonnablement être regroupées sous le terme de musiques électroacoustiques [Cross, 68].

2.5.1 Musique électronique

Il faut tout d'abord préciser qu'on parle ici de musique électronique telle qu'elle était pratiquée à ses origines au sein de la WDR [3]. Une des idées de départ était de constater qu'il est possible de reproduire par synthèse additive, n'importe quel son ce qui permet de ne plus du tout avoir recours à des instruments de musique. En pratique, l'utilisation des synthétiseurs et de tous les dispositifs de traitement de son et d'enregistrements qu'il était nécessaire de manipuler a mis en évidence les limites techniques à cette idée de départ. Les compositeurs se sont trouvés devant la nécessité d'acquérir des compétences en traitement du signal, en électronique, la manipulation des appareils étant souvent très complexe. C'est pour ces raisons que les compositeurs ont très vite été demandeurs de réponses techniques à ces problèmes de complexité de mise en œuvre qui dans certains cas peuvent être un frein à la créativité. En ce sens ce sont les premiers à poser la question du contrôle de la synthèse.

3. WestDeutscher Rundfunk, la radio de l'Allemagne de l'Ouest à Cologne. On parlait à cette époque d'*Elektronische music* [Eimert, 57].

L'utilisation de la synthèse et le fort intérêt pour la recherche que portaient ces pionniers [4] ont amené des réflexions très pertinentes sur ce que nous appelons ici les invariants [Meyer-Eppler, 58].

2.5.2 Musique concrète

De même que pour la musique électronique, la musique concrète s'est dès le début inscrite dans une dynamique d'expérimentation et de recherche, notamment à travers les travaux menés autour de P. Schaeffer au GRM [5] qui sont développés dans [Schaeffer, 66].
Comme on l'a mentionné plus haut, musique concrète et musique électronique se distinguent principalement sur le type de sons utilisés à la base des compositions. Cette distinction est loin d'être anodine puisqu'elle met en avant le problème du sens véhiculé par ces sons et de l'interaction entre ce sens et l'intention musicale du compositeur et ce, principalement quand la source peut être reconnue. Cette distinction met également en avant un problème de complexité de ce qui est donné à entendre. Dans [Dack, 02], l'auteur écrit :
"Schaeffer's decision to use recorded sounds was based on his realization that such sounds were often rich in harmonic and dynamic behaviors and thus had the most potential for his project of musical research"
Que l'on peut traduire par : *La décision de Schaeffer d'utiliser des sons enregistrés était due à une prise de conscience du fait que ces sons étaient souvent riches en termes de caractéristiques spectrales et de dynamiques et par conséquent, étaient les plus prometteurs pour son projet de recherche musicale"*
On trouve en perception sonore une démarche similaire appelée approche "écologique" issue des travaux (dans le domaine de la vision) de Gibson [Gibson, 78] et qui évite les stimuli synthétiques simples (sons pures, bouffées de bruit,...) dans les études perceptives car ils ne correspondent pas à ce qui peut être entendu dans la vie quotidienne. L'approche que nous souhaitons suivre dans cette thèse se réclame de ce souci de "validité écologique".
C'est avec l'expérience du sillon fermé [6] que Schaeffer s'est rendu compte de la possibilité de modifier son écoute et de faire passer au second plan ce qui serait en d'autres circonstances le but premier de l'audition : reconnaître la source. C'est dans ce cadre que la notion d'écoute "acousmatique" a été introduite, c'est-à-dire l'écoute des caractéristiques intrinsèques des sons sans se préoccuper de la source qui a produit le son. Les travaux de Schaeffer ont montré que ce type d'écoute dépend de l'intention de l'auditeur autant que du contexte et qu'il évolue au cours du temps (notamment quand un son est répété plusieurs fois de suite). Il est très intéressant de constater qu'une notion très proche se retrouve dans le domaine de la perception et de la synthèse avec l'écoute "musicale" définie par Gaver [Gaver, 93a]. Pour notre problème, les différents types d'écoutes que va adopter un auditeur en fonction du contexte et de sa volonté, sont très importants et devront être prises en compte aussi bien dans les tests d'écoute que lors de la conception même du synthétiseur.

Un autre point central des travaux menés autour de la musique concrète, a été la mise en place de tout un vocabulaire pour décrire les sons et en particulier ceux dont l'origine n'est pas identifiable. Même si les matériaux sonores utilisés par les compositeurs sont des enregistrements qui peuvent être décrits en termes d'objets et d'interactions entre objets qui produisent du son, trouver une façon de décrire les sons est devenu un problème incontournable. Comme on l'a mentionné précédemment, le sens véhiculé par les sons peuvent interférer avec l'intention musicale ce qui a

4. Werner Meyer-Eppler qui fut avec Robert Beyer et Herbert Eimert l'un des pionniers de cette musique, était mathématicien, physicien et fut directeur de l'institut de phonétique de l'université de Bonn.

5. Groupe de Recherche Musicale fondé par P. Schaeffer au sein de la Radio Television Française (RTF) et qui est actuellement rattaché à l'Institut National de l'Audiovisuel (INA).

6. En mettant en boucle par inadvertance une petite partie d'un enregistrement sur disque micro-sillon, Schaeffer s'est rendu compte que l'extrait sonore prenait un intérêt musical particulier ce qui a d'ailleurs été à la base de sa composition "Bidule en Ut".

poussé nombre de compositeurs à utiliser des techniques d'enregistrement et des transformations permettant de rendre impossible toute tentative d'identification d'une quelconque source. De plus, l'écoute acousmatique qui peut être adoptée rend totalement secondaires, les informations qui concernent la source en mettant en avant les qualités "musicales" des objets sonores [7], qualités qui doivent nécessairement être décrites et organisées pour aider le compositeur à s'y retrouver, dans l'infinité des choix qu'il peut faire. Schaeffer a donc introduit les critères de morphologie et de typologie, qui, suivant que le but est de décrire ou de produire de la musique, permettent de définir n'importe quel son. Même si ces critères seront développés plus loin (dans les chapitre 3 et 4), on peut d'ors et déjà préciser qu'ils se basent sur les caractéristiques acoustiques globales des sons et prennent en compte ce qui est perçu et notamment le rôle de la mémoire dans la perception.

Malgré ces efforts d'organisation et de formalisation du matériau sonore, les musiques électroacoustiques semblent toujours poser problème pour ce qui est de les décrire [Stroppa, 84][Bossis, 06]. En particulier, ces musiques résistent à la mise en place d'un système commun de notation. Ces problèmes sont directement posés aux chercheurs travaillant en analyse des sons et en perception [Zattra, 05]. Les compositeurs de musiques électroacoustiques sont à l'origine de nombreux travaux menés en informatique musicale, et ont très souvent été les premiers (voir les seuls) utilisateurs des résultats obtenus dans ce domaine. La conception d'un synthétiseur doit prendre en compte les utilisateurs potentiels qui dans notre cas risquent beaucoup d'être issus des musiques électroacoustiques.

Des réflexions et des travaux qui ont été présentés dans cette section, découle un aspects originaux du travail qui suit : l'utilisation de sons abstraits. Ces sons dont la source ou plus généralement l'origine physique n'est pas facilement identifiable ont toutes les chances de favoriser une écoute acousmatique et donc limiter les problèmes de sens (comme le suggère [Fastl, 01]). La complexité de leurs caractéristiques spectro-temporelles répond à l'exigence d'une approche écologique de la perception et promettent des résultats riches en termes d'analyse du signal.

2.6 Conclusion

Comme discuté dans l'introduction, le contrôle est un enjeu essentiel pour la synthèse sonore. La diversité des approches décrites dans ce chapitre met en évidence la grande complexité du problème et sa dépendance aux objectifs de départ. La mise en place d'une méthode générale qui pourrait s'appliquer aussi bien à des dimensions reliées à la physique qu'aux aspects des sons qui sont de l'ordre de l'évocation, est donc un problème complexe et jusqu'à maintenant, relativement peu de réponses ont été apportées. On a mis en évidence deux problèmes majeurs, l'évaluation perceptive de sons et la recherche de corrélats acoustiques aux dimensions perceptives. L'évaluation perceptive pose de nombreux problèmes méthodologiques et notamment celui du choix des sons à utiliser ainsi que comme on l'a vu, celui de donner une définition claire aux futurs paramètres de contrôle. Enfin il est évident que la mise en œuvre dans un outil de synthèse, des caractéristiques acoustiques identifiées est également un problème complexe dont on a vu qu'il existe des solutions suffisamment générales pour ne pas être dépendant d'une méthode de synthèse particulière. Le chapitre suivant est une première approche qui se concentre sur les deux problèmes principaux.

7. Par convention, on nomme "corps sonore" un objet matériel qui produit du son et "objet sonore" le son une fois enregistré et qui est utilisé pour faire de la musique. Ce dernier n'a pas forcément de lien avec un corps sonore.

Chapitre 3

Approche par catégorisation

3.1 Quels stimuli et quel protocole **38**
3.1.1 Quels stimuli ? 38
3.1.2 Quel protocole ? 40
3.2 La catégorisation **41**
3.2.1 Théorie de la catégorisation 41
3.2.2 Catégorisation de sons 41
3.3 Test 1 : Classification libre **42**
3.3.1 Sujets 42
3.3.2 Stimuli 42
3.3.3 Protocole 43
3.3.4 Résultats 43
3.4 Test 2 : Classification contrainte **48**
3.4.1 Sujets 48
3.4.2 Protocole 48
3.4.3 Résultats 48
3.4.4 Discussion 49
3.5 Analyses des signaux **50**
3.5.1 Descripteurs de timbres “classiques” 50
3.5.2 Descripteurs spécifiques 52
3.5.3 Sélection de descripteurs 53
3.5.4 Résultats 54
3.6 Conclusion **57**

Les travaux qui font l'objet de ce chapitre sont également décrits dans [Merer et al., 08][1]

Cette première étude a été menée non seulement dans le but de “défricher” les grandes dimensions perceptives du mouvement évoqué par les sons, mais aussi d'aller le plus loin possible dans l'analyse des sons en lien avec ces dimensions. Le terme mouvement définit une thématique et il ne correspond en aucun cas à un seul paramètre de contrôle dans un synthétiseur. Identifier les grandes dimensions perceptives revient donc à définir les futurs paramètres de contrôle et on verra que dans plusieurs cas, ce problème est loin d'être trivial.
De nombreux travaux (par exemple [Dubois *et al.*, 06, Guastavino, 07]) ont montré l'intérêt de l'approche par catégorisation. N'ayant pas de connaissance *a priori* sur les grands types de mouvements, et partant avec peu d'hypothèses sur leurs corrélats acoustiques, il a paru assez logique d'utiliser ce type de méthodologie. Nous avons donc opté pour une tâche de catégorisation libre où il est demandé aux sujets de regrouper des sons qui partagent une ou plusieurs caractéristiques. Dans notre cas, il s'agissait de regrouper les sons qui évoquaient le même type de mouvement.
Un des principaux avantages réside dans la grande liberté qui est laissée aux sujets, leur permettant d'évaluer les stimuli de façon très générale. Cette grande liberté peut également être problématique car elle conduit indéniablement à une grande variabilité entre les réponses données par différents auditeurs. C'est pourquoi, dans un second temps, nous avons souhaité contraindre le test pour

1. Les stimuli utilisés durant l'expérience sont disponibles sur la page internet de l'article : http://www.sensons.cnrs-mrs.fr/CMMR07_semiotique/

avoir des données obtenues avec une base commune à toutes les personnes interrogées. Nous avons donc construit un protocole en deux phases successives la première étant un test de catégorisation libre où les sujets regroupent les sons par "mouvements identiques", la deuxième étant un test de catégorisation contrainte où les mêmes sons doivent être classés dans des catégories définies par l'analyse des données du premier test.
Nous avons souhaité mener l'analyse des résultats le plus loin possible de façon à avoir une vision globale de la méthode et identifier les verrous que l'on devra lever.

3.1 Quels stimuli et quel protocole

Comme à notre connaissance aucune étude ne s'est posé les questions qui nous intéressent, il est nécessaire de mener une réflexion autour du choix des stimuli et du protocole d'évaluation de ces stimuli. Il faut préciser que ces deux choix sont intimement liés et qu'ils ne peuvent pas être faits séparément. En effet, un protocole établi sur la base d'hypothèses et à l'aide d'un plan d'expérience dictera précisément les stimuli qui devront être utilisés. Inversement, un ensemble de stimuli déterminés orientera les choix, ne serait-ce que par des considérations sur les durées ou les formes, sur des méthodes particulières.
Dans notre cas, il semble intéressant, au moins dans une première approche, de ne pas faire d'hypothèses sur les dimensions perceptives qui sont liées au mouvement. Par ce choix (ou plutôt cette absence de choix), on s'oriente donc vers une sélection de stimuli la plus large possible, susceptible de représenter “tous les sons du monde” et par là même, à un protocole le moins réducteur possible. Ensuite, nous avons le double objectif de concevoir une méthode générale d'évaluation perceptive en vue de contrôler un synthétiseur et celui de mieux caractériser la notion de mouvement.

3.1.1 Quels stimuli ?

La sélection du corpus de stimuli est une étape extrêmement importante pour plusieurs raisons :
- C'est la première étape, elle influence donc toutes les suivantes
- De la composition du corpus dépend la généralisation possible des résultats obtenus et des analyses statistiques qui pourront être effectuées
- Elle introduit un biais en orientant les réponses par rapport à un ensemble restreint de stimuli

Il est souvent difficile de définir et justifier les critères de sélection des sons. Par exemple dans [Lavandier et Defreville, 06] où les auteurs s'intéressent aux ambiances sonores urbaines, les vingt stimuli sont décrits comme recouvrant toutes les conditions "typiques" d'événements pouvant être perçus dans la rue (passage de voiture, de cyclomoteur, discussions,...). On peut alors se demander quels autres critères ont amené à choisir spécifiquement ces séquences parmi plusieurs observations d'un même événement et surtout quelle est l'influence de ces choix sur le reste de l'étude. Autre exemple dans [Tardieu *et al.*, 08], le choix des ambiances (de halls de gare) est basé sur la "qualité des prises de sons", "l'originalité des événements" et la présence d'éléments "attendus" par rapport à une pré-étude sémantique. Dans [McAdams *et al.*, 95] l'étude qui s'intéresse à des sons d'instrument de musique utilise un ensemble de "presets" fourni par un synthétiseur du commerce. Il n'y a donc pas de stratégie précise à suivre ni de méthode pour estimer *a posteriori*, à quel point ces choix ont influencé les résultats de ces études. On peut malgré tout définir un certain nombre de critères qui orientent la sélection de stimuli :

1. La représentativité du corpus
2. La taille du corpus
3. L'homogénéité

Le corpus se doit de représenter aux mieux, les différentes situations et/ou les différentes configurations acoustiques qui découlent de la problématique de départ.

Le nombre d'éléments du corpus est nécessairement limité par la durée de test et toutes les conséquences pratiques qui en découlent.
Le jugement porté par les sujets sur chaque élément du corpus, se fait généralement relativement aux autres éléments du corpus. Il est donc nécessaire d'éviter d'avoir quelques éléments très différents des autres et qui pourraient faire l'objet d'un traitement différent.

Les sons utilisés dans des tests d'écoute sont généralement issus d'une des 4 méthodes suivantes de sélection :

Banques de sons Cette approche consiste à s'appuyer sur des banques de sons de grande taille qui contiennent des sons aux caractéristiques et aux origines très diverses. En utilisant des banques de sons qui ont servi dans d'autres expériences, on aura la possibilité de faire des comparaisons et de prendre en compte d'autres attributs (valence émotionnelle, qualité,..). Il est en revanche souvent difficile de justifier le choix d'une banque de son particulière notamment pour celles qui ont été crée dans un cadre commercial pour des applications particulières (cinéma, radio,...).

Enregistrements spécifiques Une première méthode consiste à enregistrer un certain nombre de sons en fonction d'un plan d'expérience [2] établi par rapport au nombre de paramètres à tester. Lorsque l'on s'intéresse au mouvement, il faudrait par exemple placer un haut-parleur sur un rail et le déplacer à différentes vitesses et sur différentes trajectoires. On présente ainsi aux auditeurs, des enregistrements dont les paramètres qui varient sont limités et contrôlés au mieux. Il s'agit donc de définir à l'avance, les situations qui doivent être enregistrées et présentées durant le test. L'inconvénient de ce type d'approche est que la mise en œuvre du dispositif peut être complexe et coûteuse suivant le dispositifs expérimental qui doit être réalisé.

Enregistrements transformés Il est également possible, en partant d'un son qui contient une information intéressante, de le transformer et identifier les transformations qui altèrent (et celles qui n'altèrent pas) la caractéristique étudiée. Il est donc nécessaire de définir à l'avance, les transformations qui sont susceptibles d'avoir un effet notoire sur l'attribut étudié.

Sons de synthèse Egalement sur la base d'hypothèses, mais qui cette fois concernent en plus les paramètres des signaux qui décrivent les dimensions perceptives, une autre approche consiste à synthétiser des sons dont seules les caractéristiques intéressantes pour l'étude varient. Par exemple, pour ce qui concerne les formes de trajectoire, on cherche une correspondance entre acoustique et lois de modulation d'amplitude pour les différentes trajectoires et on applique ces lois de modulation à un même signal initial. Cette approche nécessite également des hypothèses à tester.

Notre démarche se situe donc au croisement de ces différentes méthodes puisque nous avons choisi d'utiliser des sons qui proviennent de banques de sons (celles des compositeurs de musiques électroacoustiques), qui ont été transformés pour rendre incertaine l'identification d'une source, et auxquels nous avons ajouté des sons de synthèse. Les sons de synthèse reprennent la plupart des considérations physiques faites au chapitre consacré au mouvement (section 1.4.1).

Au delà du simple compromis, et suite aux discussions menées dans la section 2.5, l'utilisation de sons abstraits a été choisie sur la base de deux hypothèses :

- L'information qui nous intéresse (sur le mouvement) risque moins d'entrer en conflit avec d'autres informations véhiculées par les sons, dans le cas de sons abstraits que dans le cas de sons "reconnaissables"

2. Les plans d'expérience font l'objet d'une vaste littérature, voir par exemple [Gauchi, 05] pour un aperçu des méthodes disponibles

– Les sons abstraits sont mieux adaptés à une écoute sur haut-parleur et aux applications qui en découlent.

Pour illustrer la première hypothèse, on peut donner l'exemple d'un son reconnu comme étant le bruit d'une voiture en circulation. On conçoit aisément que les sujets ne vont pas associer ce son à un déplacement sur l'axe vertical, mouvement qui est impossible pour une voiture, alors qu'un son avec des caractéristiques acoustiques similaires mais non identifié peut théoriquement être associé à un mouvement dans n'importe quel direction de l'espace.

La seconde hypothèse concerne également le problème de la reconnaissance de source et le fait que les résultats obtenus sur ce type de sons ont des chances d'être plus facilement applicables notamment en design sonore où il s'agit de véhiculer une information précise sur un support sonore nouveau. Même si on est capable d'utiliser comme sonnerie de téléphone des sons enregistrés les plus divers comme des sons d'animaux et donc de modifier le sens qu'on leur attribue (la signalisation d'un appel téléphonique), il semble que les sons abstraits sont susceptibles de s'adapter plus facilement à des utilisations où ils doivent véhiculer un nouveau sens. Il est certain que les questions soulevées par l'utilisation des sons abstraits sont complexes, on aura largement l'occasion d'y revenir dans la suite de ce document (notamment section 6.1).

3.1.2 Quel protocole ?

Les approches qui se basent sur des hypothèses quant aux liens entre signal et perception ne peuvent s'utiliser que dans les cas où les attributs étudiés sont objectifs et peuvent s'appuyer sur des considérations physiques. Notre volonté est ici d'avoir une méthode qui ne se limite pas à des attributs objectifs ce que nous permettent d'ailleurs d'envisager les sons abstraits. On exclura donc les protocoles qui nécessitent la mise en place de plan d'expérience (par exemple les tests de comparaisons par paire ou de comparaisons à une référence [3]). Les possibilités de protocoles en restent malgré tout relativement larges et la seule méthode qui permette de les comparer est de toutes les mettre en œuvre pour une même problématique et un même corpus de son comme l'ont montré [Parizet *et al.*, 05].

Les premières méthodes envisagées et qui semblent être les moins restrictives sont les méthodes de type "interrogatoire" avec les sujets. Il s'agit simplement de collecter à l'oral et/ou à l'écrit les informations perçues lors de l'écoute de sons. En pratique, ces approches sont basées sur des suppositions théoriques concernant les catégories cognitives et leurs liens avec le langage [Dubois, 00]. Cependant, le traitement des données nécessite des analyses psycho-linguistiques des commentaires qui dépassent largement notre domaine de compétences. La méthode du différentiel sémantique (jugement sur des échelles définies par des adjectifs ou des adverbes) est une alternative potentielle des interrogatoires mais demande de définir des échelles et par conséquent de connaître *a priori* les attributs subjectifs, et ce de façon exhaustive. De plus, le différentiel sémantique demande également d'avoir une expertise en psycho-linguistique. On risque également de se confronter à d'éventuelles incompréhensions entre sujets et expérimentateur ne serait-ce qu'à cause des imperfections du langage que nous utilisons.

C'est pourquoi nous nous sommes tournés vers les tests de catégorisation et plus particulièrement la catégorisation libre. A la différence des tâches de catégorisation libre habituellement utilisées en audio où les sujets doivent regrouper les sons qui leur semblent "similaires", il a été demandé de regrouper les sons en fonction des types de mouvements qu'ils évoquent. Cette méthode doit nous permettre de déterminer un maximum de catégories de mouvement en contraignant le moins possible les sujets. Dans un second temps nous proposons une tâche de catégorisation contrainte, les catégories étant définies non pas à l'aide de labels mais à l'aide de sons typiques. On vérifie ainsi que les résultats obtenus dans la tâche de catégorisation libre sont valables pour l'ensemble des sujets sans avoir recours à des descriptions sémantiques.

3. Ces méthodes sont envisageables sans plan d'expérience en demandant aux sujets d'évaluer si deux sons ont des mouvements "proches" mais il serait alors impossible de remonter aux mouvements correspondants

3.2 La catégorisation

3.2.1 Théorie de la catégorisation

Les recherches qui s'intéressent à la catégorisation et à la perception catégorielles sont très nombreuses et forment à elles seules une branche importante de la psychologie. De ces recherches, ressortent un certain nombre de questions fondamentales ainsi que des concepts y afférent. Il s'agit donc de relever les hypothèses et les problèmes inhérents à une approche par catégorisation pour mieux les prendre en compte dans l'étude qui suit.

Un des premiers principes de la catégorisation semble être de fournir un maximum d'informations en faisant le moins d'efforts (cognitifs) possible [Rosch, 78]. De cette notion d'économie découle la notion de classes d'équivalence [Goldstone et Kersten, 03] (p. 600) qui indique qu'au sein d'une catégorie, des éléments différents peuvent être traités de façon identique.

Comme décris dans [Rogosky et Golstone, 05], la catégorisation se base sur la réduction du problème sous forme de caractéristiques. Dans le domaine de la représentation des objets, les catégories sont souvent distinguées selon trois niveaux d'abstraction [Rosch, 78]. Ces trois niveaux sont inclusifs et les distances inter-catégories sont maximales pour des catégories de même niveau.

Également à travers les travaux de Roch ainsi que ceux de Lakoff, est apparue la notion de prototype et toute une théorie qui en découle. Dans ce cadre, le prototype est un élément qui représente bien une catégorie et qui peut ne pas avoir d'existence concrète. Cet élément sert donc de référence pour déterminer l'appartenance à la catégorie en fonction des propriétés qu'un objet partage avec le prototype.

3.2.2 Catégorisation de sons

[Kawai *et al.*, 04] ont mené une étude sur les sons d'environnement où les sujets devaient catégoriser non pas des sons mais leur description. Pour cela, ils leur fournissaient des cartes sur lesquelles était écrit un mot décrivant le son, la tâche étant réalisée dans différents contextes (laboratoire, différents lieu d'une ville et chez les sujets). Ils obtiennent ainsi trois grands facteurs qui sont un jugement qualitatif, l'activité qui est liée au son et le rôle du son, l'importance relative de ces facteurs étant significativement influencée par le contexte. [Guastavino, 07] s'est intéressée aux catégories de sons d'environnement et à leur description à l'aide d'une tâche de catégorisation libre. Les deux catégories principales identifiées concernent le lien ou l'absence de lien avec l'activité humaine avec un jugement positif ou négatif sur le son.
Egalement dans le cadre des sons d'environnement, [Gygi *et al.*, 07] se sont intéressés à relier les catégories obtenues à des indices acoustiques et ont montré qu'un petit nombre d'indices (harmonicité, quantité de silence et présence de modulations d'amplitude) permet d'expliquer ces catégories. Ils ont également montré que les jugements perceptifs étaient moins bien expliqués par ces indices acoustiques quand il s'agissait de catégorisation que pour des jugements de dissemblance (obtenus par comparaison par paire).

Dans un cadre très proche de celui qui nous intéresse, [Miranda *et al.*, 00] ont cherché des catégories permettant de décrire les sons complexes issus de leur logiciel de synthèse granulaire pour donner des points de repères aux utilisateurs et ainsi en permettre un meilleur contrôle. Même si la méthode utilisée pour proposer leur catégories est empirique, il est intéressant de voir qu'une partie de la taxonomie est inspirée des travaux de [Schaeffer, 66], et que les auteurs s'intéressent à des catégories de sons que l'on peut qualifier de sons abstraits pour un problème qui est celui du contrôle d'un algorithme de synthèse sonore. A noter également que plusieurs catégories ont été introduites pour combler les celles qui ne sont que peu détaillées par la typologie de Schaeffer (sons "chaotiques", sons "explosifs", *etc.*).

Enfin [Bergman *et al.*, 09] ont montré que dans le cas de sons d'alarmes, les catégories sont faites aussi bien sur la base de critères acoustiques (notamment temporels) que sur la base de critères émotionnels.

Ces études montrent que des critères très différents peuvent être utilisés pour catégoriser des sons. Il faudra donc évaluer dans quelle mesure ces critères varient entre les sujets.

3.3 Test 1 : Classification libre

3.3.1 Sujets

Vingt-six sujets (17 hommes et 9 femmes) ont participé au test. Ils avaient en moyenne 23,9 ans (min 19 ans, max 30 ans). Parmi eux, 20 ont une pratique instrumentale régulière et deux ont pratiqué la musique électroacoustique (musique que pratiquent les compositeurs ayant fourni les sons à l'origine des stimuli). La majorité des sujets travaillaient sur le campus du Groupement de Laboratoire de Marseille - CNRS.

3.3.2 Stimuli

Sons abstraits

Les sons ont été sélectionnés à partir d'un corpus de plus de 2000 fichiers fournis par les étudiants de la classe d'électroacoustique du conservatoire de Marseille ainsi que son professeur P. Gobin. Un premier travail de tri a permis d'éliminer les sons trop facilement identifiables, ceux qui paraissaient similaires (même corps sonore manipulé) et ceux de moins bonne qualité (bruit de fond, choc sur le micro, ...) ce qui a réduit le corpus de départ à environ 200 sons. Pour répondre aux trois critères de sélection énoncés précédemment, nous avons procédé à une phase de pré-test identique à celui proposé aux sujets sur ce corpus. L'analyse des résultats nous a permis d'optimiser le corpus en termes de taille, d'homogénéité et de représentativité, ce qui nous a conduit à conserver au final 62 sons.

Sons de synthèse

6 sons de synthèses ont été ajoutés pour étudier les relations entre les mouvements évoqués et les transformations décrites par les lois de la physique. Pour minimiser les différences acoustiques avec les sons abstraits, ces stimuli ont été construits sur la même base, c'est-à-dire en transformant des sons (tous différents) provenant des échantillons collectés au près des compositeurs. Dans un premier temps, l'évolution de leurs paramètres au cours du temps a été gelée en utilisant une re-synthèse par vocodeur de phase [Dolson, 86]. Dans un second temps, 2 sons ont été obtenus en appliquant des variations d'intensité pour une trajectoire rectiligne et une trajectoire circulaire. 2 autres sons ont été obtenus en appliquant des variations de hauteur sur les deux mêmes trajectoires. Nous avons ensuite simulé un mouvement rectiligne et un mouvement circulaire en appliquant séparément les variations de hauteur et d'intensité sur quatre échantillons différents. Le mouvement rectiligne est décrit pour une source qui se déplace de -50 à 50 mètres en 6 secondes devant un observateur fixe situé à 1 mètre de l'axe. Le mouvement circulaire est celui d'une source qui parcourt un cercle de 9 mètres de rayon à la vitesse de 18 tours par minute et dont l'observateur est à 1 mètre à l'extérieur du cercle.
Enfin les deux derniers sons ont été obtenus en appliquant des variations du taux de réverbération et de la fréquence de coupure d'un filtre passe-bas simulant le mouvement d'un objet qui s'approche d'un observateur. Pour la variation du taux de réverbération, le son initial a de plus été modulé en amplitude à la fréquence 2,3Hz car sur le son original la variation du taux de réverbération n'était pas perceptible.
Le détail du calcul et de la mise en œuvre de ces transformations sera largement développé section 5.2.

Les stimuli étaient tous au format "wav" 44100 Hz 16bit. Les 68 sons ont été égalisés en sonie par l'expérimentateur en comparant les sons un à un à l'aide du logiciel Audacity [4].

3.3.3 Protocole

Le test s'est déroulé au casque (Stax 3R202 avec préampli Stax SRM310) dans une cabine audiométrique. Les sons étaient présenté à travers la carte son interne de l'ordinateur ("imac G4").

Le test se base sur une interface graphique [5] qui permet aux sujets de passer de la phase d'entraînement au test de manière totalement autonome et de prendre le temps nécessaire pour lire les consignes affichées sur l'écran avant de commencer.

Phase d'entraînement

Le test débutait par une phase d'entraînement qui consistait à écouter des échantillons sonores et dire s'ils évoquaient ou non un mouvement/déplacement. La consigne précisait de ne pas s'intéresser à l'origine du son. En effet il est apparu dès le début des pré-tests qu'il s'agissait d'une consigne difficile à suivre et qui devait être bien expliquée.
Un autre intérêt de cette phase était de donner la possibilité de juger qu'un son n'évoque aucun mouvement pour que les sujets fassent clairement la distinction entre les sons qui sont pertinents par rapport à la tâche demandée et ceux qui ne le sont pas.
Enfin cette phase a permis à l'expérimentateur de voir si les sujets avaient bien compris les consignes et arrivaient à utiliser l'interface (les sujets pouvaient également déplacer les sons comme lors du test principal).

Test

Au début du test formel, les 68 stimuli sont répartis de manière aléatoire sur l'écran. Dans l'interface, les sons sont représentés par des carrés qui peuvent être écoutés et déplacés avec la souris (cf. figure 3.1). Il était demandé aux sujets de regrouper sur l'écran, les sons qui leur évoquaient des mouvements similaires et de ne pas s'intéresser aux autres aspects des sons. Aucune limite n'est fixée concernant le nombre de groupes à faire, la durée du test et le nombre d'écoutes de chaque son. Pour chaque sujet outre la position de chaque son sur l'écran, on enregistre la durée totale du test (entraînement non compris).
En début de séance il est possible de modifier le volume d'écoute de manière à être confortable et ne pas risquer de se fatiguer par un niveau trop élevé ou trop faible (ce niveau est ensuite fixe durant tout le test). Il était précisé qu'un questionnaire serait rempli à la fin du test sans en préciser le contenu. Enfin, du papier et des crayons étaient mis à disposition des sujets sans préciser d'utilisation potentielle.
Le détail des consignes et des questionnaires est donnés en annexe A.

3.3.4 Résultats

A la fin du test on obtient simplement les coordonnées de chacun des 68 sons sur l'écran. Il s'agit dans un premier temps de retrouver les groupes par le biais d'un algorithme d'optimisation qui prend en entrée le nombre de groupes réalisés par le sujet et le bi-vecteur des coordonnées des sons. L'algorithme commence par répartir aléatoirement autant de points que de groupes souhaités dans l'espace contenu par les quatre points extrêmes des données. L'algorithme calcule les distances entre chaque point et chaque centre de groupe et attribue à chaque son, le numéro du groupe dont le centre est le plus proche. On recalcule ensuite la nouvelle position des centres pour

4. Logiciel libre d'édition et de montage : http://audacity.sourceforge.net/
5. Interface ©Matlab utilisant en partie du code développé par J. Marozeau Ircam-CNRS et M. Barthet LMA-CNRS

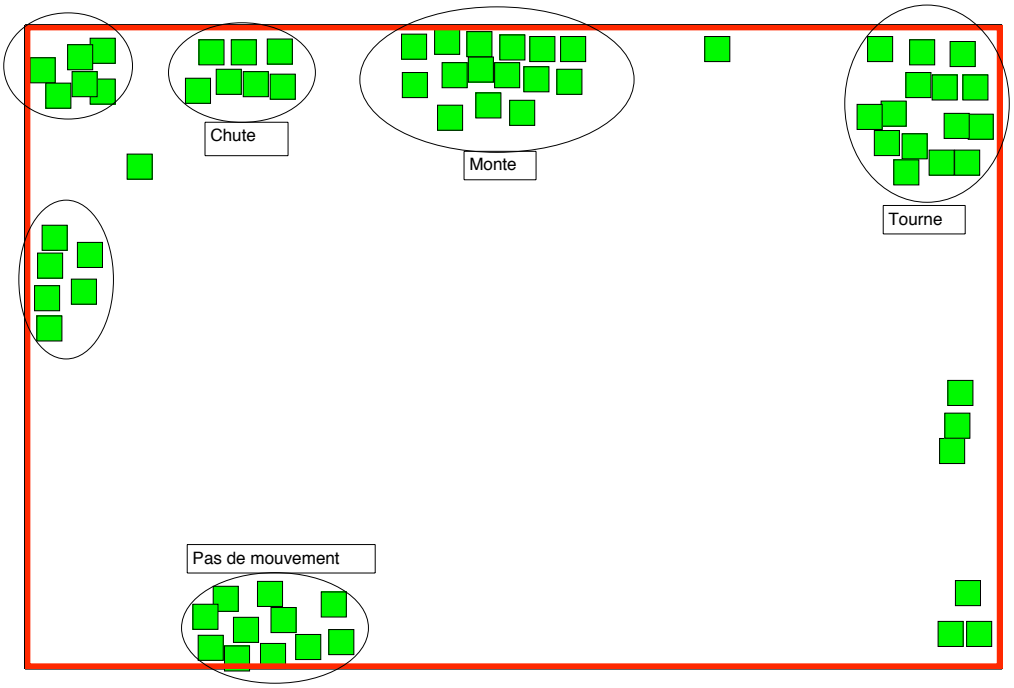


Figure 3.1 – Interface graphique pour le test de classification libre. Chaque carré représente un son qu'il est possible d'écouter et de déplacer à l'aide de la souris de l'ordinateur (clique gauche sur le carré pour écouter et clique droit pour déplacer). A la fin du test, les sujets doivent décrire les groupes. Les groupes et les labels présentés ici sont donnés à titre d'exemple et ne correspondent pas aux résultats d'un sujet.

minimiser la somme des distances centre-son. Le calcul complété est réitéré 200 fois avec différents "sets" de positions initiales des centres des groupes. Les données sont finalement stockées sous forme de vecteur comportant pour chaque son, le numéro du groupe auquel il appartient. A noter qu'une vérification systématique de l'attribution des groupes pour chaque son a été effectuée par l'expérimentateur, la procédure automatique n'ayant été mise en place que pour faciliter la mise en forme des données.

Durée du test et difficulté

Le test a duré de 23 minutes à 1h40 et en moyenne 43 minutes. Cette durée est calculée à partir du temps total du test y compris le temps utilisé pour remplir le questionnaire où les sujets réécoutent les sons avant de nommer les groupes. Dans certains cas, le test s'est poursuivi par une discussion entre le sujet et l'expérimentateur ce qui a parfois beaucoup augmenté la durée mesurée. Les sujets ont majoritairement trouvé que le test demandait une attention soutenue et même s'il était possible de faire des pauses, aucun sujet n'en a fait ou du moins n'est sorti de la salle durant le test. Globalement la tâche a été jugée difficile et surtout longue par une majorité de sujets.
Un sujet n'a pas été jusqu'au bout du test : n'ayant pas classé la moitié des sons au bout de 1h30, l'expérimentateur a interrompu le test.
La consigne "essayez de ne pas vous intéresser à l'origine du son" a posé problème à cinq sujets dont deux l'ont mentionné en commentaire et trois ont fait une majorité de groupes dont le nom n'est pas associé à un mouvement mais à une source sonore (exemple : voiture, avion, bruit d'eau...).
Les deux sujets qui ont pratiqué la musique "concrète" ont particulièrement éprouvé des difficultés à faire ce test. Ils ont tous les deux fait très peu de catégories et surtout ont clairement porté leur attention sur les techniques avec lesquelles les stimuli avaient été construits.

Nombre de groupes

Le nombre de groupes effectués par les sujets illustre de façon probante la variabilité des réponses. Comme le montre la figure 3.2, toutes les valeurs entre 3 et 15 groupes ont été observées avec une moyenne de 8,8 groupes. On remarquera également qu'un sujet se démarque avec un nombre de groupes (21) nettement supérieur aux autres sujets. Les données démographiques (âge, niveau d'étude, *etc.*) ne permettent pas de d'expliquer ces différences.

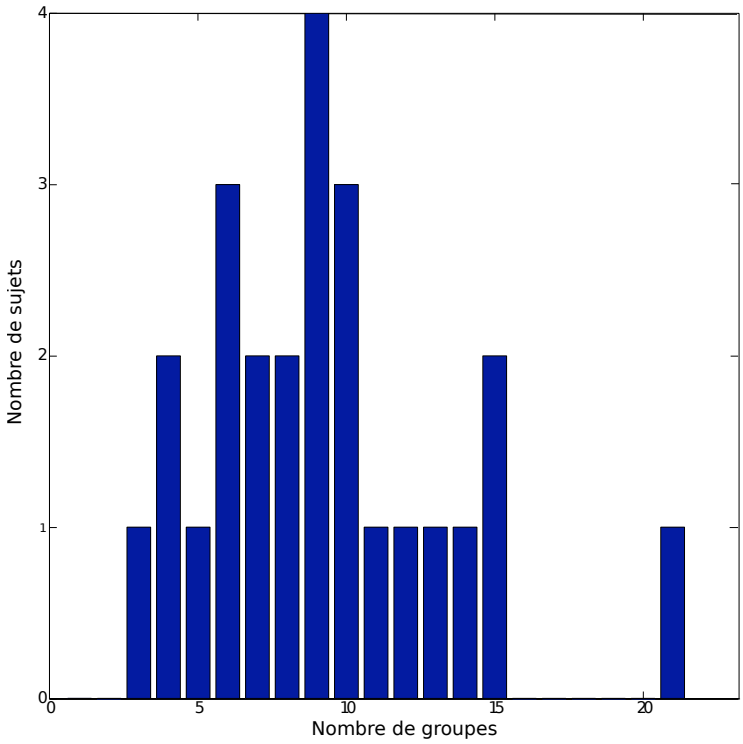


Figure 3.2 – Nombre de groupes effectués par les sujets.

Détermination des catégories à partir du questionnaire

Le questionnaire complété par les sujets à la fin du test demandait notamment de fournir quelques mots pour décrire chaque catégorie effectuée. L'interprétation de ces descriptions sémantiques, est une étape essentielle pour pouvoir regrouper et comparer les réponses des sujets. Pour cela, nous avons simplement compté le nombre de fois que chaque mot (ou mot similaire) a été utilisé et sélectionné ceux qui étaient utilisés par une majorité de sujets. Ces données sont regroupées dans le tableau 3.1 (p. 46).
Chaque sujet ayant son propre vocabulaire et n'attribuant pas forcément le même sens à un mot, nous nous sommes limités aux regroupements qui peuvent être faits sans ambiguïté (exemple : tombe - chute). D'autres expressions comme "sons aériens " et "mouvement vertical" sont beaucoup plus ambiguës et n'ont pas été regroupées. Si on considère les catégories effectuées par 50% ou plus des sujets, on obtient 4 catégories à savoir "S'éloigne", "Approche", "Tourne" et "Descend". Deux autres catégories sont proches de ce seuil des 50% à savoir "Monte" (46% des sujets) et la catégorie "Passe" (42%).
Il est certain que les intitulés des catégories ainsi que la décision de retenir ou non un mot pour une catégorie donnée sont des problèmes discutables. Le nombre de fois que des sujets ont fait une catégorie est donc une donnée qui n'a qu'une valeur qualitative ce qui dans un premier temps nous incite à retenir les 6 catégories. A noter que pour chaque son, on détermine le nombre de fois qu'il a été classé dans chacune des 6 catégories. Tous les sons ont été cités au moins une fois dans l'une

des 6 catégories, celles-ci sont donc susceptibles de bien représenter le corpus.
Ces analyses basées sur les descriptions sémantiques des groupes ont ensuite été complétées et validées par des analyses statistiques.

Analyse de proximité

Dans un premier temps les données ont été mises sous forme de matrices de dissemblance en considérant que deux sons qui auront été classés par une majorité de sujets dans la même catégorie sont très semblables, deux sons qui ne sont que rarement classés dans la même catégorie sont très dissemblables. On obtient ainsi une matrice symétrique de 68×68 qui contient des valeurs comprises entre 0 et 1 où 1 correspond à deux sont qui n'ont jamais été classés dans le même groupe et 0 à deux sons qui ont toujours été classés ensemble (les éléments de la diagonale sont donc tous nuls). Mettre les données sous cette forme nous donne accès à de nombreuses méthodes statistiques dans la limite du constat que ces distances ne sont pas euclidiennes puisqu'elles sont déduites d'un nombre de sujets ayant classés deux sons ensemble. Nous avons retenu l'analyse en cluster ascendante hiérarchique en utilisant la distance moyenne entre groupes (*cf.* figure 3.3).

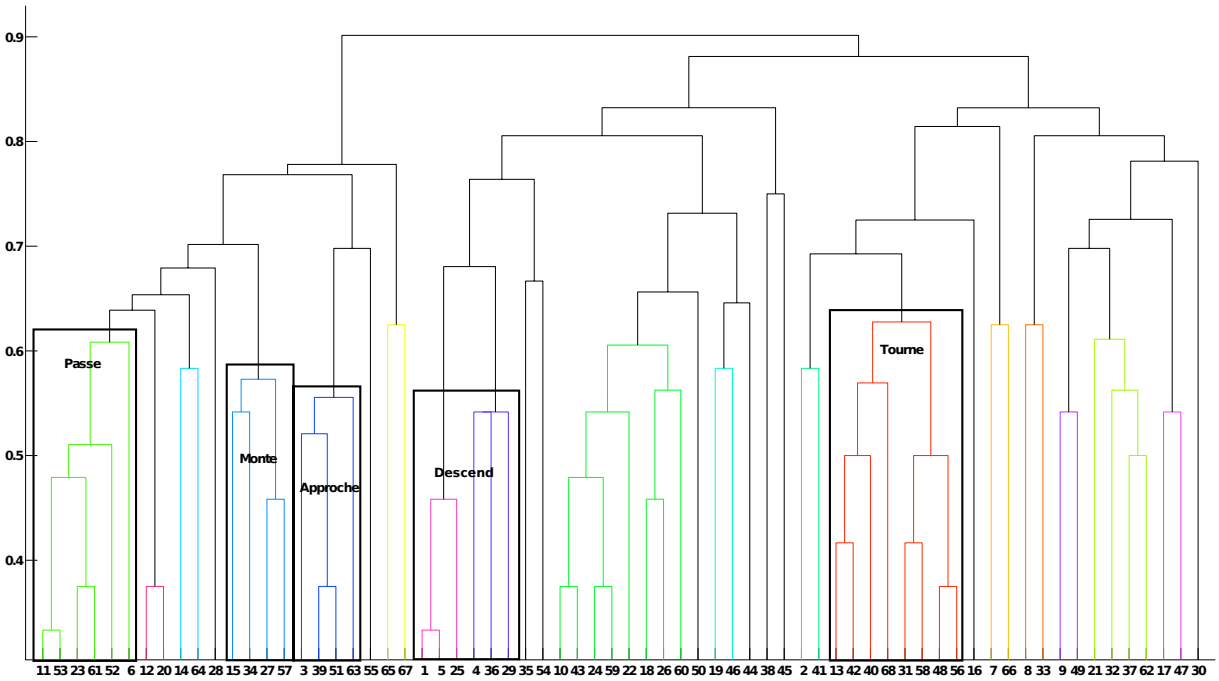


Figure 3.3 – Dendrogramme des réponses obtenues par catégorisation libre et transformées en dissemblance par le nombre de fois que deux sons ont été classés dans la même catégorie. Les sons encadrés correspondent aux groupes identifiés par l'analyse des descriptions sémantiques.

Il n'existe pas de critères pour définir *a priori* le seuil en dessous duquel on considère que des sons appartiennent au même groupe. Une première approche consiste à chercher des correspondances entre les résultats de l'analyse en cluster et des résultats obtenus par ailleurs. L'interprétation des mots que les sujets ont utilisés pour décrire les catégories nous a donné pour chaque son, le nombre de fois qu'il a été classé dans une catégorie. Ainsi on peut identifier des sons qui ont été classés par plus de 50% des sujets dans l'une des catégories et qui sont donc susceptibles de la représenter. En isolant ces sons sur le dendrogramme, on retrouve ainsi 5 des 6 groupes précédemment définis (les groupes sont encadrés sur la figure 3.3). **Ces 5 catégories ont été retenues pour le test 2**. A noter que la catégorie qui ne se retrouve pas dans l'analyse de proximité est la catégorie "S'éloigne". Une explication possible pour cette différence entre les analyses est que ce mouvement est inclut dans la majorité des autres mouvements (un son qui monte s'éloigne également de l'auditeur).

Sujet	Monte	S'éloigne	Passe	Descend	Tourne	Approche	autres
1				Qqchose qui tombe			grincement
2		Sons qui s'éloignent				qui se rapprochent	Sons statiques
3	montée démarrage, montée rapide		Doppler	Chute	sinusoïdale vague*		constants
4			passe ultrarapide				
5	sons aériens*					se rapproche	lents - rebonds
6	mvt vers le haut		mvt longitudinal qui passe devant moi		mvt circulaire	mvt longitudinal qui arrive sur moi	circulaire longitudinal, va et viens
7	montée	éloignement			rotation forcée, lourde		
8	montée en tension puis envol	qui part : mvt filant et rapide		chute qui dégouline	choses qui tournes	rapprochement	rebonds, en fin de course qui s'arrête
9	qui monte	qui s'éloigne	de gauche à droite ou de droite à gauche	amorti sa chute, tombe, descend	qui tourne	qui arrive (bref) qui arrive (long)	s'ouvre, se ferme rebondit/sautille
10	décollage*	étirement*	écoulement*	chute	rotation		vitesse/accélération
11		éloignement rapide	passage rapide	chute+choc	rotation	approche lente, +ralentissement+arrêt	accélération
12			passe lentement, passe rapidement		roule qui se répète roule sec rapide		
13	monte	s'éloigne	flotte*	descend	tourne, vibre	se rapproche	rebondis, flotte
14	objets qui montent			descendent	bouge autour d'1 position d'équilibre		
15	latéral ascendant				vibration*		amplification à partir d'un centre
16	vertical bas en haut			haut en bas	mvt rotatif	se rapproche	
17		éloignement accélérat°		Chute	tourne, oscillant	rapprochement	ralenti décéléré
18		s'éloigne	passe	tombe	tourne	se rapproche	s'ouvre, se ferme
19		corps qui s'éloigne	mvt rapide qui passe devant soi		sons circulaires		
20	ça monte	part, s'éloigne	passe devant ou à côté de nous	descend/tombe	ça roule/tourne	ça vient jusqu'à soi	sintille/gigote
21		sons qui partent	arrivent passent et s'éloignent	descendent	avance recule et ainsi de suite*	arrivent et zig zag, arrivent se stoppent	se dispersent
22		s'éloignent			va et viens av/ar	se rapprochent	
23	montée ouverture			chute rapide, descente	rotation sur soi même	avancée chaotique*	
24		éloignement		objets qui chutent			mouvement de corde
25	translation bas-haut	objet poussé*	passe à côté de moi	tombe sur le sol	roule sur le sol*	arrivant en face de moi	
26					roulement/vibrat°tenue		frappé/déchiré

Tableau 3.1 – Labels utilisés par les sujets pour décrire les groupes, organisés suivant les 6 catégories retenues. Les groupes notés *correspondent à des noms qui n'ont pas été utilisés comme appartenant à la catégorie définie en haut de colonne. Les mots séparés par une virgule indiquent qu'il s'agit d'un groupe différent.

En revenant, pour chaque son, au nombre de fois qu'il a été cité dans chacune des catégories, on détermine ceux qui ont obtenu les plus fort pourcentages ($> 70\%$) et qui sont donc susceptibles de représenter au mieux une catégorie donnée. On évalue en plus, le fait que ces sons ne soient pas également représentatifs d'une autre catégorie. On peut finalement, pour chacune des 5 catégories, déterminer un son représentatif qu'on peut qualifier de son "prototypique". Ces sons ont été utilisés dans le test 2 pur définir les catégories.

Remarques

Sur le dendrogramme, les sons du groupe du milieu (10, 43, *etc.*, en vert) correspondent à des sons courts ($< 1s$) et c'est un point que plusieurs sujets ont soulevé dans le questionnaire en expliquant que ces sons étaient trop courts pour pouvoir être comparés à d'autres.
La plupart des catégories sont "non-exclusives" et par exemple, plusieurs éléments de la catégorie "s'éloigne" se retrouvent aussi dans la catégorie "monte" ainsi que d'autres dans la catégorie "passe". Comme l'ont fait remarquer deux sujets dans le questionnaire, un même son peut très bien avoir deux types de mouvements simultanément. On peut donc se poser la question de ce qui détermine le choix, pour un sujet, de classer un son dans l'une ou l'autre des catégories quand 2 sont possibles. Il était relativement prévisible que devant la liberté qu'offre la tâche demandée, la variabilité entre les réponses serait relativement importante et c'est en grande partie pour cette raison que le second test a été mené.

3.4 Test 2 : Classification contrainte

3.4.1 Sujets

16 sujets (6 femmes, 10 hommes) ayant en moyenne 24,4 ans (min 21, max 30) ont participé au second test. Tous ont participé au premier test. 12 sont musiciens dont les deux sujets ayant pratiqué la musique concrète.

3.4.2 Protocole

Le second test a été réalisé avec les mêmes stimuli que le test 1. Entre le premier et le second test nous avons observé un intervalle de temps d'au moins deux semaines. L'interface est une version modifiée de celle du test 1 (figure 3.4) de manière à découper verticalement l'écran en deux parties et découper la partie supérieure en cinq catégories. Chaque catégorie était représentée par un son prototypique définie à l'issue du test 1.

Les consignes étaient très proches de celles données au premier test dans le sens où il s'agissait toujours de regrouper les sons évoquant des mouvements similaires avec cette fois un nombre de groupes fixe et un son de référence pour chaque groupe. La consigne était précisément de "regrouper les sons de la partie inférieure de l'écran dans l'une des 5 catégories de la partie supérieure en comparant les mouvements respectifs qu'ils évoquent". Il leur était précisé que dans le cas où un son ne correspondait à aucune des catégories, il fallait le laisser dans la partie inférieure de l'écran et qu'il était possible qu'une ou plusieurs des catégories reste vide.

3.4.3 Résultats

Onze sujets sur seize ont trouvé le second test plus simple et dix sur seize ont trouvé le nombre de catégories suffisantes pour classer les sons. Seul un sujet a laissé une des 5 catégories vide, le son prototypique ne lui évoquant aucun mouvement. Neuf sujets sur seize ont choisi de nommer les catégories sur la feuille alors que ce n'était pas précisé dans la consigne. Pour les analyses, on se limite à évaluer pour chaque son le pourcentage de fois qu'il a été classé dans une des 5 catégories. Les sons classés à plus de 50% dans l'une des catégories ainsi que les sons typiques sont retenus

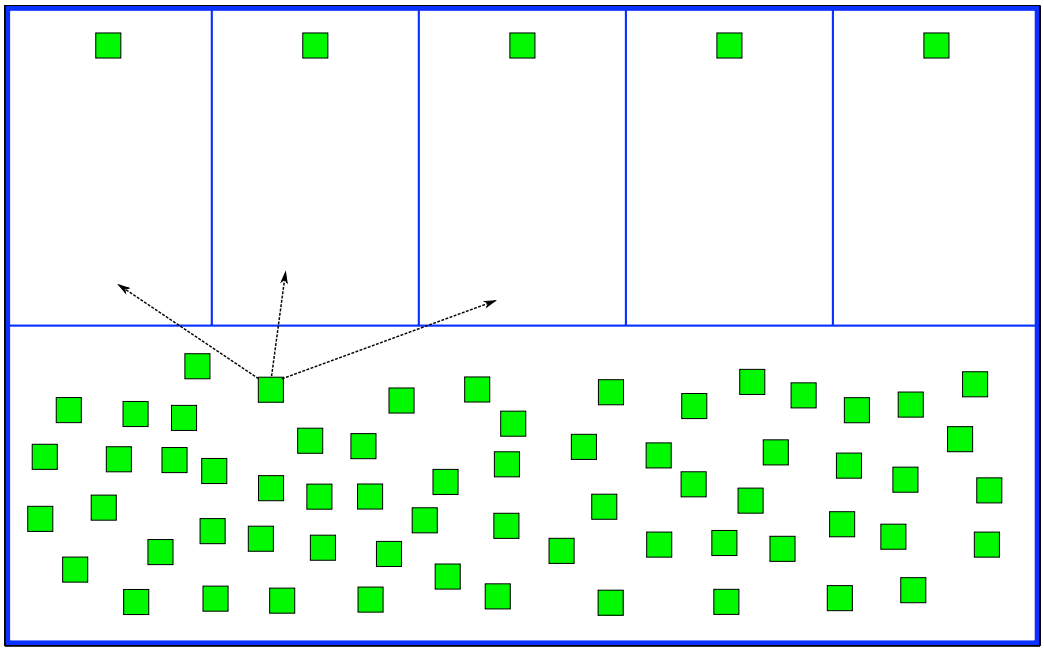


Figure 3.4 – Interface du test 2. Les sons de la partie inférieure de l'écran doivent être déplacés vers l'une des 5 catégories de la partie supérieure de l'écran. De même que pour le premier test, il est possible d'écouter et de déplacer les sons à l'aide de la souris (en cliquant sur les carrés). Les 5 sons représentatifs des catégories (en haut) ne peuvent pas être déplacés

pour être analysés (résultats regroupés dans le tableau 3.2). On notera qu'aucun son n'a finalement été retenu dans la catégorie "Approche" et qu'un seul son se trouve dans la catégorie "monte".

Tourne	Passe	Approche	Descend	Monte
68 : 100 %	53 : 100 %	66 : 100%	38 : 100%	27 : 100%
13 : 75 %	59 : 81,25 %		1 : 68,75%	28 : 68,75%
58 : 75 %	11 : 62,5 %		29 : 68,75%	
42 : 68,75 %	16 : 56,25 %		5 : 62,5%	
21 : 62,5 %	61 : 56,25 %		4 : 62,5%	
48 : 56,25 %				
56 : 56,25 %				
41 : 56,25%				
2 : 56,25 %				

Tableau 3.2 – Numéro des sons sélectionnés suite au test 2 pour être analysés et pourcentage de classification dans la catégorie mentionnée. Les valeurs 100% correspondent aux sons prototypiques

Cette analyse est également intéressante en comparaison avec le premier test dans le sens où elle permet de comparer à la fois la cohérence des réponses des sujets et la validité de la méthode d'analyse utilisant les descriptions sémantiques des groupes effectués.

3.4.4 Discussion

La première remarque concerne le fait que les sons précédemment appelés sons de synthèse (sons numérotés de 63 à 68) n'ont pas été traités différemment des autres par les sujets. Aucune des analyses ne fait apparaître une catégorisation spécifique à leur égard. Il se trouve que deux des

sons prototypiques utilisés pour le second test (pour les catégories "Tourne" et "Approche") sont des sons de synthèse.

Pour la catégorie “Tourne”, le son correspondait à la trajectoire circulaire dont l’observateur se situe à un mètre à l’extérieur du cercle et où seule la variation d’intensité est appliquée à l’échantillon initial. Au premier test, le son simulant la même trajectoire avec la variation de hauteur due à l’effet Doppler, se trouve classé dans la même catégorie à moins de 30% et ni l’analyse en cluster exclusif, ni l’analyse en cluster hiérarchique ne les regroupe. La même remarque peut se faire concernant les deux stimuli obtenus avec les mêmes transformations mais pour une trajectoire rectiligne d’une source à vitesse constante qui passe à un mètre de l’observateur. Le son transformé avec une variation d’intensité se retrouve bien à plus de 50% dans la catégorie “Passe” (de l’analyse du questionnaire) ce qui n’est pas le cas pour celui transformé pour simuler la variation de hauteur. De même, ces sons ne sont jamais regroupés dans les résultats de l’analyse en cluster. La comparaison avec l’étude [Lutfi et Wang, 99] est intéressante car les auteurs montrent que l’indice qui globalement est le plus utilisé pour différentier deux mouvements simulés pour des sources acoustiques se déplaçant à des vitesses différentes ou ayant une accélération différente, est la variation de hauteur due à l’effet Doppler. Les auteurs indiquent que les échelles de temps ont une grande place dans l’importance relative des différentes transformations utilisées pour simuler le mouvement et notamment pour des faibles vitesses pour lesquelles les variations d’intensité deviennent plus utilisées par les sujets que les variations de hauteur. En revanche les auteurs ne posent pas la question de savoir si la simulation donne bien une sensation de mouvement, ni si la trajectoire est bien reconnue. Ce sont deux aspects qui sont tenus pour vrai *a priori* et non validés *a posteriori*.

Pour la catégorie "approche", le son prototypique est celui qui correspond à la trajectoire simulée avec une variation du taux de réverbération. En revanche, pour la même trajectoire, la variation de timbre effectuée avec un filtrage passe bas avec fréquence de coupure variable n’a pas été concluante. Le son obtenu se trouve classé dans cette catégorie (au second test) à 43% mais aussi à 31% dans la catégorie "Monte". Ce résultat doit être confronté au fait que l’effet a été exagéré par rapport au filtrage réel de l’air qui n’est perceptible que pour des sons riches en hautes fréquences (au-dessus de 3000Hz) et pour des distances source-observateur relativement grandes (de l’ordre de la centaine de mètres).

3.5 Analyses des signaux

Le problème de décrire des signaux aussi variés que ceux de notre expérience est relativement complexe, notamment parce que pour qu’un outil d’analyse soit efficace il faut souvent utiliser des connaissances *a priori*, qui sont forcément relatives à une classe de signaux particulière. La diversité des sons de notre corpus ne nous permet pas d’utiliser des méthodes qui s’appliquent à des sons harmoniques ou des signaux ayant une structure formantique marquée. L’analyse des signaux présentée ici utilise la démarche classique basée sur le calcul de descripteurs de timbre qui ont fait leur preuve et sont devenus incontournables. Comme on l’a discuté section 1.4.2, la vaste littérature concernant les descripteurs a été introduite avec des objectifs précis qui ne sont pas forcément liés aux nôtres (compréhension de la notion de timbre pour les instruments de musique, reconnaissance automatique,...). Pour nous l’objectif est au final d’être en mesure de concevoir un synthétiseur ce qui implique d’identifier des descripteurs qui non seulement décrivent un comportement valable pour un corpus de son donné mais en plus donnent une meilleure compréhension du phénomène. De plus, la majorité des descripteurs utilisés dans les études sur le timbre ne sont en général pas adaptés à une description dynamique du signal. Toutes ces raisons font qu’il a été nécessaire d’introduire des descripteurs spécifiques à notre problématique et également de définir des variables pour décrire l’évolution temporelle des descripteurs calculés. En calculant un nombre important de descripteurs, on espère également augmenter les chances d’avoir dans notre corpus, des descripteurs permettant de comprendre les caractéristiques des signaux utilisées par les sujets pour donner leur jugement. Cette section présente donc le calcul des descripteurs et leur mise en relation avec les résultats des

tests perceptifs.

3.5.1 Descripteurs de timbres "classiques"

Nous avons utilisé les définitions et l'implémentation Matlab de M. Barthet [Barthet, 08]. Les descripteurs utilisés sont : le centre de gravité spectrale, l'étalement spectral, la variation spectrale (ou flux spectral), enveloppe d'amplitude et le centre de gravité temporel.

Le centre de gravité spectral d'une densité spectrale de puissance (DSP) à K coefficients c_k a été calculé en utilisant la définition suivante :

$$Cgs = \sum_{k=1}^{K} \frac{k c_k}{\lambda + \sum c_k} \tag{3.1}$$

où λ est un paramètre de régularisation. Dans ce qui suit, on a pris $\lambda = 10^{-6}$.
Le centre de gravité spectral sert à décrire le caractère "brillant" du son [Schubert et Wolfe, 06].

L'étalement spectral est calculé grâce à la relation :

$$Es = \sqrt{\frac{\sum_k (k - Cgs)^2 c_k}{\sum_k c_k}} \tag{3.2}$$

L'étalement spectral sert à décrire la "richesse" du son. Son calcul comme fonction du temps permet de détecter les transitoires [Dixon, 06].

La variation spectrale (ou Flux spectral) d'une DSP calculée à un instant n :

$$Vs(n) = 1 - \frac{\sum_k c(n-1,k) \times c(n,k)}{\sqrt{\sum_k c(n-1,k)^2}\sqrt{\sum_k c(n,k)^2}} \tag{3.3}$$

La variation spectrale décrit les changements de la répartition spectrale d'énergie à une échelle d'observation qui est celle choisie pour le calcul des spectres à court terme.

L'enveloppe d'amplitude est extraite du signal analytique par filtrage passe-bas (de type Butterworth du second ordre) avec une fréquence de coupure de 20Hz. A noter que cette fréquence de coupure a été choisie car elle correspond à la limite supérieure des modulations d'amplitude correspondant à la force de fluctuation [Zwicker et Fastl, 90].

Le centre de gravité temporel d'une enveloppe d'amplitude $A(n)$ calculée sur N pas de temps discrets est calculé en utilisant :

$$Tc = \frac{1}{N} \frac{\sum_{n=1}^{n=N} n \times A(n)}{\sum_{n=1}^{N} A(n)} \tag{3.4}$$

Le centre de gravité temporel permet de distinguer les sons d'impacts (pour lesquels une grande partie de l'énergie est au début du son) d'autres sons.
A noter qu'on intègre également au corpus de descripteurs, la durée du signal (en seconde).

Tous les descripteurs à part le centre de gravité temporel, sont calculés comme fonction du temps. Pour les descripteurs spectraux, on utilise une fenêtre de Hanning de 512 points et un taux de recouvrement de 50% entre deux fenêtres successives. Pour chaque courbe de descripteur dépendant du temps, on calcule la **valeur moyenne** et l'**écart type**. Nous avons de plus introduit deux scalaires : la **monotonie** et le **taux de variation**. Ces deux scalaires calculés sur les fonctions temporelles des descripteurs ont été introduits pour mieux décrire leurs évolutions.

La monotonie est définie par :

$$Mn = \frac{1}{N}\sum_{n=1}^{N} sign\big(va'(n)\big) \tag{3.5}$$

où $va'(n)$ est la dérivée de la variable discrète $va(n)$ de longueur N. $Mn \approx 1$ correspond à une courbe croissante (resp. décroissante pour -1). $Mn \approx 0$ correspond à une courbe globalement horizontale (ou oscillante).

Le taux de variation est défini par :

$$Vr = \frac{1}{N-1}\sum_{n=1}^{N-1} abs\Big(sign\big(va'(n+1)\big) - sign\big(va'(n)\big)\Big) \tag{3.6}$$

qui correspond à un taux de croisement par zéro de la dérivée du signal et qui est donc corrélé avec le moment du second ordre. Le taux de variations permet donc de décrire l'irrégularité de la courbe indépendamment de l'évolution globale.
On obtient ainsi 4 descripteurs dépendant du temps $\times$ 4 scalaires (moyenne, écart type, monotonie et taux de variation)+ 2 descripteurs indépendant du temps (centre de gravité temporel et durée du signal) = 18 variables.

3.5.2 Descripteurs spécifiques

Il est évident que d'un point de vue physique, les variations d'intensité sont très liées aux déplacements de la source et ont donc de grandes chances d'être également importantes d'un point de vue perceptif. Il nous a donc semblé important de décrire au mieux les variations de l'enveloppe d'amplitude et notamment sur des échelles de temps assez longues (de l'ordre de la seconde). Il a donc été nécessaire de calculer des descripteurs spécifiques a cet aspect et qui sont détaillés ci-dessous.

Caractérisation des modulations d'amplitude Pour la catégorie "Tourne", il nous a semblé important de prendre en compte les variations périodiques et basses fréquences de l'enveloppe d'amplitude. Comme pour la plupart des méthodes de recherche de périodicité (tempo, hauteur, *etc.*)[Gouyon et Dixon, 05], notre méthode est basée sur un calcul d'autocorrélation. En pratique, à partir de l'enveloppe d'amplitude, on calcule l'autocorrélation à laquelle on vient soustraire une interpolation linéaire ou quadratique pour supprimer les composantes statiques (*cf.* figure 3.5). Finalement on extrait les pics de la transformée de Fourier de cette autocorrélation "ajustée". Les "pics" dont la largeur dépasse un certain seuil ne sont pas pris en compte dans la mesure où ils ne correspondent pas à des périodicités marquées. En pratique le seuil utilisé est de 5Hz calculé à 75% du maximum d'énergie du pic.

Au final les scalaires qui sont utilisés pour l'analyse sont les amplitudes et les fréquences des 2 pics les plus marqués, le rapport de ces fréquences et amplitudes ainsi que le nombre de pics au-dessus de la moyenne de la transformée de Fourier de l'autocorrélation ajustée dans la bande

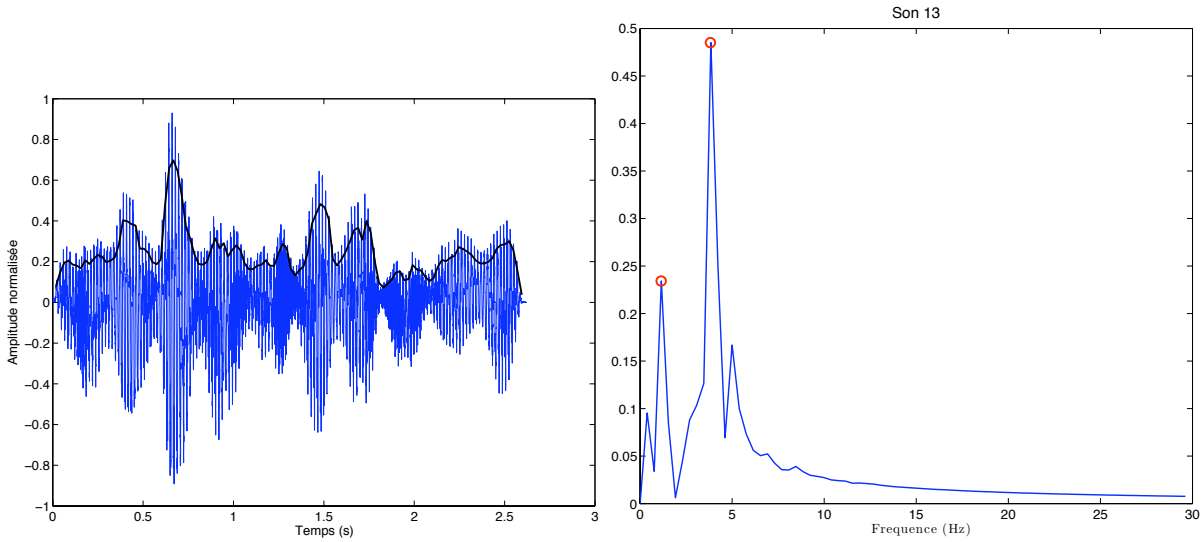


Figure 3.5 – Estimation de l'enveloppe d'amplitude (à gauche) et des fréquences de modulation (à droite) par extraction de pics de la transformée de Fourier de l'autocorrélation.

0-20Hz (soit 7 variables). Ce dernier descripteur permet, dans le cas d'une modulation d'amplitude qui varie en fréquence, de mettre en évidence ces variations.

Caractérisation des variations globales d'intensité Parmi les variations d'intensité qui nous semblent importantes de décrire on trouve en particulier celles qui correspondent aux variations se produisant à l'échelle de la durée du son, notamment pour la catégorie "Passe". Toujours à partir de l'enveloppe d'amplitude, il s'agit dans un premier temps de détecter le maximum global d'énergie. Ensuite l'enveloppe d'amplitude est séparée en deux parties avant et après le maximum d'énergie. Pour chacune de ces parties, on procède enfin à une interpolation linéaire du logarithme de l'enveloppe. Les variations globales d'énergie seront donc décrites par deux durées $t1$ et $t2$ qui correspondent à la position du maximum d'énergie par rapport au début du son (resp. à la fin), deux coefficients directeurs α_1 et α_2 qui décrivent le taux de croissance et de décroissance de l'énergie (*cf.* figure 3.6). On calcule aussi le rapport de ces coefficients directeurs pour caractériser le caractère symétrique ou non de l'enveloppe. A noter que les erreurs quadratiques E_{rr1} et E_{rr2} de l'interpolation linéaire sont utilisées car en plus de caractériser la pertinence de l'interpolation, elles fournissent une estimation des fluctuations de l'enveloppe sur les deux portions du son. Les variations globales d'intensité sont donc décrites par 7 variables.

3.5.3 Sélection de descripteurs

L'étape précédente nous a permis de calculer 32 variables scalaires pour chaque son. Etant donné la diversité des sons et les définitions de certains descripteurs, le calcul sur certains sons a donné des valeurs aberrantes qui n'ont pas été prises en compte pour la suite. Il s'agit maintenant de les confronter aux catégories pour déterminer les points communs entre les sons qui ont été classés ensemble. L'objectif est donc de sélectionner les descripteurs qui expliquent le mieux les catégories et qui pourront par la suite nous permettre de synthétiser ou transformer des sons. Dans la littérature, on trouve ce problème sous le nom générique de "fouille de données" (en anglais "Data mining") et qui consiste à retrouver par le calcul et la création de modèles basés sur des descripteurs, des informations pertinentes dans des données. En particulier, des méthodes ont été développées pour sélectionner au mieux les variables descriptives utilisées en entrée des modèles de prédiction (en anglais "feature selection methods"). Les critères de sélection à la base de ces méthodes peuvent être très différents (*cf.* [Blum et Langley, 97] pour une revue des différents critères et des méthodes

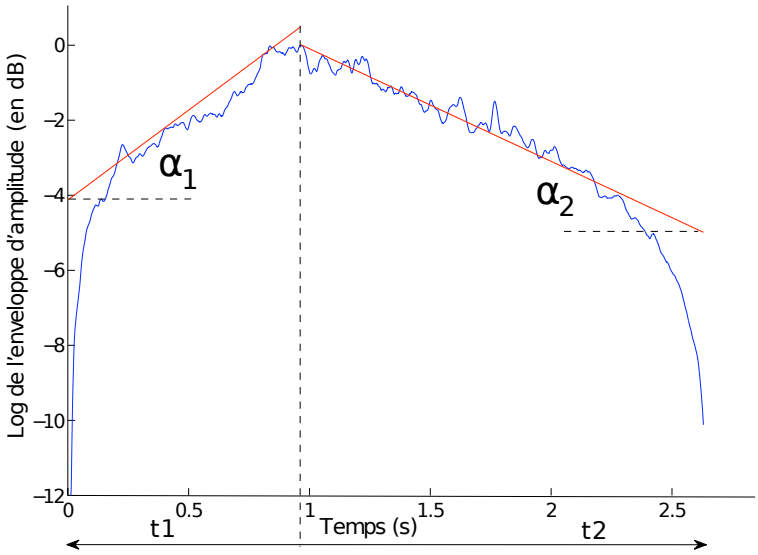


Figure 3.6 – Illustration du calcul des variables α_1, α_2, $t1$ et $t2$ à partir du logarithme de l'enveloppe d'amplitude.

de sélection). C'est pourquoi nous nous limitons à la définition suivante : "un descripteur pertinent permet de différentier deux observations différentes" (définition 1 de [Blum et Langley, 97]).

En pratique il existe deux types de méthodes de sélection de descripteurs les "wrappers" et les "filters". Le premier type étant forcément associé à une méthode d'apprentissage, puisqu'on évalue la sélection en comparant les performances du modèle en fonction des descripteurs utilisés en entrée, il ne sera donc pas détaillé ici. Pour les méthodes dites "filters" il en existe un grand nombre et il est souvent difficile de les comparer. Par exemple dans [Herrera *et al.*, 02], les auteurs ont comparé deux méthodes qui ont fourni des sets de descripteurs très différents (seulement 2 descripteurs sur une dizaine coïncidaient). Il est possible d'évaluer les descripteurs un par un ou de les combiner et de prendre en compte leurs éventuelles corrélations. A ce propos, dans [Guyon et Elisseeff, 03], les auteurs indiquent que "deux variables qui sont inutiles par elles-mêmes peuvent être utiles ensembles". Un autre problème qui se pose est la possibilité ou non de mélanger dans le corpus de descripteurs, des variables discrètes et des variables continues ou plus simplement des variables qui varient sur des échelles très différentes.

Il faut enfin préciser que le but est de ne retenir que des descripteurs "pertinents" qui pourront être contrôlés dans un outil de synthèse ce qui constitue une différence fondamentale par rapport aux approches MIR [6] dont le but est de retrouver une information sans contrainte de compréhension des liens acoustique-perception sous-jacents. En définitive, nous avons choisi deux méthodes dont la première procède à une discrétisation des variables en trois niveaux [Fayyad et Irani, 93]. La seconde méthode fonctionne directement à partir des variables continues et toutes deux se basent sur le calcul du critère F de Fisher (d'où le nom de la méthode en anglais : "Fisher filtering"). En pratique, il s'agit pour un set de variables et un ensemble de catégories donné, de construire un modèle de régression multiple et de voir, quand on enlève ou rajoute une variable comment évolue le F partiel de Fisher (*cf.* [Blum et Langley, 97] et [Guyon et Elisseeff, 03] pour le détail de fonctionnement de cette méthode). Nous avons utilisé les implémentations fournies par le logiciel TANAGRA [7].

6. Music Information Retrieval, domaine déjà mentionné section 1.4.2
7. http://chirouble.univ-lyon2.fr/~ricco/tanagra/fr/tanagra.html

3.5.4 Résultats

Pour évaluer la robustesse des méthodes, nous avons ajouté au corpus de descripteurs, une variable aléatoire de distribution uniforme avec des valeurs comprises entre 0 et 1. Contrairement à certaines autres méthodes testées dans une étude préliminaire, cette variable n'a bien évidemment pas été sélectionnée par nos deux méthodes.
Les tableaux 3.3 et 3.4 regroupent les résultats obtenus par la méthode de discrétisation et ceux obtenus par la méthode qui utilise les descripteurs sans transformation.

Pour la catégorie "Approche", aucun descripteur n'a été significativement mis en évidence par l'une ou l'autre des méthodes et pour la catégorie "Monte", seul un descripteur a été relevé par la méthode de discrétisation (le taux de variation de l'enveloppe d'amplitude). Ces problèmes de sélection s'expliquent probablement par le faible nombre de sons dans ces catégories (resp. 4 et 6 sons). Les résultats qui sont discutés ne concernent donc que les catégories "Tourne", "Passe", et "Descend". On s'intéressera en particulier aux descripteurs qui sont significatifs pour les deux méthodes.

Catégorie "Tourne" : Le paramètre a_1 qui ressort en premier est tout à fait cohérent avec la physique dans le sens où il indique la nécessité d'une modulation d'amplitude. Le rapport f_0/f_1 indique quand à lui que deux modulations d'amplitude semblent nécessaire.

Catégorie "Passe" : Err_2 correspond à la partie décroissante de l'enveloppe et indiquerait la nécessité que cette décroissance soit log-linéaire. En revanche, il est assez difficile de donner une explication de la pertinence du paramètre $Npeak$ qui caractérise les modulations d'amplitude.

Catégorie "Descend" : Une première constatation est que la grande majorité des sons de cette catégorie sont des impacts. Cette constatation se retrouve avec le descripteur ls qui indique des sons courts (moins de $0,94s$) ainsi que le descripteur $t1$ pour lequel la discrétisation indique qu'il doit être inférieur à $0,36$ et qu'on peut donc interpréter comme un temps d'attaque court caractéristique des sons d'impact. Il est intéressant de remarquer ici que dans la trajectoire d'un objet, l'impact ne correspond qu'à la conséquence du mouvement et qu'a priori, le son ne contient pas d'information sur la trajectoire qui est donc probablement "imaginée" par l'auditeur.

"Tourne"		"Passe"		"Descend"	
Descripteur	Seuil	Descripteur	Seuil	Descripteur	Seuil
cgs_std	224 (Hz) [-]	**Npeak**	3,5 [-]	t_2	0,875[-]
$a0/a1$	0,1 [+]	$\mathbf{Err_2}$	0,26 [-]	**ls**	0,94[-]
$f1$	0,69 (Hz) [-]			α_1	6,1 (db/s) [+]
a1	0,06 [+]			$\mathbf{t_1}$	0,36(s) [-]
f0/f1	0,05 [+]				

Tableau 3.3 – Discrétisations des descripteurs et valeurs des seuils. [+] signifie que la valeur du descripteur pour les sons qui appartiennent à la catégorie est au-dessus du seuil ([-] au-dessous). *Les abréviations correspondant aux descripteurs sont données dans le tableau 3.5*

"Tourne"			"Passe"			"Descend"		
Descripteur	F	P-Value	Descripteur	F	P-Value	Descripteur	F	P-Value
a1	19,94	0,000076	$\mathbf{Err_2}$	8,03	0,00748	**ls**	14,60	0,0005
$Npeak$	8,32	0,00658	**Npeak**	7,54	0,00934	Err_1	14,38	0,00055
f0/f1	8,23	0,00684	Err_1	6,31	0,01663	$a1$	9,98	0,0032
			En_mn	4	0,0531	$\mathbf{t_1}$	9,67	0,0036

Tableau 3.4 – Résultats pour les variables non discrétisées. En gras les descripteurs qui se retrouvent dans les deux analyses.
Les abréviations correspondant aux descripteurs sont données dans le tableau 3.5

cgs_std :	Ecart type du centre de gravité spectral
ls :	longueur du signal
$f1$:	Fréquence de la deuxième modulation d'amplitude
$f0/f1$:	Rapport des fréquences de modulation
$a0$:	Taux de modulation (de la première modulation)
$a1$:	Taux de modulation (de la seconde modulation)
$a0/a1$:	Rapport des taux de modulation
$Npeak$:	Nombre de pics dans la transformée de Fourier de l'autocorrélation modifiée
t_1 :	Temps jusqu'au maximum d'énergie de l'enveloppe d'amplitude
t_2 :	Temps entre le maximum d'amplitude et la fin du son
Err_1 :	Erreur quadratique dans l'interpolation linéaire de la première moitié du signal.
Err_2 :	idem Err_1 pour la deuxième moitié
α_1 :	Coefficient directeur de l'interpolation linéaire de la première moitié
α_2 :	idem α_1 pour la deuxième moitié
En_vr :	Taux de variation de l'enveloppe d'amplitude
En_mn :	Monotonie de l'enveloppe d'amplitude

Tableau 3.5 – Listes des abréviations correspondant aux descripteurs.

3.6 Conclusion

Le protocole qui a été mis en place ainsi que les analyses qui ont suivi ont soulevé un certain nombre de difficultés. Dès la première étape qui consiste à sélectionner des sons et construire un protocole, on a mis en évidence la grande quantité de possibilités à envisager et la difficulté qu'il y a pour appuyer ces choix sur des bases solides. Il a néanmoins été possible de mettre en évidence des grandes catégories de mouvements ainsi que des sons représentatifs de ces catégories. L'analyse des sons est également une étape complexe qui fait à elle seule l'objet de recherches. Notre approche basée sur le calcul de descripteurs scalaires et la sélection de descripteurs qui séparent au mieux les sons d'une catégorie, donne des résultats encourageants, mais pas suffisants pour aborder la conception d'un synthétiseur. En d'autres termes, même si nous sommes partis d'un corpus qui contenait 68 sons, nous n'avons pas collecté suffisamment de données nécessaires à la réalisation d'analyses suffisamment robustes pour faire de la synthèse.
A ce stade, il semble donc nécessaire de faire le point sur les aspects qu'il faut améliorer pour être en mesure d'atteindre les objectifs que nous nous sommes fixés. D'abord il semble que l'approche par catégorisation bien que pertinente par la grande liberté qu'elle offre aux sujets, ne permet pas de prendre en compte la diversité des attributs du mouvement. Autrement dit, on a observé des degrés de raffinement très différents pour les catégories effectuées par les sujets, et la méthode en elle-même ne permet pas de prendre en compte ce raffinement. La catégorisation, qu'elle soit libre ou contrainte, ne permet pas non plus d'évaluer simultanément deux attributs d'un même son. Par exemple, un son peut monter et tourner en même temps et l'approche par catégorisation contraint les sujets à choisir une catégorie au détriment de l'autre. Ces problèmes ont eu pour conséquence de restreindre le nombre de sons qui ont pu être sélectionnés sans ambiguïté et par là même les analyses des signaux qui ont suivi. Enfin les catégories proposées nous donnent une base intéressante pour mieux définir la notion de mouvement et la façon dont elle est perçue mais ne fournit en aucun cas un espace de paramètres de contrôle pour la synthèse.

Il est donc nécessaire de définir un nouveau protocole plus directement adapté à notre problématique de synthèse sur la base de ce que nous a enseigné ce test. La conception et la mise en œuvre d'un tel protocole font l'objet du chapitre suivant.

Chapitre 4

Approche par caractérisation graphique

4.1 Dessiner les sons, perception multimodale **58**
4.1.1 Dessiner les sons 59
4.1.2 Interactions son-image 61
4.2 Interface de caractérisation graphique **62**
4.2.1 Choix des paramètres de contrôle 62
4.2.2 Implémentation 62
4.3 Constitution et caractérisation d'un corpus de sons abstraits **66**
4.3.1 Présélection des sons 66
4.3.2 Corpus final 69
4.4 Protocole **69**
4.4.1 Sujets 69
4.4.2 Conditions 69
4.4.3 Test 3 : Caractérisation du corpus 70
4.4.4 Test 4 : Caractérisation graphique 70
4.5 Résultats **74**
4.5.1 Test 3 74
4.5.2 Test 4 76
4.5.3 Analyse des trajectoires 76
4.5.4 Résumé 84
4.6 Conclusion **85**

Les travaux qui font l'objet de ce chapitre sont également décrits dans [Merer et al., 11]. Pour les stimuli et les interfaces graphiques des tests, le lecteur est invité à visiter la page web de l'article : `http://www.lma.cnrs-mrs.fr/~kronland/Movement/`

Les tests décrits au chapitre précédent nous ont permis de définir des catégories perceptives de mouvements évoqués par les sons. Les conclusions de ce test nous ont incité à mettre au point un nouveau protocole d'évaluation perceptive des sons qui soit davantage adapté à la problématique du contrôle de la synthèse.
Lors de l'analyse des questionnaires de fin du test de catégorisation libre, nous avions remarqué que plusieurs sujets avaient spontanément utilisé des dessins pour décrire les catégories qu'ils avaient effectuées (*cf.* figure 4.1) alors que les consignes indiquaient simplement de décrire les catégories avec quelques mots et ne parlaient pas de dessin. L'utilisation de dessin est donc apparue comme une manière intuitive et incontournable pour décrire le mouvement évoqué par les sons. Ce chapitre s'intéresse donc à la mise en place d'un protocole original qui utilise le dessin pour décrire le mouvement évoqué par les sons.

4.1 Dessiner les sons, perception multimodale

Dessiner les sons ou générer des sons à partir de dessins est une problématique qui a intéressé beaucoup de chercheurs et de musiciens. Il s'agira donc de recenser ici quelques-unes des approches qui peuvent nous apporter des informations importantes dans la mise en place de notre protocole et

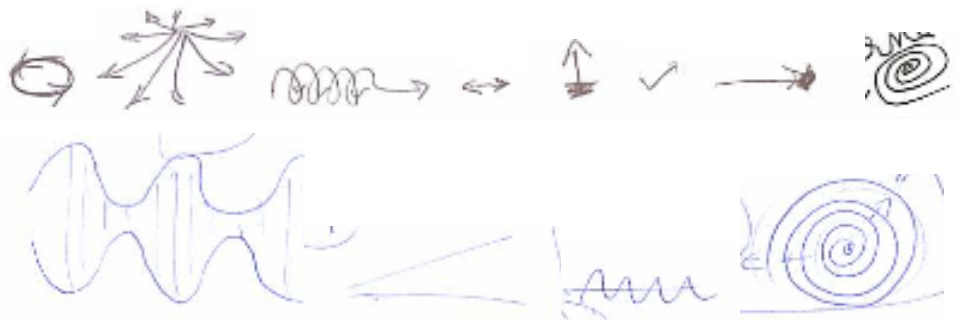


Figure 4.1 – Dessins effectués spontanément par les sujets pour décrire les catégories à la fin du test de catégorisation libre (*cf.* section 3.3)

l'analyse des données récoltées.
L'étude des interactions entre les différentes modalités et notamment entre le sonore et le visuel est un domaine de recherche à part entière. Mettre en place un protocole qui vise à décrire le sonore à l'aide de dessins va nécessairement induire des interactions multimodales et requiert une grande prudence.
Le dessin met également en jeu des processus qui ne se limitent pas au visuel mais également aux aspects gestuels et moteurs. Ces aspects doivent également être pris en compte dans la conception du protocole, pour minimiser les éventuelles interactions qui pourraient perturber la validité de nos résultats.

4.1.1 Dessiner les sons

L'utilisation conjointe de son avec du dessin est très ancienne, ne serait-ce que dans le domaine de la musique avec la notation musicale. Dans un document de synthèse concernant la musique et le mouvement [Repp, 93] décrit les travaux des trois "pionniers allemands" qui les premiers, se sont intéressés aux liens mouvement-musique et surtout à l'utilisation du dessin pour décrire les mouvements. Les travaux menés par Sievers dans les années 20 consistaient à analyser les gestes utilisés lors de la lecture de textes qu'il retranscrivait à l'aide d'un dictionnaire de courbes de type "boucle". A la même période et en utilisant les mêmes méthodes que Sievers, Becking s'est intéressé à des compositeurs et plus particulièrement aux accentuations dynamiques caractéristiques de ces compositeurs ce qui l'a amené à transcrire un tableau de ces profils (*cf.* figure 4.2). Il est intéressant de constater que ces travaux mettaient en avant différentes échelles de représentation. Dans les années 30, Trulist s'est intéressé aux mouvements (impliquant le corps entier) suggérés par une pièce musicale et qui se décrivent grâce à quatre types de courbes (dent de scie, lacets supérieurs et inférieurs et courbe en huit). Il faut également mentionner les travaux de Clynes (également rapportés par [Repp, 93]) qui a travaillé à partir de courbes obtenues par mesure de la pression d'appui sur un capteur dédié, lors de l'écoute de musique. Dans le cadre d'un travail autour des UST, [Rousset, 08] a proposé une notation graphique (*cf.* figure 4.3). Il est très intéressant de remarquer que cette notation utilise systématiquement des repères pour marquer la direction ou le début et la fin de la notation.
Plusieurs logiciels de composition assistée par ordinateur ont été conçus pour permettre à l'utilisateur de représenter le son avec des dessins afin de contrôler des modules de transformation ou de synthèse [Lesbros, 96, Lesbros, 99], [Farbood *et al.*, 07], [Wu et Li, 08]. Le plus célèbre de ces systèmes est probablement le UPIC [1] de Iannis Xenakis composé d'une tablette graphique associée à un moteur de synthèse additive. De par le nombre de systèmes qui ont fait suite au UPIC, il apparaît clairement

1. Unité Polyagogique (agogique désigne de légères modifications de rythme ou de tempo dans l'interprétation d'un morceau) Informatique du CEMAMu (Centre d'Etudes de Mathématiques et Automatique Musicales)

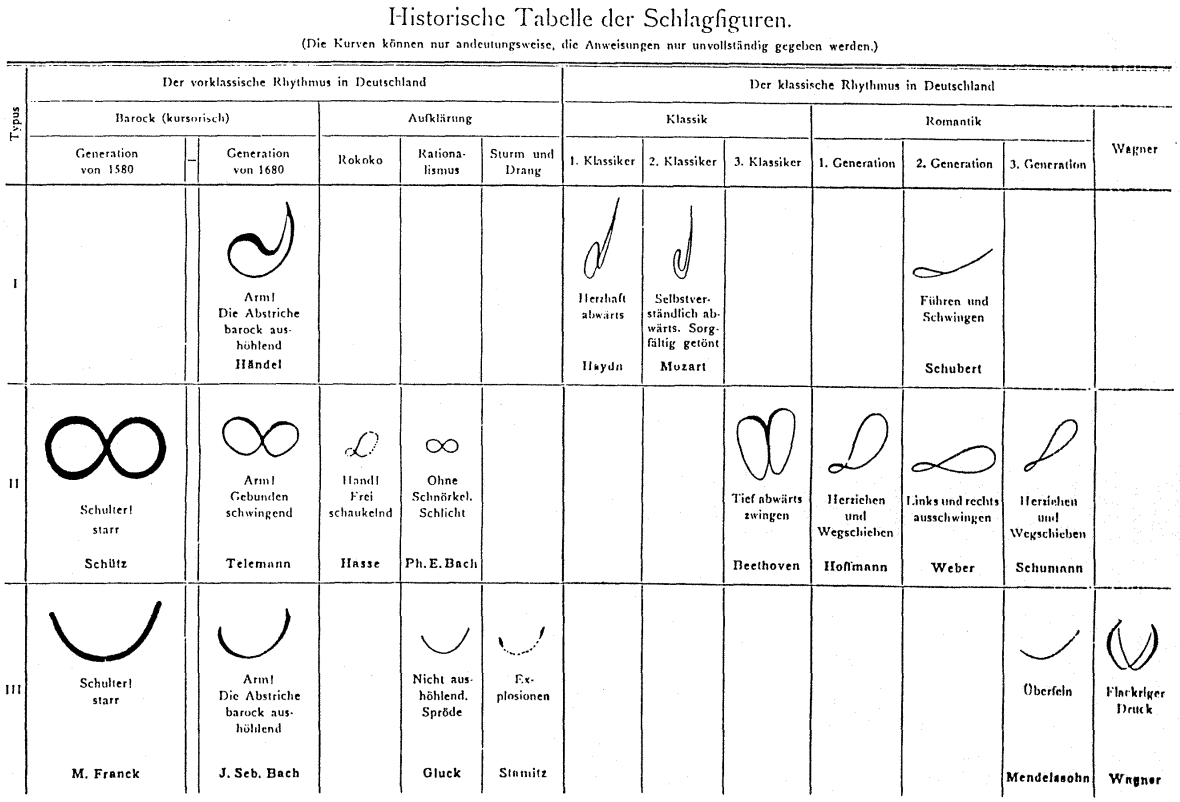


Figure 4.2 – Courbes caractéristiques de certains compositeurs allemands de Becking (reproduit à partir de [Repp, 93])

qu'il existe un fort intérêt pour les dispositifs permettant de générer des sons à partir de dessin. En dehors du domaine de la musique, il faut citer des travaux qui ont été réalisés dans le but de permettre à des aveugles de se représenter des courbes par le biais de la sonification (projet TeDub [2]). C'est dans ce contexte que l'on trouve un des rares exemples de situation inverse c'est-à-dire où des auditeurs dessinent pour décrire ce qu'ils ont entendu. En effet, [Brown et Brewster, 03] ont mené une étude pour évaluer un dispositif de sonification de courbes (avec deux courbes superposées) en demandant aux sujets de produire les courbes qui leur semblaient être à l'origine des sons qui leur étaient donnés à entendre. Les sons étaient tous obtenus en reliant l'axe vertical des courbes à la hauteur de notes jouées, les notes étant déclenchées en parcourant les courbes. Les résultats montrent que les courbes d'origine ne sont jamais parfaitement reproduites mais un certain nombre d'éléments sont correctement reproduits comme l'endroit où se croisent les courbes ainsi que la forme globale. Il est très intéressant de constater que même avec une stratégie de sonification aussi simple et explicite, la reconnaissance des courbes n'est pas évidente.
Enfin, la seule autre étude demandant aux sujets de décrire le son par des dessins est celle de [Minard *et al.*, 10] qui ont utilisé des dessins libres dans le but de catégoriser des sons d'environnement. Dans cette étude, l'objectif était plutôt de relier les dessins obtenus aux lois d'évolutions de descripteurs de timbre ce qui a contraint les auteurs à ne s'intéresser qu'à des courbes planes et continues. Ce projet qui a été mené simultanément au nôtre sans que l'on en ait connaissance, a l'intérêt de montrer la pertinence d'une description graphique dans un tout autre contexte, celui de la catégorisation de sons d'environnement.

2. http://projects.dedicon.nl/tedub/

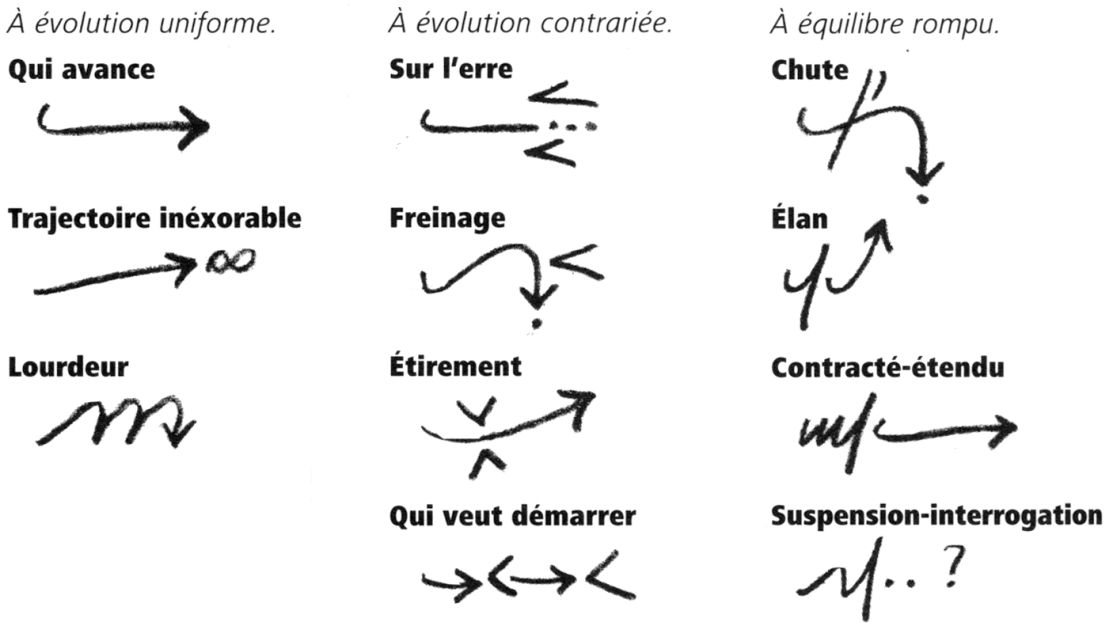


Figure 4.3 – Proposition de notation graphique des Unités sémiotiques temporelles d'après [Rousset, 08]

4.1.2 Interactions son-image

Dans le domaine de la vision, il existe un effet de persistance de mouvement (en anglais motion after-effect) qui fait que l'on a l'impression d'observer un mouvement en sens inverse après avoir fixé un objet mobile et que regarde un objet statique (par exemple après avoir fixé une chute d'eau quelques minutes et qu'on regarde un rocher à côté). Dans [Kitagawa et Ichihara, 02], les auteurs ont mis en évidence un phénomène similaire où un son stationnaire est perçu comme étant non-stationnaire (augmentation d'intensité) en présence d'un stimulus visuel mobile. Inversement [Shams *et al.*, 02] ont montré qu'en présentant un flash de lumière unique simultanément avec une succession de bips sonores on percevra plusieurs flashs dans la limite d'une échelle de temps caractéristique d'environ 100 ms. Dans une étude célèbre de 1974 (rapportée par [Zwaan *et al.*, 04]) Loftus et Palmer ont présenté des vidéos d'accidents de voiture et demandaient aux sujets d'estimer la vitesse de la voiture avant impact avec différentes formulations. Selon que la question parlait de "contact", de "choc" ou de "crash", les vitesses estimées augmentaient. Il n'est nullement nécessaire de rappeler que de manière générale, la perception est multimodale et qu'il existe des interactions entre le visuel et le sonore. On doit malgré tout mentionner le célèbre effet "ventriloquist" qui décrit des modifications de la localisation d'une source quand les informations visuelles et auditives ne coïncident pas. Des études sur ce phénomène ont montré que l'intégration multimodale est presque optimale [Alais et Burr, 04] et que, contrairement aux idées reçues, il ne met pas en évidence la supériorité d'une modalité par rapport l'autre.
Par contre, les interactions multimodales peuvent prendre une forme particulière quand il s'agit de mouvement [Alaerts *et al.*, 09] du fait des implications motrices (comme on l'a discuté dans la section 1.5.2). Dans une expérience visuelle mettant en jeu des disques de couleur identiques se rapprochant l'un de l'autre, se superposant puis s'éloignant, [Sekuler *et al.*, 97] ont montré que la présence de son 150ms avant, pendant ou 150ms après le moment où les disques se superposent, augmente très significativement l'impression que les disques rebondissent ensemble. Dans une situation ambiguë comme celle-ci, le son même s'il n'est que peu concordant avec l'image a donc une forte influence sur la perception globale de la situation. En plus des problèmes d'interaction multimodale, on doit donc prendre en compte le fait de travailler sur une dimension particulière

pour la perception.

4.2 Interface de caractérisation graphique

L'étude de la littérature a été complétée par des essais (*cf.* figure 4.4) de mise en place de protocoles pour évaluer la faisabilité d'un test d'écoute où les sujets dessinent le mouvement évoqué par des sons abstraits. Un des premiers problèmes qui est apparu rapidement lors de ces essais est dû aux différences d'aptitudes qui existent entre deux individus pour dessiner une seule et même chose. En effet, les premiers prototypes proposaient de réaliser des dessins libres pour décrire les formes et au-delà des difficultés d'analyse que pose le dessin libre, il est apparu qu'ils ne fournissaient pas une base d'évaluation des sons qui soit commune à tous les sujets. C'est donc sur la base de ces différents essais qu'une grande partie les choix de conception ont été faits par la suite.

La solution que nous proposons et qui est l'une des grandes originalités de notre projet, est de fournir une interface de dessin paramétrique qui permet aux sujets de dessiner en agissant sur des curseurs linéaires et différents autres contrôles graphiques. Bien entendu l'action sur ces curseurs doit se traduire par la visualisation instantanée de la trajectoire. Les sujets sont donc en mesure de produire des dessins parfaitement identiques, le problème principal étant pour nous de définir les paramètres qui leur seront mis à disposition.

4.2.1 Choix des paramètres de contrôle

Le choix des paramètres de contrôle est une étape critique puisqu'elle nécessite de faire des compromis entre les possibilités laissées aux sujets et la complexité du dispositif. En effet, une interface trop complexe nécessitera un temps d'apprentissage qui sera trop long pour que l'on puisse l'utiliser dans un test perceptif mettant en jeu des dizaines de sujets et de sons. Pour choisir les paramètres, nous nous sommes principalement basés sur les résultas obtenus au chapitre précédent (chapitre 3). Ainsi, l'interface doit permettre de reproduire les 5 grandes catégories identifiées ("Tourne", "Monte", "Passe", "Descend", "Approche"). En conséquence de quoi, l'interface doit permettre de contrôler la forme et l'orientation de la trajectoire pour pouvoir distinguer des trajectoires circulaires de trajectoires linéaires, des trajectoires qui montent de trajectoires qui descendent. Les dessins que les sujets ont faits spontanément (figure 4.1) nous donnent plus précisément des formes qui doivent être reproduites : spirale, cercle, sinusoïde, hélice et ligne droite. On remarque également que les sujets ont presque systématiquement ajouté des flèches pour indiquer le sens de parcours de la trajectoire ainsi que des références donnant l'orientation de la trajectoire (par exemple un trait horizontal sous la trajectoire en spirale pour indiquer le sol). Nous avons testé des protocoles avec un maximum de possibilités de contrôle (incluant tous les éléments décrits précédemment) et nous avons notamment mis en évidence la complexité induite par la manipulation d'une interface permettant de positionner la trajectoire n'importe où dans un espace à trois dimensions autour de l'auditeur. Cette possibilité a ainsi été rejetée, les trajectoires ont été placées face à l'auditeur à une distance moyenne fixe. De plus certains sujets ont créé des catégories qui font état de vitesses différentes, c'est donc un paramètre qui doit être contrôlable. Au final, nous avons abouti à 9 paramètres qui permettent de dessiner une trajectoire, la dynamique étant représentée par une sphère qui se déplace le long de cette trajectoire de façon synchrone avec la lecture du son. Le détail de chaque paramètre est donné dans la section suivante.

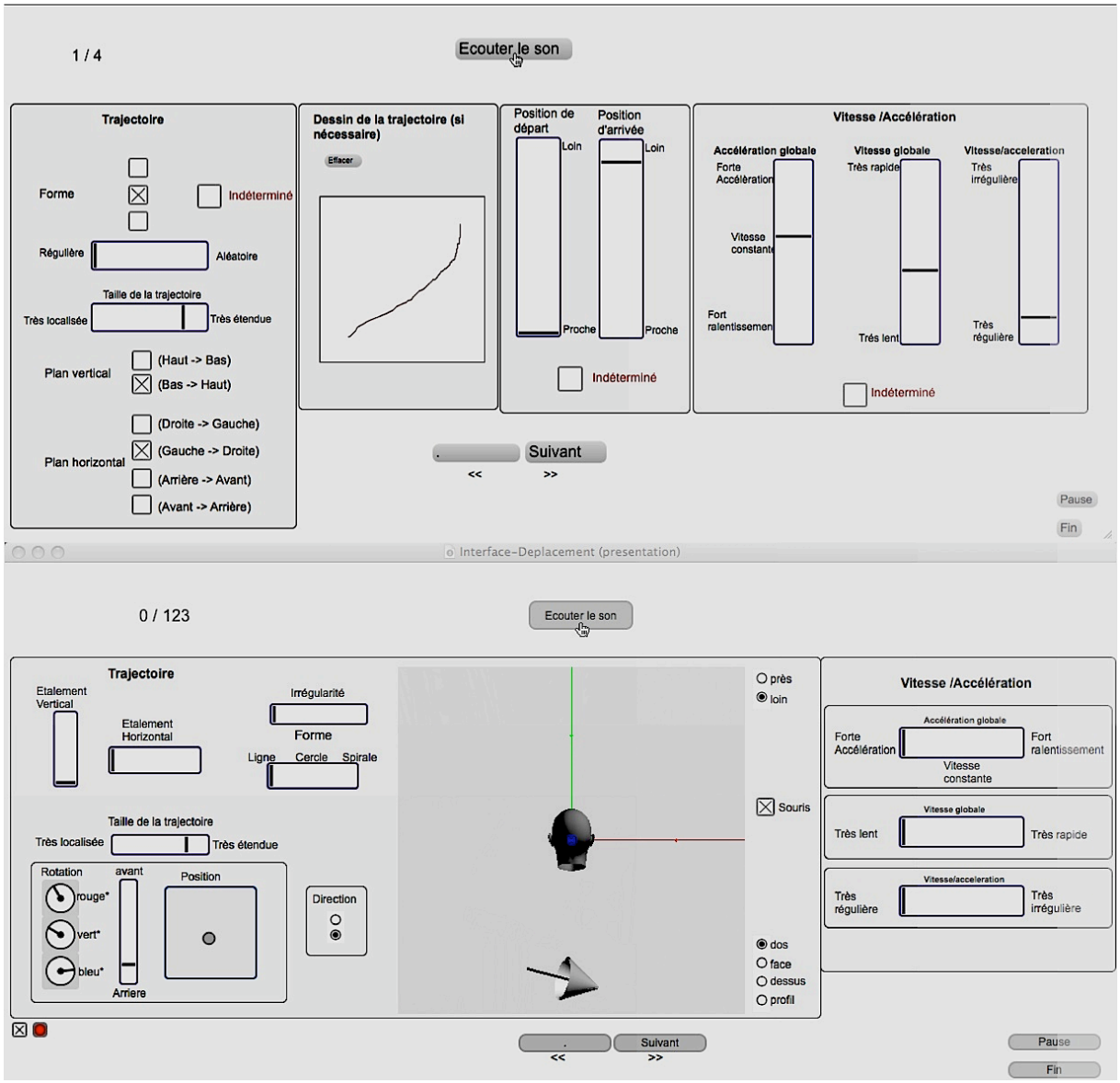


Figure 4.4 – Deux exemples d'interfaces graphiques pour l'évaluation perceptive qui ont été envisagées durant la conception du protocole. L'interfaces du bas permettait le positionnement des trajectoires à 3 dimensions autour de l'auditeur. Ces exemples font partie d'environ 40 versions différentes qui ont été développées dont une dizaine représentent des changements majeurs.

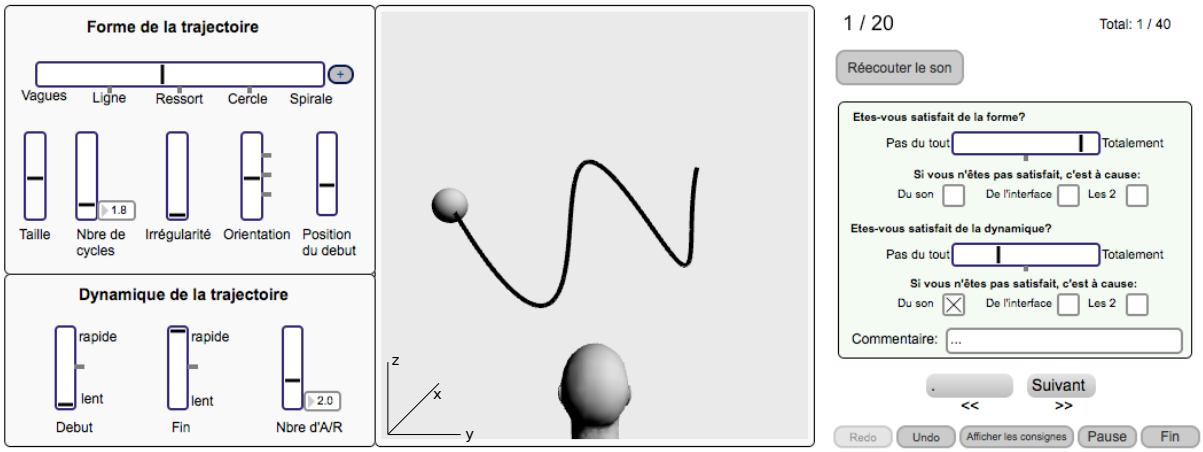


Figure 4.5 – Interface graphique du Test 4. A gauche les curseurs de forme et dynamique de la trajectoire. Quand on clique sur "Réécouter le son", la sphère se déplace le long de la trajectoire en fonction des paramètres de dynamique. La partie droite de l'écran est réservée à l'auto-évaluation et la navigation dans le test (lancer l'évaluation d'un autre son, faire des pauses, ...).

4.2.2 Implémentation

Pour limiter au maximum le nombre de variables manipulées, toutes les formes sont obtenues à partir de l'équation d'une hélice à laquelle on rajoute une partie aléatoire :

$$\begin{cases} x(u) = A_1(1 - A_2 u)cos\Big(2\pi(A_3 u + A_4)\Big) + A_5 Rnd_1(u) \\ y(u) = A_6(1 - A_2 u)sin\Big(2\pi(A_3 u + A_4)\Big) + A_5 Rnd_2(u) \\ z(u) = A_7 u + A_5 Rnd_3(u) \end{cases} \tag{4.1}$$

- A_1, A_6 et A_7 correspondent donc à l'étalement de la forme dans chacune des 3 directions de l'espace
- A_2 ajoute une dépendance de l'amplitude suivant le début ou la fin de la trajectoire (ce qui permet de faire des spirales)
- A_3 correspond à la fréquence, A_4 à la phase à l'origine
- A_5 est l'amplitude de la partie aléatoire du signal
- $Rnd_i(u)$ sont différentes séquences d'un processus aléatoire

On se place dans un repère x-y-z où les axes correspondent respectivement à l'axe du regard, l'axe des oreilles et l'axe vertical pour un observateur regardant une trajectoire situé devant lui.

Mapping entre paramètres de l'interface graphique et paramètres de la courbe

La partie "Forme de la trajectoire" de l'interface graphique comporte 6 curseurs :
- "Forme"
- "Taille"
- "Nombre de cycles"
- "Irrégularité"
- "Orientation"
- "Position du début"

Ces curseurs sont reliés aux paramètres de l'hélice de la manière suivante :

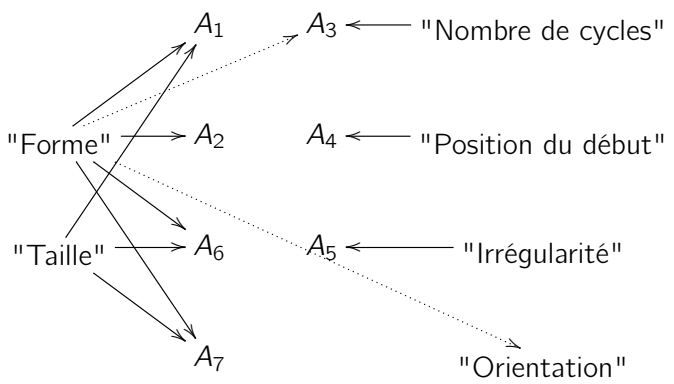


Figure 4.6 – Mapping entre les paramètres de l'interface et les paramètres de l'hélice

- Le curseur linéaire de forme est relié aux paramètres de façon à obtenir cinq formes pré-définies équi-réparties le long du curseur de passer continûment d'une forme à l'autre. Les 5 formes sont : "Vague" ($A_6, A_2 = 0$), "Ligne" ($A_1, A_2, A_6 = 0$), "Ressort" ($A_2 = 0$), "Cercle" ($A_2, A_7 = 0$) et "Spirale" ($A_7 = 0$).
- Le paramètre "Taille" est défini entre 0 et 1 où 1 correspond à une trajectoire qui occupe toute la fenêtre et 0 à une trajectoire qui occupe 10% de la fenêtre. Ce paramètre est multiplié aux variables

contrôlant l'étalement de l'hélice suivant les trois axes (A_1, A_6 et A_7).
- Le paramètre "Nombre de cycles" est directement relié au paramètre A_3 et peut varier de 1/2 à 8 périodes. Comme ce paramètre n'a pas de sens pour une trajectoire de type ligne droite, il est grisé quand le paramètre "Forme" est sur "Ligne".
- Le paramètre "Irrégularité" est directement relié au paramètre A_5 et peut varier de 0 à 1, où 1 correspond à des variations aléatoires dont l'amplitude est de même échelle que la taille de la trajectoire.
- Le paramètre "Orientation" contrôle une fonction globale d'openGL [3] qui permet à l'affichage de faire tourner les objets dans les trois directions de l'espace. Cependant, l'utilisateur ne peut agir que sur des rotations autour de l'axe y, de $-180°$ à $+180°$, pour décrire le caractère horizontal ou vertical de la trajectoire.
Les cercles font l'objet d'un mapping particulier concernant l'orientation (flèche pointillé sur la figure 4.6). En effet, le paramètre "Orientation" tel qu'il a été défini n'a aucun effet sur un cercle inscrit dans le plan x-z et un cercle inscrit dans le plan x-y risquerait d'être confondu avec une ligne droite. Le paramètre de forme agit donc sur l'orientation pour que les trajectoires "cercle" et "spirale" s'affichent par défaut dans le plan vertical et redéfinit le mapping de façon à ce que le paramètre "Orientation" fasse tourner les trajectoires autour de l'axe x (les rotations sont autour de l'axe y pour les autres formes).
-Le paramètre "Position du début" modifie la phase à l'origine des cosinus et sinus dans l'équation de l'hélice (paramètre A_4) entre 0 et 2π. Comme pour "Nombre de cycles", ce paramètre n'a pas de sens pour des trajectoires en ligne droite, il est donc désactivé dans ce cas.

La partie "Dynamique de la trajectoire" de l'interface comporte 3 curseurs :
- "Vitesse de début" et "Vitesse de fin" chacun variant de 0 à 1 et dont le détail est donné par la figure 4.7. Etant donné que les sons n'ont pas la même durée, la vitesse de la sphère est ajustée pour chaque son de manière à ce que son déplacement soit synchrone avec la lecture du fichier son.
- "Nombre d'A/R" qui varie de 0.5 à 5 et permet de parcourir la trajectoire plusieurs fois.

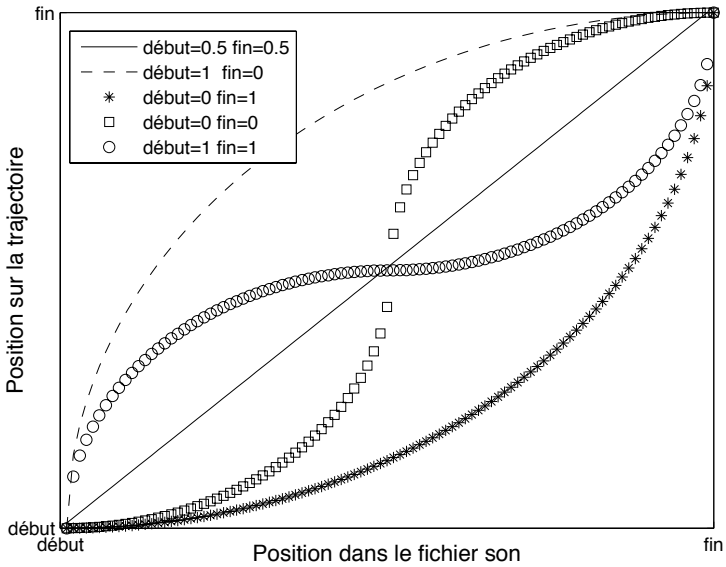


Figure 4.7 – Contrôle de la vitesse à partir des deux paramètres "Vitesse de début" et "Vitesse de fin" et de la lecture du fichier son.

3. Open Graphics Library, spécification et bibliothèque de fonctions pour l'affichage d'éléments graphique dans les logiciels. Le paramètre "Orientation" est donc complètement dépendant de l'implémentation choisie (décrite dans la suite de cette section).

A part pour “Nombre de cycles” et “Nombre d’A/R”, les valeurs numériques des curseurs ne sont pas affichées à l’écran. L’interface de dessin (figure 4.5) ainsi que les fonctionnalités nécessaires à la mise en œuvre d’un test perceptif ont été implémentés en utilisant Max/Msp/Jitter [4] qui permet entre autres la modification et l’affichage temps réel des trajectoires. La courbe de trajectoire est échantillonnée sur 300 points pour une matrice image de 300 × 300 pixels.

4.3 Constitution et caractérisation d’un corpus de sons abstraits

Dans la suite logique des réflexions menées dans le chapitre précédent concernant le type de stimuli utilisés (section 3.1.1), nous avons souhaité approfondir l’étape de sélection et de caractérisation du corpus. En particulier, il nous a semblé important de pouvoir justifier la représentativité du corpus en termes de caractéristiques spectraux-temporelles c’est pourquoi nous avons utilisé les critères de typologie de Schaeffer. Cette section décrit donc la méthode systématique que nous avons suivi pour sélectionner les stimuli.

4.3.1 Présélection des sons

Comme pour l’expérience décrite au chapitre 3, nous avons fait appel aux élèves de la classe de composition élecroacoustique du Conservatoire de Marseille pour nous fournir des échantillons qu’ils utilisent dans leurs compositions. Cette “collecte” a permis de constituer une banque d’environ 200 sons qui ont subi une procédure de sélection composée des 4 étapes suivantes (chaque étape sera détaillée par la suite) :
1) Classification suivant la typologie de Schaeffer
2) Sélection et transformations pour n’obtenir qu’un seul “flux auditif”
3) Sélections des sons “abstraits”
4) Égalisation en sonie et fenêtrage.

Typologie de Schaeffer

La classification des sons sur la base de critères "typologiques" a avant tout été introduite afin de faciliter l’utilisation (musicale) future des sons, c’est-à-dire avec pour objectif premier, la composition. Cet objectif n’est pas très éloigné du problème de sélectionner des sons pour une expérience perceptive quand celle-ci est conçue sans hypothèse de départ et où il s’agit donc de maximiser les chances de résultats.
Schaeffer part du principe que tous les sons peuvent être utilisés par le musicien, il s’est donc posé la question de savoir ceux qui sont plus pertinents et adaptés à une utilisation donnée. Concrètement, les critères morphologiques sont basés sur l’idée que notre mémoire des sons impose une durée optimale : un son trop court sera naturellement regroupé avec d’autres, un son trop long sera décomposé en plusieurs éléments. Il se base aussi sur le rapport durée/variations : une variation trop rapide sera difficilement perçue, une variation trop lente sera lassante et trop "prévisible". Enfin il fait la distinction habituelle entre caractéristiques spectrales (critères de masse) et temporelles (critère de facture). A noter que pour l’analyse, Schaeffer a développé une autre classification sur la base de critères dits "morphologiques".

Il est donc possible de classer n’importe quel son dans l’une de ces 21 cases pour obtenir un ensemble de sons aux caractéristiques acoustiques les plus diverses possibles. Pour faciliter la classification des 200 sons, une interface graphique a d’ailleurs été développée (*cf.*figure 4.9).
De nombreuses critiques ont été faites à cette typologie notamment par Schaeffer lui même car elle ne prend pas en compte des paramètres qui sont très importants en musique. Par exemple, une note jouée pianissimo et cette même note jouée forte seront toutes les deux classées N’ dans la typologie alors que leur utilisation en contexte musicale est très différente. On se doit également de

4. Cycling’74 http://cycling74.com. Le calcul des trajectoires utilise les fonctionnalités Javascript intégrées

	Durée démesurée (macro objets) Pas d'unité temporelle		**Durée mesurée** Unité temporelle → **Durée réduite** Micro-objets			**Durée démesurée** (macro objets) Pas d'unité temporelle		
	Facture imprévisible	Facture nulle				Facture nulle	Facture imprévisible	
			Tenue formée	Impulsion	Itération formée			
Hauteur définie **Masse fixe**	En	Hn	N	N′	N″	Zn	An	Hauteur définie **Masse fixe**
Hauteur complexe	Ex	Hx	X	X′	X″	Zx	Ax	Hauteur complexe
Masse peu variable	Ey	Tx Tn Trames particulières	Y	Y′	Y″	Zy Pédales particulières	Ay	**Masse peu variable**
	Unité causale				Causes multiples mais semblables			
Variation de masse imprévisible	E Cas général	T Cas général	W	F	K	P Cas général	A Cas général	**Variation de masse imprévisible**

Objets équilibrés:	**Objets redondants ou trop brefs:**	**Objets excentriques**
N: masses toniques X: masses complexes Y: masses variables ′: impulsions ″: itérations	H: sons homogènes continus Z: sons homogènes itératifs T: trames K: cellules F: fragments	E: échantillons A: accumulations P: pédales W: grosses notes

Figure 4.8 – Tableau de la typologie de Schaeffer. Les critères de "Masse" (lignes) correspondent à des aspects spectraux, les critères de "Facture" (colonnes) correspondent aux aspects temporels. *Reproduit à partir de* `http://cours.musique.umontreal.ca/MUS1217/Typologie.html`

préciser que cette classification n'est pas totalement objective dans le sens où certains sons peuvent correspondre à plusieurs cases. De plus les extrémités gauche et droite du tableau se rejoignent et il est parfois difficile de distinguer les sons dont la durée est indéfinie. Enfin, les sons constitués de plusieurs éléments simultanés qui peuvent être dissociés à l'oreille ne peuvent logiquement pas être classés suivant les critères de la typologie puisque deux éléments peuvent correspondre à deux cases différentes. C'est en partie pourquoi il a été nécessaire de considérer la notion de flux auditifs développée ci-dessous.

Flux auditifs

Introduite par Bregman [Bregman, 90], la notion de flux auditifs décrit la capacité à regrouper et séparer certains éléments du son ce qui est l'une des étapes les plus importantes dans l'analyse de "scènes auditives". Les regroupements peuvent être faits entre des éléments qui se succèdent dans le temps aussi bien que pour des éléments qui sont simultanés mais qui occupent des régions spectrales différentes. Dans notre cas, en plus du problème de classification des sons, il est nécessaire de limiter au maximum le nombre d'éléments qui pourront être perçus séparément par les auditeurs et sur lesquels leur attention pourra se porter sans contrôle possible de la part de l'expérimentateur. Même dans le cas de sons bien calibrés avec peu de composantes, l'attention de différents auditeurs, peut se porter sur différents aspects comme c'est le cas pour l'effet de confusion [Neff, 85] en masquage temporel.
Nous avons donc procédé à une vérification systématique de la présence d'un ou plusieurs flux en ajoutant une évaluation dédiée dans l'interface graphique utilisée pour la classification suivant la typologie de Schaeffer (figure 4.9). Certains sons présentant plusieurs flux ont par la suite été transformés par des filtrages (filtres à réjection de bandes) et parfois des transpositions.

Egalisation en sonie

L'égalisation en sonie des sons d'un corpus est une procédure systématique dans les tests d'écoute afin d'éviter que les sujets se basent sur cet aspect essentiel des sons, pour donner leur

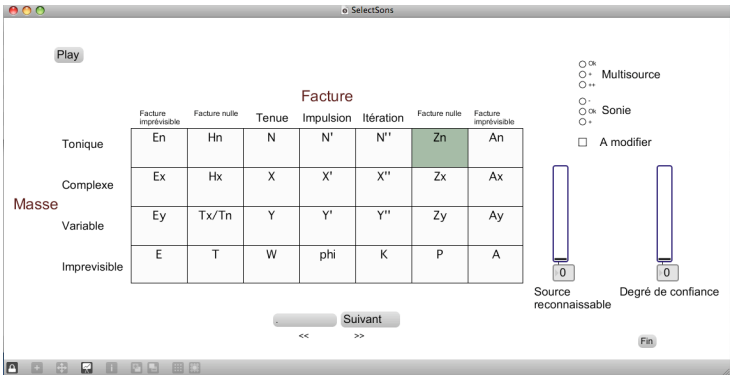


Figure 4.9 – Interface MaxMsp pour faciliter la classification d'une grande banque de sons suivant la typologie de Schaffer ainsi que l'évaluation subjective de la sonie et de la présence simultanée de plusieurs flux auditifs.

jugement. Dans le cas de sons de durées très différentes, qui sont globalement très instationnaires et avec des caractéristiques spectrales volontairement très différentes, l'égalisation en sonie est un problème relativement complexe.
D'abord il n'existe aucun modèle qui permet de comparer la sonie de tels sons. Nous avons notamment testé la Loudness Toolbox [5] et les fonctions de calcul de sonie pour les sons instationnaires qu'elle propose (modèles de [Zwicker et Fastl, 90] ainsi que celui de [Glasberg et Moore, 02]) sans qu'il soit possible de définir de façon automatique le facteur d'atténuation à appliquer aux sons pour qu'ils aient une même sonie (dont l'évaluation reste discutable).
Même lorsque l'on réalise l'égalisation en sonie "à l'oreille", plusieurs stratégies s'offrent à l'expérimentateur notamment en se basant sur le maximum d'intensité du son ou sur une "moyenne" sur le son. Enfin il est connu depuis [Neuhoff, 98] que la sonie des sons dont l'intensité croît au cours du temps fait l'objet d'un traitement particulier ce qui est le cas pour plusieurs sons de notre corpus. La stratégie que nous avons choisie, bien que toujours discutable, est de considérer une sonie "d'ensemble" c'est-à-dire une impression globale que nous a donnée le son et ce, en réitérant la procédure plusieurs fois pour éviter d'éventuels effets de fatigue auditive et en utilisant différentes références parmi les sons étudiés. Pour simplifier la tâche sachant que dans notre cas, elle s'applique à une centaine de sons, une interface graphique a été programmée (cf. figure 4.10). Le programme permet de sélectionner un dossier et de traiter automatiquement les sons au format ".wav" qu'il contient. On règle alors le gain pour être au même niveau qu'une référence choisie parmi les sons du dossier et qui peut être changée à tout moment pendant la procédure de comparaison. Une fois tous les gains réglés, le programme considère deux cas :
- Tous les gains sont négatifs ou nuls ; dans ce cas les fichiers ayant un gain différent de zéro sont modifiés et sauvegardés dans un sous dossier.
- Il existe au moins un gain positif ; dans ce cas, le programme considère que les fichiers ont tous été normalisés et qu'il faut mettre à 0db [6] le fichier qui a le plus grand gain et abaisser tous les autres fichiers de ce gain, ce qui permet d'éviter tout problème de saturation.

Fenêtrage

Une bonne partie des sons du corpus a été obtenue par découpage de sons de durée importante (parfois de l'ordre de la minute) ce qui implique nécessairement un fenêtrage. Il est apparu que l'écoute change quand ce fenêtrage n'est pas adapté à l'évolution temporelle du son et semble non

5. http://www.genesis-acoustics.com/sonie_en_ligne-32.html
6. décibel "FullScale". 0dbFS correspond à l'utilisation de toute l'échelle permise par le taux de quantification du signal

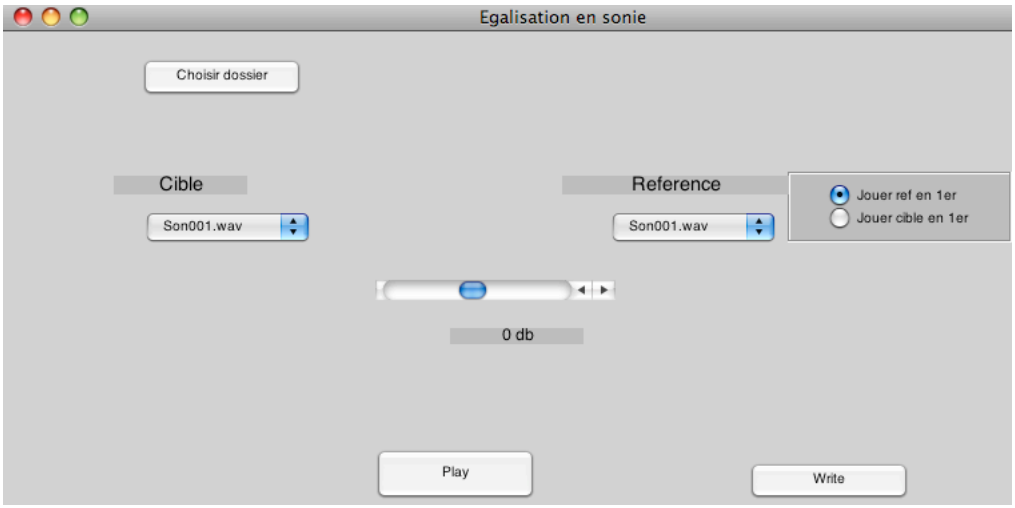


Figure 4.10 – Interface graphique Matlab pour égaliser des sons en sonie de manière subjective

"naturel". L'ajout de rampe de début et de fin identique pour chaque son, donne parfois l'impression que le son a été coupé (par celui qui l'a réalisé) et à l'écoute, cette transformation peut prendre le dessus sur un éventuel mouvement qui nous intéresse précisément.
La durée des rampes de début et de fin (de 4ms à 200ms) ainsi que le type de rampe (linéaire - exponentielle) a donc été adapté à chaque son en fonction de ses caractéristiques (attaque, présence de modulations d'amplitude,*etc.*).

4.3.2 Corpus final

Les 150 sons sélectionnés ont été soumis à une dernière étape de pré-test sur 4 sujets collaborant à l'étude. Ce pré-test est identique au test 4 (*cf.* section 4.4.4) mais avec un corpus plus grand qu'il serait impossible de faire évaluer aux sujets, le but étant d'obtenir un nombre optimal de stimuli qui se répartissent le plus uniformément possible dans l'espace des paramètres (du test). En effet, en effectuant le test sur plus de sons que nécessaire, on peut ne sélectionner qu'un sous ensemble de sons qui ont obtenus des réponses les plus différentes lors de ce pré-test. La durée globale a également été prise en compte pour faire en sorte que le test final ne dépasse pas les 3 séances d'une heure. Le corpus obtenu au final est constitué de 40 sons dont les durées varient entre 0,54 et 5,56 secondes (2,2 secondes en moyenne).

4.4 Protocole

Le protocole que nous avons mis en place est en fait constitué de deux tests (Test 3 et 4). Le test 3 s'est imposé suite à l'étude de la littérature qui a montré l'existence de spécificité perceptive quand il s'agit de mouvement. Ce test est donc basé sur un questionnaire informatique et vise à caractériser précisément le corpus notamment en étudiant si les sujets se sentent impliqués dans la scène auditive que représentent ces sons.
Le test 4 est le test principal où les sujets ont décrit les mouvements évoqués par les sons du corpus à l'aide de l'interface présentée section 4.2. Les résultats du test 3 n'étant pas nécessaire à la conduite du second, les protocoles et les résultats des deux tests seront présentés en parallèle.

4.4.1 Sujets

29 sujets ont participé au test : 10 femmes, 19 hommes de 21 à 53 ans (moyenne d'âge 29,8). Parmi les participants, 22 étaient musiciens et/ou experts en acoustique et audio. Seulement 2

sujets n'avaient jamais participé à un test d'écoute. Tous les sujets ont assuré utiliser régulièrement un ordinateur et 25 participants sont titulaires d'un diplôme bac+4 ou supérieur.

4.4.2 Conditions

Tous les stimuli ont été présentés au casque par écoute diotique avec un casque Stax 3R202, préampli Stax SRM310, en utilisant la carte son interne de l'ordinateur (Apple MacBook 13") dans une cabine audiométrique.
Le niveau d'écoute est laissé à l'appréciation de l'auditeur, mais il lui est indiqué qu'il ne peut le régler que pendant la phase d'entraînement et qu'il ne doit en aucun cas être modifié durant le test.
En plus de la valeur de chacun des paramètres de l'interface graphique, il est également enregistré pour chaque stimulus :
- Le temps de réponse
- Le nombre de fois qu'il a été écouté
- L'ordre de présentation

Les tests sont répartis en trois séances :
- Séance 1 : Test 3 + entraînement
- Séance 2 : Test 4 (première partie) + entraînements (parties 1 et 2 successivement)
- Séance 3 : Test 4 (deuxième partie)

Le détail des consignes et des questionnaires est donné en annexe B.

4.4.3 Test 3 : Caractérisation du corpus

Entraînement

Dans un premier temps, le sujet est invité à lire en détail les consignes qui sont affichées à l'écran. Chaque question est complétée par une phrase explicative (cf. annexes) avec à la fin, des recommandations générales (donner des réponses spontanées, ne pas passer trop de temps sur chaque son,...)
La phase d'entraînement est identique au test mais avec seulement trois sons différents des 40 précédemment sélectionnés. Deux de ces sons n'avaient pas été retenus dans la dernière étape de sélection pour cause de redondance avec d'autres sons, le dernier avait été rejeté à cause de sa trop grande complexité et fourni un exemple "extrême" de ce que vont rencontrer les sujets dans la suite du test.

Test

Une fois la phase d'entraînement terminée, les sujets lancent eux-mêmes le test et peuvent faire des pauses quand ils le souhaitent (cf. figure 4.11).
A la fin du test, ils doivent compléter par écrit un deuxième questionnaire en deux parties. La première partie concerne les renseignements personnels (âge,..) la seconde porte sur le test et les éventuelles difficultés rencontrées.

4.4.4 Test 4 : Caractérisation graphique

Entraînement

Etant donné la complexité du protocole, il a été nécessaire de passer par une étape où l'auditeur peut se familiariser avec l'interface sans autre contrainte que celle de comprendre à quoi correspond chaque paramètre. Les auditeurs doivent pouvoir faire la correspondance entre paramètres de contrôle et visualisation. Les paramètres de contrôle de l'interface ont été séparés en deux catégories "Formes

de la trajectoire" et "Dynamique de la trajectoire" et sont présentés séparément à l'utilisateur, dans deux phases d'entraînement où il est demandé de reproduire respectivement des formes et des dynamiques prédéfinies. Les interfaces graphiques de ces deux entraînements sont présentées dans les figures 4.12 et 4.13. A la fin de ces entraînements, les sujets sont donc en mesure de reproduire les trajectoires qu'ils imaginent. A noter que durant ces entraînements, aucun son n'est donné à entendre.

Test

Pour des raisons de durée et de concentration nécessaire pour compléter le test, celui-ci a été découpé en deux parties durant lesquelles 20 sons sont évalués. La première fois que le test est lancé, un tirage aléatoire est fait pour déterminer l'ordre de présentation des sons et quand le test est relancé pour la deuxième série, cet ordre est conservé ce qui permet éventuellement de revenir sur les sons évalués lors de la séance précédente. Si le sujet estime qu'il ne peut plus continuer faute de concentration, il est possible d'interrompre et de reprendre le test à tout moment.

Auto-évaluation

Pour chaque son, le sujet doit évaluer la qualité de ses réponses à l'aide de deux curseurs linéaires allants de “Pas du tout satisfait” à “Totalement satisfait”. Le premier curseur concerne la partie des contrôles de forme, le second concernant la dynamique. Dans les cas où le sujet n'est pas satisfait de sa réponse (le curseur est plus proche de “Pas du tout satisfait”), il est également demandé de préciser, si cette non-satisfaction est due au son, à l'interface ou aux deux (il est également possible de ne pas répondre).

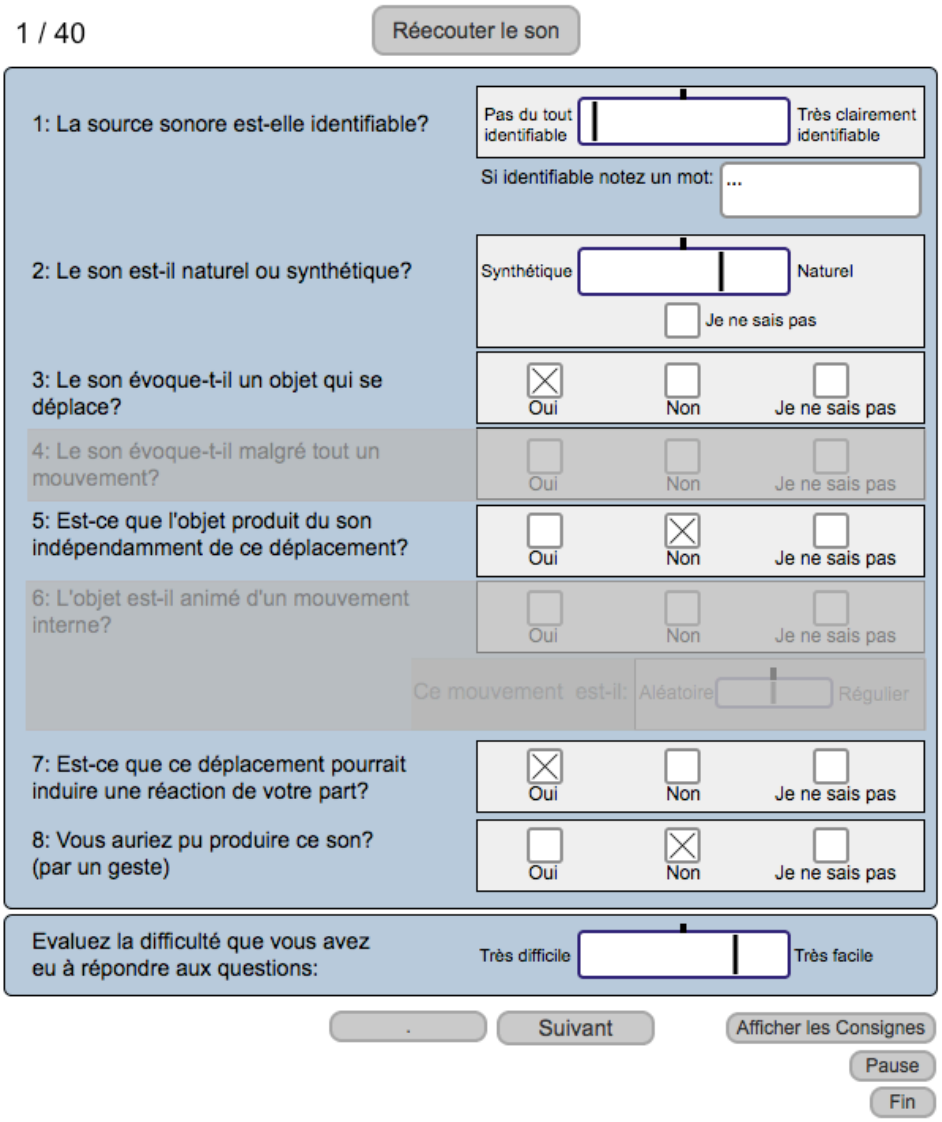


Figure 4.11 – Interface graphique pour le test 3. Les questions grisées sont celles auxquelles il n'est pas possible de répondre en fonction des réponses aux questions précédentes. Dans la partie navigation en bas à droite de l'écran, on voit que le sujet peut faire une pause (ces pauses sont soustraites des temps de réponse), revenir sur un son ou passer au suivant.

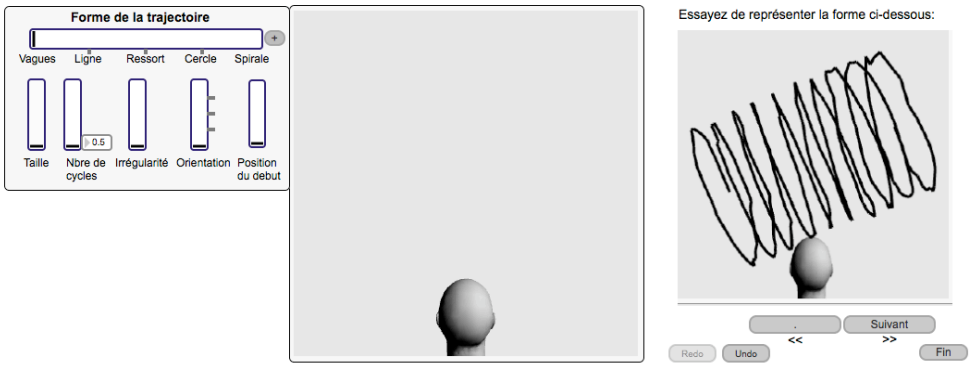


Figure 4.12 – Interface graphique de la première phase d'entraînement qui concerne les paramètres permettant de dessiner la forme sur l'interface. Les sujets doivent reproduire les formes proposées sur la parties droite de l'écran.

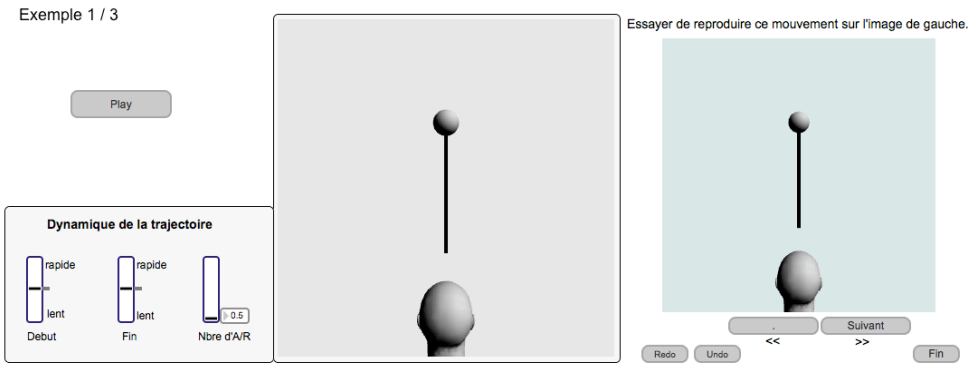


Figure 4.13 – Interface graphique de la seconde phase d'entraînement qui concerne la dynamique. Sur une trajectoire qui est à chaque fois verticale, les sujets doivent régler les trois curseur pour obtenir la même loi de déplacement de la boule au cours du temps que sur l'animation de la partie droite de l'écran.

4.5 Résultats

Etant donné le nombre de sujets inférieur à 30 et le nombre de sons analysés également inférieur à 30, les analyses de variance ont été réalisées en utilisant le test non-paramétrique de Kruskall-Wallis avec un seuil de significativité à 0.01. Même dans les cas où l'on considère les 40 sons simultanément les réponses ne remplissent généralement pas la condition de normalité permettant de faire des analyses de variance paramétriques. Les analyses sont reportées sous la formes $\chi^2(ddl) =$ où ddl est le nombre de degrés de liberté des sources de variation.
Les numéros qui ont été attribués aux sons l'ont été de manière totalement arbitraire, en revanche les sons 36 à 40 correspondent respectivement aux sons 68, 29, 27, 21 et 11 du chapitre 3. De même les numéros des sujets sont arbitraires mais correspondent à l'ordre dans lequel les sujets se sont présentés au test 3.

4.5.1 Test 3

Durée et difficultés

Le test a duré en moyenne 32 minutes (écart-type 10 minutes) chaque son étant écouté 5 fois en moyenne. L'analyse de variance sur les temps de réponse ($\chi^2(39, 1120) = 70.29$, $p < 0.01$) indiquent que certains sons ont pris plus de temps que la moyenne pour être évalués (les sons 5, 22 et 31) et d'autres moins de temps (les sons 28, 29 et 33). En revanche, en classant les sons par ordre de présentation, l'analyse de variance ($\chi^2(39, 1120) = 32.28$, $p = 0.76$) ne révèle aucun effet d'habituation ou de fatigue. Le test n'a pas été jugé trop long (par 77% des sujets) ni trop difficile (61%). Ce dernier résultat est confirmé par l'évaluation de la difficulté (que les sujets ont complété pour chaque son) qui est en moyenne plus proche de "Très facile". En revanche, 48% ont indiqué qu'une question leur avait semblé particulièrement difficile et en majorité il s'agit de la question 6 ("Le son est-il animé d'un mouvement interne ?").

Sons qui "évoquent un mouvement"

Il a été nécessaire d'identifier les sons qui "évoquent du mouvement" et qui seuls doivent être considérés dans la suite des analyses. 18 sons ont obtenu plus de 50% de réponses "oui" à la question 3 "Le son évoque-t-il un objet qui se déplace ?" et 4 plus de 75% (*cf.* figure 4.14). Nous avons également pris en compte les réponses positives à la question 4 "Le son évoque-t-il malgré tout un mouvement ?" question pour laquelle il n'était possible de répondre que suite à une réponse négative à la question précédente. Pour cela, nous avons sélectionné les sons qui ont obtenu plus de 21 réponses positives sur l'ensemble des questions 3 et 4 (en vert sur la figure 4.14). On obtient ainsi 29 sons qui "évoquent un mouvement" à plus de 75% des sujets. Toutes les analyses et notamment celles qui font suite au test 4 seront effectuées sur ces 29 sons.
A noter que ces sons sont cohérents avec la définition des sons abstraits dans le sens où ils ont tous obtenu un score inférieur à la moyenne à la première question "La source sonore est-elle identifiable ?". On notera enfin que les sons ont majoritairement été jugés "synthétiques" et notamment ceux qui ont été jugés les moins facilement identifiables [7]

Détails sur les mouvements

Concernant les questions 5 à 8, les réponses sont globalement moins consensuelles mais il existe systématiquement un ou plusieurs sons qui ont obtenu une majorité de réponses positives (resp. négatives).
Pour la question 5 ("Est-ce que l'objet produit du son indépendamment de son déplacement ?"), il n'était possible de répondre à cette question que si le son évoquait un objet qui se déplace (question

7. Ce résultat sera discuté en détail ultérieurement (section 6.1)

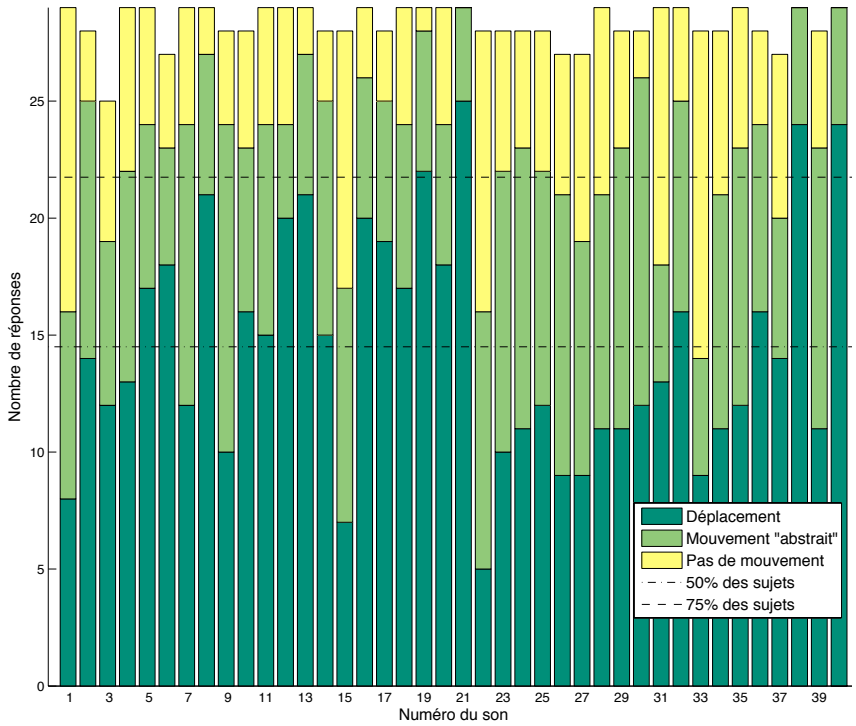


Figure 4.14 – Nombre de réponses à la question 3 “Le son évoque-t-il un objet qui se déplace ?” et à la question 4 “Le son évoque-t-il malgré tout un mouvement ?”

3) et malgré cette contrainte, 3 sons (les sons 8, 19 et 38) ont obtenu plus de 50% de oui et surtout 4 sons (les sons 5,13, 16 et 17) ont obtenu plus de 50 % de non. Pour ces 4 derniers sons, on peut estimer qu'ils correspondent à un cas où il existe un lien particulier entre le déplacement d'un objet et le son produit, il sera donc nécessaire de prendre en compte cette particularité lors de l'étape d'analyse. En effet, on peut supposer qu'il existe de fortes relations entre les paramètres acoustiques de ses sons et les caractéristiques du mouvement qui leur a donné naissance.
La question 6 (“Le son évoque-t-il un objet animé d'un mouvement interne ?”) a permis de dégager 6 sons avec plus de 50% de réponses “oui” (les sons 7, 8, 9, 10, 30 et 36) qui tous font partie des sons “qui évoquent un mouvement”. Cette différence entre mouvement interne et mouvement globale sera prise en compte dans les analyses qui suivent.
La question 7 (“Vous auriez pu produire ce son (par un geste) ?”) n'a fait ressortir qu'un seul son (le son 5) avec plus de 50% de réponses positives et seulement trois sons (4, 23 et 24) n'ont pas obtenu plus de 50% de réponses “non”. Cette question a été posée dans le but de prendre en compte d'éventuels effets perceptifs dus à la présence de composantes liées à la motricité dans les sons (les sujets s'imaginent en train de réaliser le geste qui produit le son et ont donc une perception différente de lorsqu'ils sont spectateur d'une scène sonore). Ce résultat indique donc que globalement, les sons du corpus ne mettent que peu ou pas en jeu de tels spécificités perceptives.
Enfin pour la question 8 (“Ce son aurait pu induire une réaction de votre part ?”) seul le son 19 a obtenu une réponse positive (à 69% des sujets) et 3 sons ont obtenu un consensus en faveur d'une réponse négative. Comme pour la question précédente, on s'intéresse ici à des particularités de la perception quand le sujet est très impliqué dans la scène sonore. On peut également conclure que ces particularités seront globalement limitées même si on se base sur des réponses à faible consensus (ceci peut probablement être expliqué par le fait que l'exemple fourni dans les consignes se rapporte à un objet qui s'approche de l'auditeur).

4.5.2 Test 4

Durée et difficultés

Le test (entraînements non compris) a duré en moyenne 86 minutes avec des disparités importantes entre les sujets, les durées totales allant de 39 à 199 minutes. Chaque son a été écouté en moyenne 14 fois (jusqu'à 25 fois pour certains sons).
Dans le questionnaire de fin, 79% des sujets ont rapporté que le test avait une durée raisonnable et 62% qu'il n'était pas trop difficile. De même que pour le test 3 l'analyse de variance sur les réponses groupées par ordre de présentation ($\chi^2(39, 1120) = 47,64$, $p = 0.16$) n'a pas mis en évidence d'effet de fatigue et l'analyse de variance par son ($\chi^2(39, 1120) = 151,73$, $p < 0.01$) a révélé des différences significatives (les temps de réponses sont significativement plus longs pour certains sons).
82% des sujets ont trouvé que les possibilités de contrôle étaient trop limitées. En particulier, les contrôles de la dynamique ont été jugés moins pertinents, certains sujets ont même précisé qu'ils auraient souhaité pouvoir modifier point par point la courbe de vitesse en fonction du temps. Ce résultat est d'ailleurs confirmé par l'auto-évaluation réalisée durant le test, qui indique que 30% des réponses (pour tous les sons et tous les sujets) de dynamique ont été jugées non satisfaisantes et ce, majoritairement à cause de l'interface graphique. 76% des sujets ont jugé que le premier test n'avait pas influencé leurs réponses et 62 % que leur stratégie de réponse n'avait pas été modifiée au cours du test. On notera enfin que la grande majorité des sujets se sont montrés très intéressés par l'étude et ont demandé à être tenus au courant des résultats.

Comme on peut le voir sur la figure 4.15, les dessins obtenus sont relativement différents d'un sujet à l'autre et à première vue, il semble assez difficile de dégager des points communs entre les réponses. Les sections qui suivent reprennent donc en détail les réponses obtenues pour chaque paramètre pour mieux mettre en évidence des caractéristiques communes à une majorité de sujets.

Sélection des réponses

Dans les analyses qui suivent, seules les réponses qui ont été jugées satisfaisantes par les sujets ont été prises en compte. Concrètement, les réponses qui ont obtenu une note inférieure à la moyenne, dans la partie auto-évaluation (réalisée par chaque sujet pour chaque son) sont considérées comme des valeurs manquantes. Dans cette sélection les contrôles de formes et de dynamiques ont été traités séparément. Pour un son et un sujet donné, on peut très bien n'utiliser que les valeurs obtenues pour les trois curseurs de dynamique (ou inversement uniquement les valeurs obtenues pour les contrôles de formes) en fonction des deux notes d'auto-évaluation données par le sujet.
L'analyse de variance des notes d'auto-évaluation a révélé des différences significatives de stratégies entre les sujets ce qui nous a incité à procéder à une mise à l'échelle des notes. En pratique, chaque note a été divisée par la différence entre la moyenne du sujet et la moyenne obtenue par le son. Nous avons ensuite sélectionné les réponses (individuelles) qui sont supérieures à 10, note qui correspond à la valeur initiale du curseur de réponse ainsi qu'à la position centrale de l'échelle de réponse. Des 841 réponses enregistrées (29 sujets $\times$ 29 sons "qui évoquent un mouvement"), nous n'avons donc pris en compte que 708 réponses pour les variables de formes et 678 pour les variables de dynamique (630 réponses sont retenues pour les 2 types de variables).

4.5.3 Analyse des trajectoires

Etant donnée la non-linéarité des relations entre les paramètres de l'interface et les trajectoires affichées, il a été nécessaire de redéfinir certaines variables. En effet, différentes combinaisons des 9 variables de l'interface peuvent conduire au même dessin comme par exemple en inversant les valeurs d'orientation et de position initiale (phase à l'origine) d'une trajectoire qui est une demi-période de sinusoïde. Dans un premier temps, une définition de la variable à partir des paramètres de l'interface

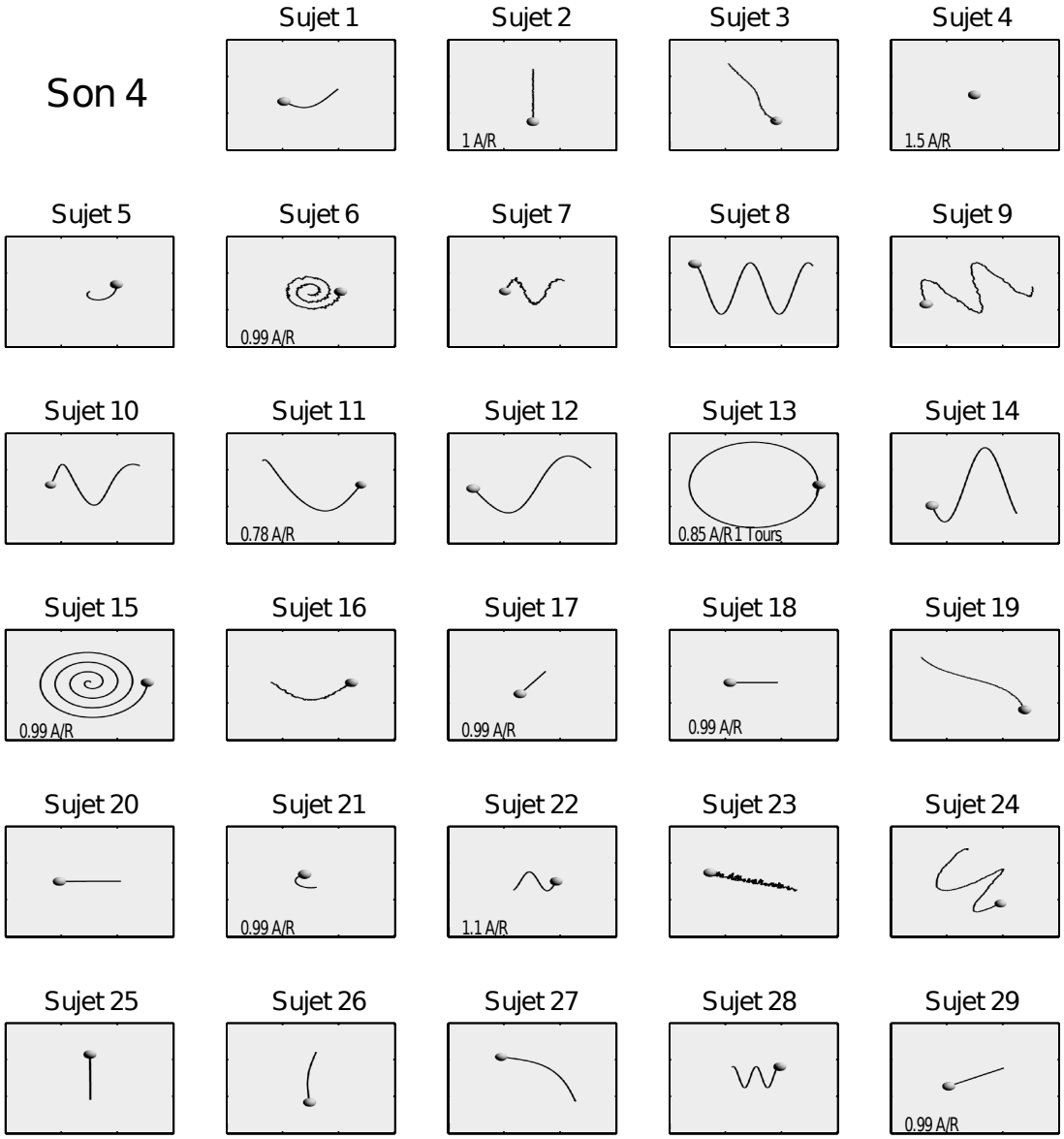


Figure 4.15 – Exemple de dessins réalisés par les sujets pour le son 4.

est donnée, puis l'analyse de la cohérence des réponses entre les sujets est faite. Ensuite on procède à l'évaluation du degré de précision pour enfin sélectionner les sons pertinents au regard de la variable.
Il faut rappeler que le nombre de réponses qui sont prises en compte est différent d'un son à l'autre du fait de la sélection en fonction de l'auto-évaluation décrite dans la section précédente. En revanche, les pourcentages sont calculés à partir du nombre de réponses totale ce qui explique les faibles valeurs qui sont parfois observées.

Forme

Les formes obtenues dépendent non seulement du paramètre du même nom, mais aussi des paramètres *Nombre de cycles*, *Orientation* et *Nombre d'A/R*. Par exemple, on définira les formes *Vagues* et *Ressort* uniquement pour un *Nombre de cycles* supérieur à 1. Suivant cette logique, en plus des 5 formes initiales (*Ligne, Vague, Ressort, Cercle* et *Spirale*), 3 nouvelles formes ont été définie : *Creux* et *Dome* qui correspondent à des vagues ou des ressort pour des nombres de cycles inférieurs à 1 et à des orientations adéquates. Enfin on définit la forme *Rebond* qui correspond à

une ligne droite verticale avec au moins 1 aller/retour (*cf.* figure 4.16).

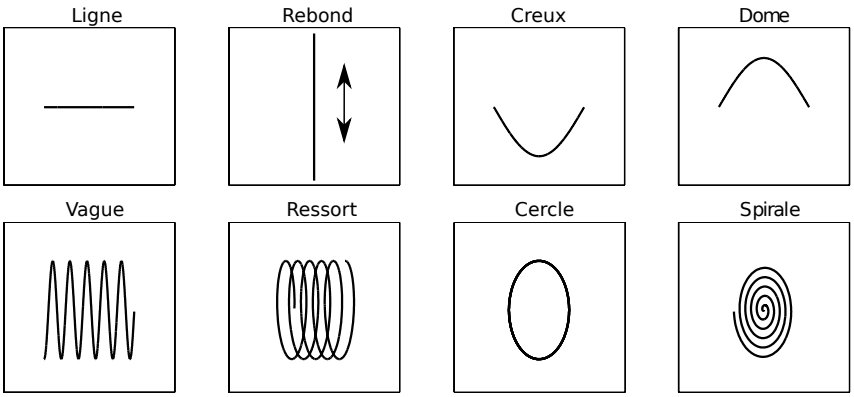


Figure 4.16 – Exemples typiques des 8 formes élémentaires définies pour l'analyse.

Nous avons ensuite procédé à une analyse factorielle (estimation à partir du maximum de vraisemblance et avec rotation type "varimax") sur le pourcentage de réponses obtenues pour chacune des ces 8 formes et pour chaque son. Le test de sphéricité de Bartlett étant significatif ($p < 0.01$) et en considérant les valeurs propres supérieures à 1, nous avons retenu les deux premiers facteurs expliquant 47 % de la variance totale. A noter que le troisième facteur n'a pas été retenu car bien qu'ayant une valeur propre légèrement supérieure à 1, il correspondait au point d'inflexion quand on représente les valeurs propres par ordre décroissant (en anglais "scree plot"). Cette analyse nous permet donc de faire des regroupements et surtout de mettre en évidence les grandes formes caractéristiques (*cf.* figure 4.17). Le premier axe sépare les formes linéaires des

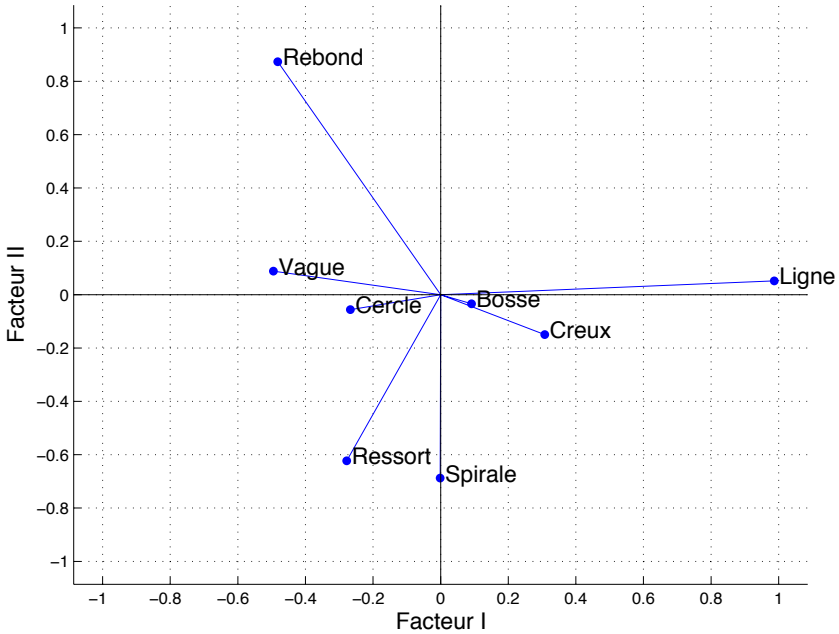


Figure 4.17 – Facteurs 1 et 2 de l'analyse factorielle réalisée sur les pourcentages de réponses obtenues pour chaque forme. Ces deux facteurs expliquent 47% de la variance totale.

formes oscillantes tandis que le second sépare le type d'oscillation (linéaire ou circulaire). A noter qu'une partie non négligeable de la variance n'étant pas expliquée par les deux facteurs, ils doivent être interprétés comme reflétant des "tendances" et non comme définissant des catégories précises.

Sur la base de ces “macro-catégories”, on peut donc regrouper les réponses et sélectionner les sons pour lesquels un type global de forme émerge pour plus de la moitié des sujets. Les sons 24, 29 et 32 ont majoritairement obtenu des trajectoires linéaires. Les sons 5 et 35 sont majoritairement des oscillations linéaires. Les sons 2, 8, 13, 25, 36 et 39 sont majoritairement des oscillations circulaires. On peut donc conclure que 3 catégories de formes semblent pertinentes d'un point de vue perceptif : linéaire, oscillation régulière et oscillation circulaire.

Fréquence d'oscillation

Comme on l'a vu dans le paragraphe précédent, les trajectoires oscillantes se distinguent des autres formes (elles concernent 21 sons) ce qui pousse à vouloir quantifier ces oscillations. L'interface permettaient aux utilisateurs de contrôler cet aspect par deux moyens différents : le *Nombre de cycles* qui représente la fréquence spatiale de la trajectoire et le *Nombre d'A/R* qui représente le nombre de fois que la trajectoire est parcourue par la sphère. Dans beaucoup de cas, les sujets ne s'accordent pas dans l'utilisation de l'une ou l'autre des stratégies ce qui nous a incité à définir une nouvelle variable. Cette variable appelée *Fréquence* est le produit du *Nombre de cycles* par le *Nombre d'A/R* divisé par la durée du son (et qui est donc bien homogène à une fréquence).

Le premier résultat concerne la variabilité inter-sujets, mesurée par la distance inter-quartil (IQR), qui est globalement très élevée. Ce résultat est relativement étonnant puisque compter des oscillations semble à première vue, être une tâche facile, dans la limite ou ce nombre d'oscillations est faible. Plus précisément, l'IQR augmente linéairement avec la médiane (les deux sont corrélés à 78%) jusqu'à 8Hz (*cf.* figure 4.18) et on observe des valeurs bien supérieures au-dessus de cette fréquence. La variabilité est donc importante même pour les basses *Fréquence* et il semble qu'au-delà de 8Hz, il devienne difficile de parler d'oscillation au sens de phénomène périodique quantifiable. Il est intéressant de constater que les sons dont la *Fréquence* mediane est supérieure à 8Hz, ont également obtenu des valeurs d'*Irrégularité* importantes, ce qui supporte l'hypothèse d'un traitement différent au-delà de ce seuil. On notera au passage que la variabilité inter-sujets est encore plus importante quand on considère les variables *Nombre de cycles* et *Nombre d'A/R* séparément.

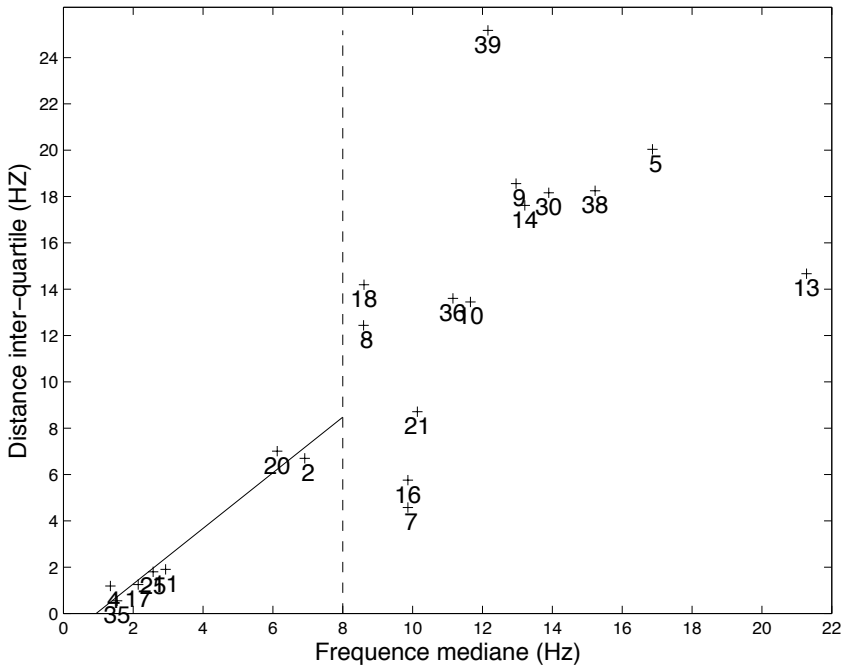


Figure 4.18 – Distance inter-quartile en fonction de *Fréquence* mediane.

Direction

Le paramètre *Orientation* de l'interface permet de faire tourner à 360 ° la trajectoire, et ce, dans le plan $x - z$. Pour les formes *Cercle* et *Spirale*, cette rotation s'effectue autour de l'axe x (puisqu'une rotation dans le plan $x - y$ n'a aucun effet). Du fait de leur propriété de symétrie, dans un premier temps, on ne considère donc pas de direction privilégiée pour ces deux formes.
Pour les formes *Dome* et *Creux*, la direction est également dépendante du paramètre *Position du début* qui contrôle la phase à l'origine des cosinus et sinus (dans l'équation de l'hélice). Dans ces cas, la direction est donc calculée en utilisant les coordonnées des extrémités de la trajectoire et non directement à partir du paramètre *Orientation*.
Dans les cas où les sujets ont utilisé un *Nombre d'A/R* supérieur à 0,5 il n'est plus possible de définir une direction, celle-ci ne sera donc pas prise en compte dans les analyses qui suivent.
Pour s'affranchir de l'indétermination $-\pi/+\pi$, les valeurs d'angle ont été ramenées aux huit grandes directions de la rose des vents les plus proches. L'analyse factorielle sur ces directions (même procédure que pour *Forme*) montre que la distinction principale se fait entre les trajectoires verticales et les trajectoires horizontales sur le premier facteur (*cf.* figure 4.19). Sur ce même axe, on peut voir que les directions droite et gauche (E et O) se retrouvent groupées ce qui concorde avec notre situation d'écoute diotique qui ne fournit aucun indice acoustique permettant de distinguer ces deux directions. Bien que l'analyse factorielle fasse ressortir un troisième facteur, son interprétation en est relativement incertaine puisqu'il est principalement supporté par les directions SE et SO. On peut cependant avancer l'hypothèse de l'existence d'une distinction supplémentaire des trajectoires inclinées par rapport aux autres trajectoires verticales.
On peut donc conclure que 3 directions semblent pertinentes d'un point de vue perceptif : "Monte", "Descend" et "Horizontal".

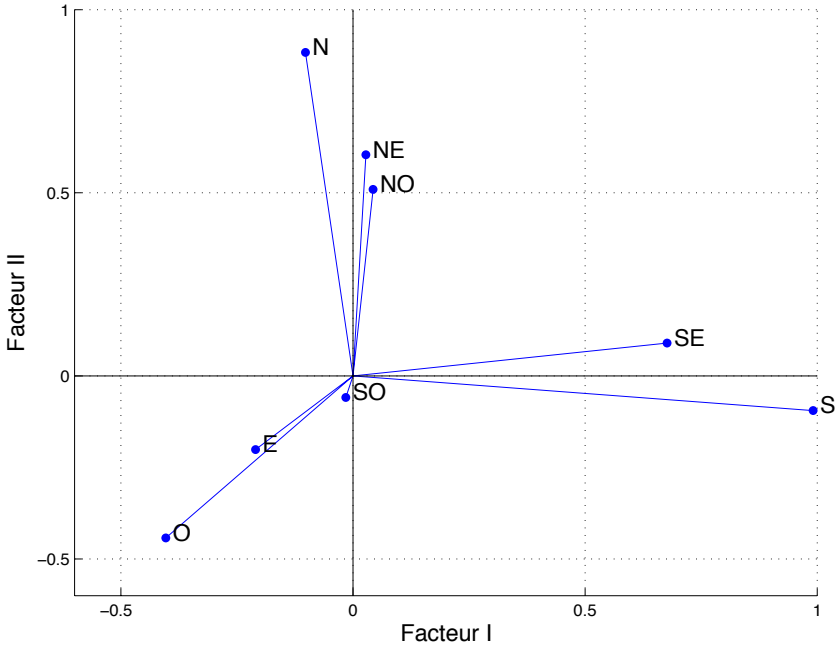


Figure 4.19 – Facteurs I et II de l'analyse en factorielle sur les pourcentages de réponse pour chacune des 8 directions de la rose des vents.

Taille

La *Taille* est une des rares variables qui a été analysée directement à partir du paramètre de contrôle de l'interface, celui-ci ayant un effet sans équivoque sur l'espace occupé par la trajectoire

dans la fenêtre. Une taille de 1 correspond à une trajectoire qui occupe toute la surface de la fenêtre tandis qu'une taille de 0 correspond à une trajectoire qui en occupe moins de 10%.
Tout d'abord, nous avons constaté des différences globales entre les réponses données par différents sujets, certains ayant tendance à produire des trajectoires qui sont en moyenne plus petites (resp. plus grandes) que d'autres. L'analyse de variance des réponses par sujets est en effet significative ($\chi^2(28, 679) = 102.24$, $p < 0.01$). Nous avons donc souhaité vérifier si au-delà de ces différences individuelles, il était possible de dégager des différences au niveau des sons ce qui nous a incité à effectuer une normalisation par rapport à la médiane des réponses de chaque sujet. En pratique pour chaque sujet, ses réponses sont divisées par leur écart avec la médiane des réponses de tous les sujets. Cette normalisation permet de faire baisser la distance inter-quartile moyenne des sons de 7,1% (de $0,28$ à $0,26$) et surtout de ne plus avoir de différences significatives entre les réponses par sujet ($\chi^2(28, 679) = 9.71$, $p = 1$). Nous avons ensuite procédé à une analyse en cluster hiérarchique avec des distances "complètes" (*cf.* figure 4.20) sur la médiane des tailles obtenues par chaque son. En considérant un seuil arbitraire fixé à 0.2, les sons se regroupent suivant trois niveaux de tailles. L'analyse de variance sur les réponses groupées suivant ces 3 tailles ($\chi^2(2, 705) = 122.81$, $p < 0.01$) est significative et les comparaisons multiples (avec correction de Bonferroni) effectuées à la suite de l'analyse de variance montrent que les sujets s'accordent pour distinguer chacune de ces trois tailles. A noter que si on choisit un seuil plus faible dans l'analyse en cluster permettant d'obtenir 4 groupes, les différences de tailles entre les groupes ne sont plus significatives (dans le test "post-hoc").

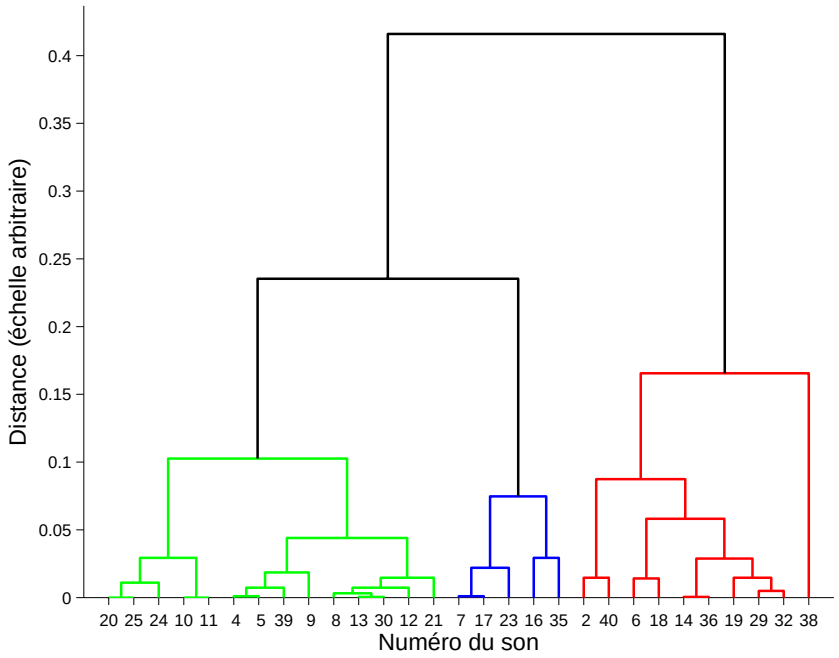


Figure 4.20 – Dendrogramme obtenu par analyse en cluster hiérarchique pour la variable *Taille* (mediane des réponses pour chaque son). A gauche se trouvent les sons correspondant à des trajectoires de taille moyenne, au milieu les trajectoires de petite taille et à droite les trajectoires de grande taille. A noter que les réponses de chaque sujet ont été normalisées pour réduire l'aspect subjectif de cette variable.

Irrégularité de la forme

Tout comme la variable *Taille*, la variable *Irrégularité* est directement issue du paramètre de l'interface. En revanche nous n'avons pas procédé à une normalisation des réponses puisque même

si l'analyse de variance par sujet est significative ($\chi^2(28, 679) = 54$, $p < 0.01$), les comparaisons multiples qui ont suivi montrent que cette différence n'est due qu'à un seul sujet et l'analyse de variance n'est plus significative si on retire les réponses de ce sujet ($\chi^2(27, 653) = 25.5$, $p = 0.54$). Pour quelques sons ayant une irrégularité proche de 0, on observe un consensus presque total entre les sujets. De même que pour la variable *Taille*, nous avons procédé à une analyse en cluster suivie d'une analyse de variance ($\chi^2(2, 705) = 198.13$, $p < 0.01$) avec comparaisons multiples. Nous avons mis en évidence trois niveaux d'irrégularité qui distinguent significativement les réponses entre les sons (*cf.* figure 4.21). Pour les sons qui correspondent aux trajectoires les plus irrégulières, on constate qu'ils ont tous obtenu une majorité de réponses positives à la question 6 du test 3 ("L'objet qui a produit le son est-il animé d'un mouvement interne ?"). On peut donc conclure que la variable *Irrégularité* traduit en partie cette notion de mouvement interne.
Du point de vue du contrôle, il est important de constater que les sujets n'ont utilisé qu'une partie de l'échelle de valeurs permises par cette variable, 95% de toutes les réponses se trouvant en dessous de 0.5 (valeur qui correspond à des variations aléatoires dont l'ordre de grandeur est la moitié de la taille de la trajectoire).

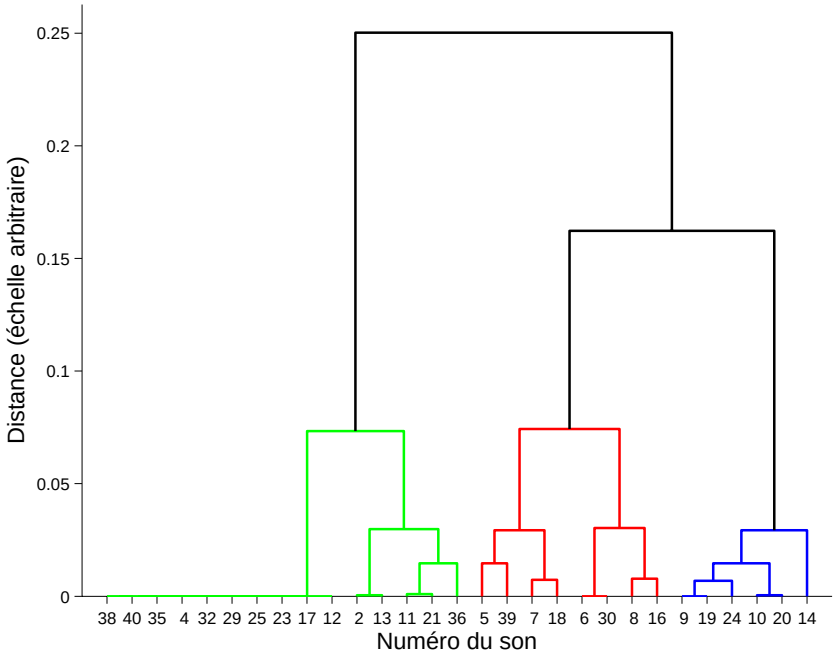


Figure 4.21 – Dendrogramme issu de l'analyse en cluster hiérarchique des réponses obtenues pour la variable *Irrégularité*. A gauche se trouvent les sons correspondant à des trajectoires parfaitement régulières, au milieu les trajectoires faiblement irrégulières et à droite les trajectoires très irrégulières.

Dynamique

Dans le questionnaire et aux travers de l'auto évaluation, les contrôles de vitesse sont apparu comme étant les plus frustrants pour les sujets (*cf.* section 4.5.2). Ce problème se retrouve dans les réponses obtenues sur les variables *Vitesse de début* (V_d) et *Vitesse de fin* (V_f) pour lesquelles il est relativement difficile de dégager des résultats valables pour une majorité de sujets. Même en considérant seulement trois niveaux (lent, constant et rapide), très peu de sons obtiennent plus de 50% des réponses dans l'un de ces trois niveaux et ce, de façon identique pour les deux variables. Nous avons donc cherché à caractériser la vitesse de façon plus globale en introduisant une nouvelle variable appelée *Dynamique* (notée Dy) définie par :

$$Dy = |V_d - 0.5| + |V_f - 0.5|$$

on rappelle que $\{V_d, V_f\} \in [0, 1]$, où 0 correspond à une vitesses lente et 1 une vitesse rapide (0.5 pour une vitesse constante). Dy quantifie donc la déviation absolue de la vitesse par rapport à une trajectoire parcourue à vitesse constante. Nous avons ensuite procédé à une analyse en cluster (*cf.* figure 4.22) en suivant la même procédure que précédemment (classification ascendante hiérarchique avec distance du plus proche voisin puis essais de différents seuils pour constituer des groupes à partir du dendrogramme). L'analyse de variance ($\chi^2(2, 614) = 65$, $p < 0.01$) et les comparaisons multiples ont mis en évidence trois niveaux de *Dynamique* significativement différents, à savoir : vitesse constante, peu de variation et beaucoup de variations.
Du point de vue du contrôle, il semble donc important d'offrir aux utilisateurs la possibilité de modifier finement la vitesse et notamment les variations de vitesse au cours du temps. En effet les sujets se sont accordés sur trois niveaux de variations malgré le peu de satisfaction que leur a donné les contrôles auxquels ils avaient accès.

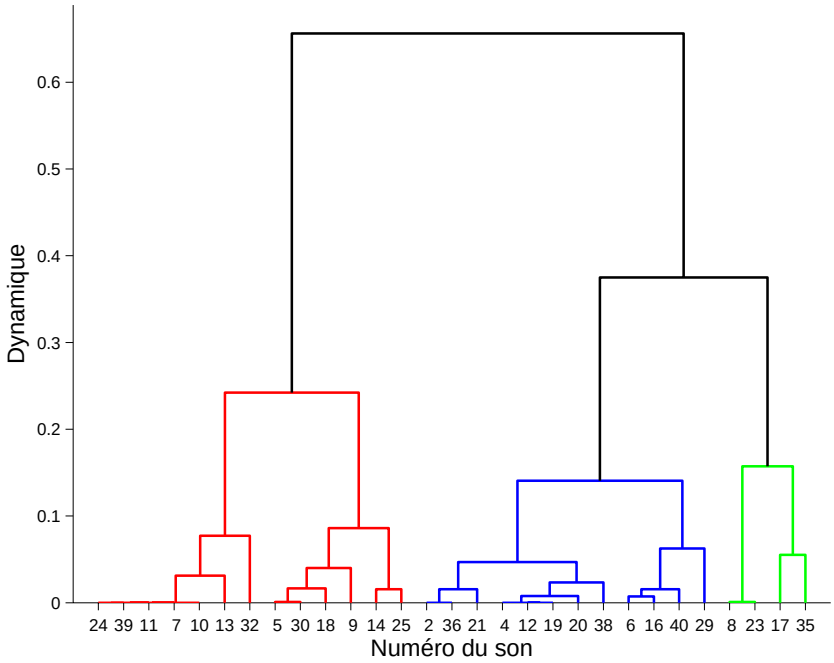


Figure 4.22 – Dendrogramme issu de l'analyse en cluster sur la variable *Dynamique* (définies à partir des vitesses de début et de fin). Les trois groupes correspondent respectivement aux trajectoires parcourues à vitesse constante (à gauche) avec peu (au milieu) et beaucoup (à droite) de variations de vitesse.

4.5.4 Résumé

Le tableau 4.1 représente l'ensemble des résultats du point de vue de chaque son. L'auto-évaluation des sujets ayant entraîné la non prise en compte d'une partie des réponses, certains résultats ne sont pas obtenus avec une majorité absolue (seuls les résultats en gras dans le tableau correspondent à une majorité absolue). Enfin, pour certains sons et certaines dimensions, il n'a pas été possible de dégager un résultat sans ambiguïté, certaines cases sont donc vides.

N° du son	Nb. rep. retenues (form / dyn)	Forme	Fréqu. oscil.	Direction	Taille	Irrég.	Vitesse
2	25 / 24	**Osc. Circ**	6,9	-	III	0	
4	25 / 24	-	1,3	**Horizontal**	II	0	
5	21 / 16	**Osc. Reg.**	16,9	-	II	+	Variable
6	24 / 25	-	-	Vertical	III	+	
7	25 / 22	Osc. Reg.	9,9	-	I	+	
8	24 / 27	**Osc. Circ**	8,6	Monte	II	+	Constante
9	22 / 19	Osc. Circ	13,0	-	II	++	Variable
10	23 / 24	Osc.	11,7	-	II	++	Constante
11	23 / 22	Osc.	2,9	-	II	0	Constante
12	24 / 26	-	-	-	II	0	-
13	26 / 24	**Osc. Circ**	21,3	-	II	0	-
14	27 / 27	Osc. Circ	13,2	**-**	**III**	**++**	Variable
16	19 / 18	Osc. Reg.	9,9	-	I	+	-
17	22 / 20	Osc.	2,1	**Horizontal**	I	**0**	-
18	26 / 26	-	8,6	-	**III**	+	Variable
19	23 / 24	Osc. Circ	-	**Horizontal**	III	**++**	-
20	22 / 26	Osc. Circ	6,1	-	II	++	Variable
21	27 / 26	Osc. Circ	10,1	**Horizontal**	II	0	-
23	21 / 16	Osc.	-	-	I	0	-
24	26 / 27	-	-	**Horizontal**	II	++	Constante
25	27 / 25	**Osc. Circ**	2,6	-	II	0	-
29	25 / 27	Line	-	**Descend**	III	0	-
30	27 / 27	-	13,9	Horizontal	II	+	Constante
32	24 / 26	**Ligne**	-	Vertical	III	**0**	-
35	23 / 17	**Osc. Reg.**	1,5	-	I	0	Variable
36	27 / 24	**Osc. Circ**	11,2	-	III	0	Variable
38	28 / 25	-	15,2	**Monte**	**III**	0	-
39	27 / 28	**Osc. Circ**	12,2	-	**II**	+	**Constante**
40	25 / 16	-	-	Horizontal	III	0	Variable

Tableau 4.1 – Récapitulatif des résultats par son.

La deuxième colonne correspond au nombre de sujets qui ont été satisfaits de leurs réponses et donc au nombre de réponses retenues pour les analyses. Pour les Formes, les abréviations correspondent à "Oscillation Régulière" et à "Oscillation circulaire". On notera les trois niveaux de *Taille* (I pour une petite taille, II pour une taille moyenne et III pour une grande taille) et les trois niveaux d'*Irrégularité* (0 pas d'irrégularité, + pour un peu d'irrégularité et ++ pour beaucoup d'irrégularité).

4.6 Conclusion

Le protocole développé dans ce chapitre est tout à fait original et comporte un certain nombre de défauts inhérents à cette originalité. En particulier, il n'existe aucune méthode standard pour l'analyse globale de telles données et nous avons conscience que les choix qui ont été faits peuvent et doivent être discutés. Dans le temps qui a été imparti à cette recherche, nous avons pris le parti de considérer chaque variable plutôt que de focaliser sur un nombre plus restreint de dimensions et ainsi aller plus loin dans le traitement des données.
L'effort qui a été fait pour réduire au maximum la complexité de l'interface a permis de faire évaluer 40 sons par 29 sujets mais n'a pas toujours donné la possibilité aux sujets de donner des réponses qui les ont pleinement satisfait. Toutefois, les résultats nous ont permis de caractériser les mouvements évoqués suivant un ensemble de dimensions qui semblent importante d'un point de vue perceptif. Une partie importante des données collectées n'a pas été exploitée, données pour lesquelles on espère qu'elles pourront servir à d'autres recherches et pas seulement des recherches qui concernent la communauté travaillant autour du son. Ce protocole remplit les objectifs fixés puisque dans une même étape, nous avons évalué l'interface graphique qui sera utilisée dans un futur synthétiseur (*cf.* chapitre 5) et collecté des sons qui représentent différentes valeurs des paramètres de contrôle. Nous avons également réussi à organiser de façon hiérarchique les différents attributs du mouvement en fonction de leur importance perceptive (et probablement de leurs occurrences dans le corpus de son évalué). Cette hiérarchisation doit maintenant être validée sur des stimuli parfaitement calibrés et c'est ce qui fait l'objet du chapitre suivant. Enfin ce test ouvre un certain nombre de perspectives pour l'analyse et la sonification de dessins qui seront abordés dans le chapitre 6.

Chapitre 5

Analyse par synthèse de mouvements

5.1 Analyse des signaux **86**
5.1.1 Analyse temps-fréquence 87
5.1.2 Test d'instationnarité 89
5.2 Moteur de synthèse temps réel **89**
5.2.1 Stratégie de contrôle 90
5.2.2 Calibration 92
5.3 Test de validation **92**
5.3.1 Stimuli 93
5.3.2 Sujets 94
5.3.3 Protocole 94
5.3.4 Résultats 95
5.4 Vers un contrôle générique du mouvement évoqué par les sons . . . **100**
5.5 Conclusion **101**

Comme pour le chapitre précédent, les travaux présentés ici sont en partie décrits dans [Merer et al., 11]. Le lecteur est invité à visiter la page web de l'article qui contient les sons et l'interface de test : `http://www.lma.cnrs-mrs.fr/~kronland/Movement/`

Les objectifs de ce chapitre sont multiples car il s'agit à la fois d'analyser les sons en fonction des résultats des précédents tests, de construire un synthétiseur de mouvement et d'évaluer les résultats obtenus au chapitre précédent. Pour cela, nous avons utilisé le paradigme d'analyse par synthèse [Risset et Wessel, 99] qui permet de mettre en évidence par des étapes successives d'analyse, de transformation, de synthèse et d'évaluation, les éléments qui sont pertinents pour un rendu perceptif donné. La priorité est donc de construire un modèle de synthèse relativement général qui permet de tester différentes transformations et éventuellement revenir sur sa conception, *a posteriori*.
L'analyse des trajectoires obtenues précédemment a mis en évidence un certains nombre de variables qui sont pertinentes pour décrire le mouvement évoqué par les sons. Il a été nécessaire dans cette dernière étape de se limiter à l'étude de trois variables à savoir *Forme*, *Direction* et *Fréquence*. En effet, nous souhaitons étudier des hypothèses déduites des précédents tests perceptifs, à l'aide de stimuli bien calibrés obtenus par synthèse ce qui limite forcément le nombre de facteurs qui peuvent raisonnablement être étudiés.

Dans ce chapitre, il est d'abord question d'analyse de signaux avant de présenter la conception d'un logiciel de synthèse temps réel utilisant la même interface graphique que le test du chapitre précédent. Ce synthétiseur a été utilisé pour générer les stimuli d'un test perceptif où les sujets ont évalué la cohérence entre des sons et des vidéos pour valider certains résultats obtenus au chapitre précédent et valider la stratégie de synthèse proposée.

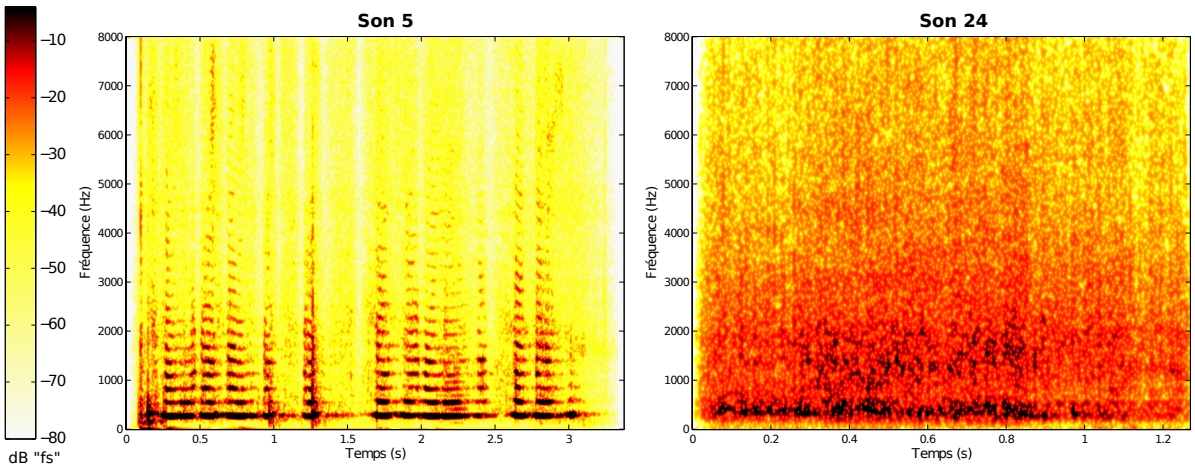
5.1 Analyse des signaux

Tout d'abord, nous tenons à préciser que cette section n'a pas pour but de fournir à elle seule les informations suffisantes pour mettre en place un modèle de synthèse complet. En effet, le travail qui vise à relier des caractéristiques du signal sonore à des dimensions perceptives pour finalement

contrôler ces caractéristiques dans un modèle de synthèse est un travail qui pourrait faire en soit, l'objet d'une thèse et dont nous avons montré qu'il peut difficilement se limiter à la mise en relation entre descripteurs de timbre et jugements perceptifs (*cf.* section 3.5).
Ces analyses sont donc menées principalement dans le but d'ouvrir des pistes pour de futures études plus poussées et de faire des hypothèses quant aux caractéristiques des sons qu'il faudra contrôler. Nous nous sommes limités à quelques sons qui ont globalement obtenu des résultats consensuels : les sons 5, 24, 29, 32, 38 et 39. Il est à noter que les données perceptives du test qui sont rappelées n'ont pas systématiquement obtenu une majorité absolue mais il a semblé qu'elles pouvaient avoir un intérêt au moins informel.

5.1.1 Analyse temps-fréquence

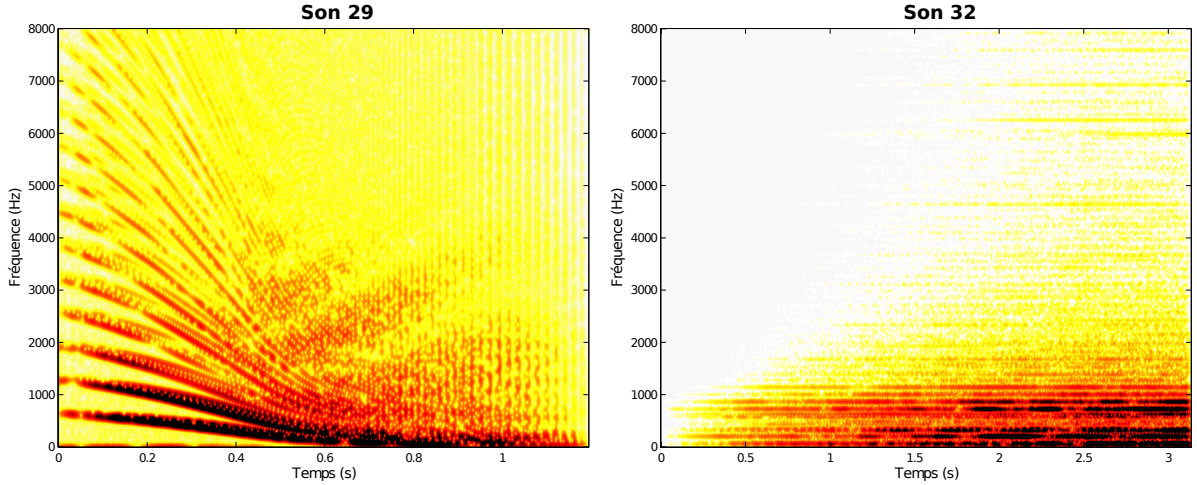
Dans ce qui suit, les représentations temps-fréquence sont des transformées de Gabor obtenues avec des fenêtres Gaussiennes périodisées de même taille que le signal analysé, pour 800 points en fréquence, et avec un pas de temps et un recouvrement temporel qui dépendent de la durée du signal [1]. L'amplitude est représentée suivant une échelle logarithmique où le noir correspond à $0dB_{FS}$ et le blanc à $-70dB_{FS}$.
Pour faciliter l'interprétation des images temps-fréquence, les résultats du test perceptif sont rappelés pour chaque son discuté.



Son 5 *Oscillation régulière rapide, orientation non-définie, taille moyenne, un peu d'irrégularité, vitesse peu variable*
L'image temps-fréquence montre une succession non régulière de motifs quasi-harmoniques qui varient en durée et dont les composantes varient légèrement en fréquence. On peut compter environ 12 impacts pour une durée de 3.5 secondes ce qui doit probablement être mis en relation avec le caractère "rapide" des oscillations.

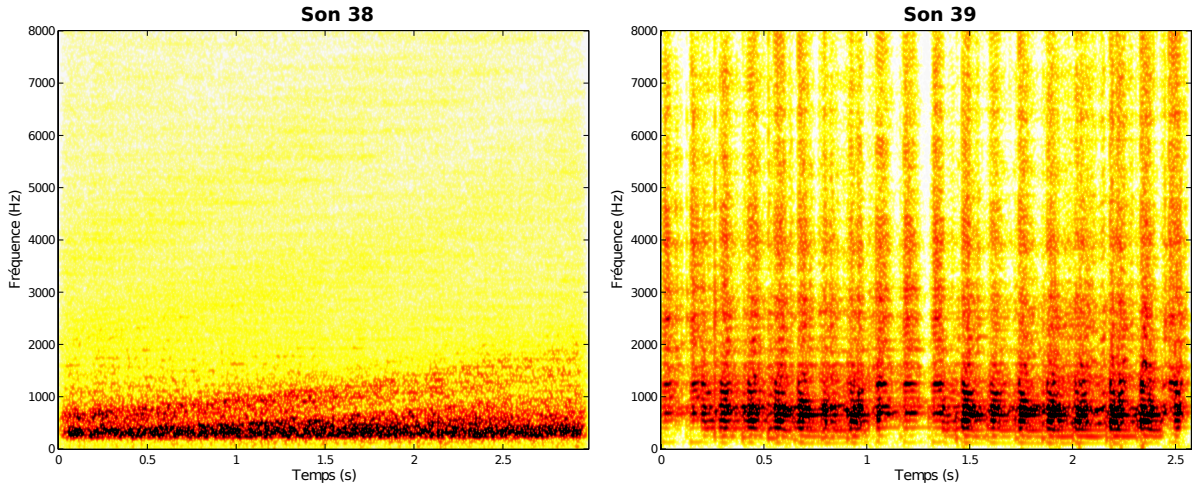
Son 24 *Forme non-définie, trajectoire horizontale, taille moyenne, beaucoup d'irrégularité, vitesse constante*
L'image temps-fréquence révèle un son globalement bruité avec de l'énergie principalement comprise entre 400Hz et 2KHz. On voit également une variation du contenu spectral haute fréquence qui peut s'assimiler à un filtrage passe bas dont la fréquence de coupure augmente progressivement jusqu'à la moitié du son (0.6s) puis diminue de façon symétrique.

1. En pratique, nous avons utilisé la fonction sgram de la toolbox LTFAT, Linear Time Frequency Analysis Toolbox, boîte à outil libre pour l'analyse de signaux `http://ltfat.sourceforge.net`



Son 29 *Ligne droite descendante, grande taille, pas d'irrégularité, dynamique non définie*
L'image temps-fréquence illustre clairement la forte variation de fréquence des différentes composantes spectrales dans la première moitié du son. Sur une durée de moins d'une seconde, la composante la plus grave varie de 650Hz à moins de 50Hz. A noter que la variation de fréquence en fonction du temps est en forme de sigmoïde et correspond à la fois à une translation et à une dilatation du spectre. On ne peut ignorer le fait que cette variation de hauteur correspond au stéréotype de l'objet qui tombe comme c'est le cas dans les bandes sons de dessins animés.

Son 32 *Ligne droite verticale, grande taille, pas d'irrégularité, dynamique non définie*
L'image temps-fréquence montre un signal dont l'amplitude globale augmente au cours du temps avec des composantes basses fréquence marquées qui varient uniquement en amplitude. On peut quasiment parler de modulation périodique d'amplitude autour de 2Hz mais avec des décalages de phase entre les composantes. L'augmentation globale d'énergie est à peu près linéaire sur toute la durée du son mais dépend en partie de la bande de fréquence considérée.



Son 38 *Forme non-définie, trajectoire montante, grande taille, pas d'irrégularité, dynamique non définie*
On voit sur l'image temps-fréquence que c'est un signal bruité avec un maximum d'énergie autour

de 200Hz. On peut presque considérer ce signal comme ayant été obtenu par filtrage passe-bas avec une augmentation linéaire de la fréquence de coupure au cours du temps.

Son 39 *Oscillation circulaire, orientation non-définie, trajectoire de taille moyenne, un peu d'irrégularité, vitesse constante*
On voit sur la transformée de Gabor, qu'un motif se répète relativement régulièrement à 8Hz. Ce motif contient de l'énergie principalement entre 500 et 1300Hz avec des modulations formantiques ainsi qu'une part non négligeable d'énergie dans les hautes fréquences. Au début, à la fin et surtout au milieu du son, on peut constater un léger déplacement de l'énergie vers le haut.

Les quelques remarques qualitatives que l'on a pu faire en considérant ces images temps-fréquence permettent déjà de dresser des premières conclusions. Le caractère oscillant peut probablement être relié à une succession relativement régulière de motifs élémentaires de type "impact". Le caractère linéaire peut être associé à des variations d'énergie, de répartition d'énergie dans le plan temps-fréquence (et notamment de hauteur perçue).

5.1.2 Test d'instationnarité

Comme on l'a vu au chapitre 1 (section 1.4.2) [Xiao *et al.*, 07b, Xiao *et al.*, 07a] ont proposé un test pour caractériser la stationnarité. Nous avons donc évalué nos 40 sons à l'aide de ce test, en utilisant pour les analyses temps-fréquence la méthode multi-résolutions [Xiao et Flandrin, 07] préconisée par les auteurs. Nous avons choisi de tester différentes tailles de fenêtre d'analyse pour pouvoir estimer la ou les échelles auxquelles se trouvent les instationnarités. Les n tailles de fenêtre retenues sont 2^n, $n \in [5, 14]$ (de 32 à 16384) points. En pratique, nous avons utilisé l'implémentation et les paramétrages fournis par les auteurs [2].

Tous les sons (y compris ceux qui n'évoquent pas de mouvement) sont significativement non stationnaires à toutes les échelles testées. Ce test met en jeu plusieurs grandeurs qui sont susceptibles d'expliquer les réponses que nous avons obtenues lors des tests perceptifs (figure 5.1). En particulier, en travaillant avec plusieurs fenêtres d'analyse, le test met en évidence l'échelle de temps pour laquelle l'instationnarité est maximale. On remarque que les sons pour lesquels l'instationnarité est du même ordre de grandeur pour toutes les échelles (aucune échelle n'émerge), correspondent globalement aux sons dont la trajectoire a été parcourue à vitesse constante. Les sons ayant une échelle d'instationnarité qui émerge pour les fenêtres de 2048 et 4096 points correspondent majoritairement à des trajectoires de petite taille.

Bien que ces résultats préliminaires nous incitent à poursuivre ce travail pour essayer d'expliquer les caractéristiques des signaux qui sont responsables des différents mouvements évoqués par les sons, ces analyses ne permettent pas de faire de la synthèse. A noter que nous avons également procédé au calcul d'autres grandeurs déduites d'analyses temps-fréquence comme par exemple des distances de type Itakura-Saito sur les spectres à court terme, que les résultats préliminaires et la complexité de mise en œuvre ne nous ont pas incité à poursuivre, ils ne seront donc pas présentés ici.
Avant de pouvoir poursuivre un travail d'analyses plus détaillées, il est important de s'assurer que les données perceptives que l'on veut pouvoir expliquer sont suffisamment générales et solides. Parmi les méthodes d'évaluation de ces données perceptives, il semble intéressant d'utiliser la synthèse qui permet de travailler avec des stimuli parfaitement contrôlés et dont la mise en œuvre peut également fournir des informations utiles à de futures analyses.

2. et disponible à l'adresse : `http://perso.ens-lyon.fr/patrick.flandrin/stat_test.html`

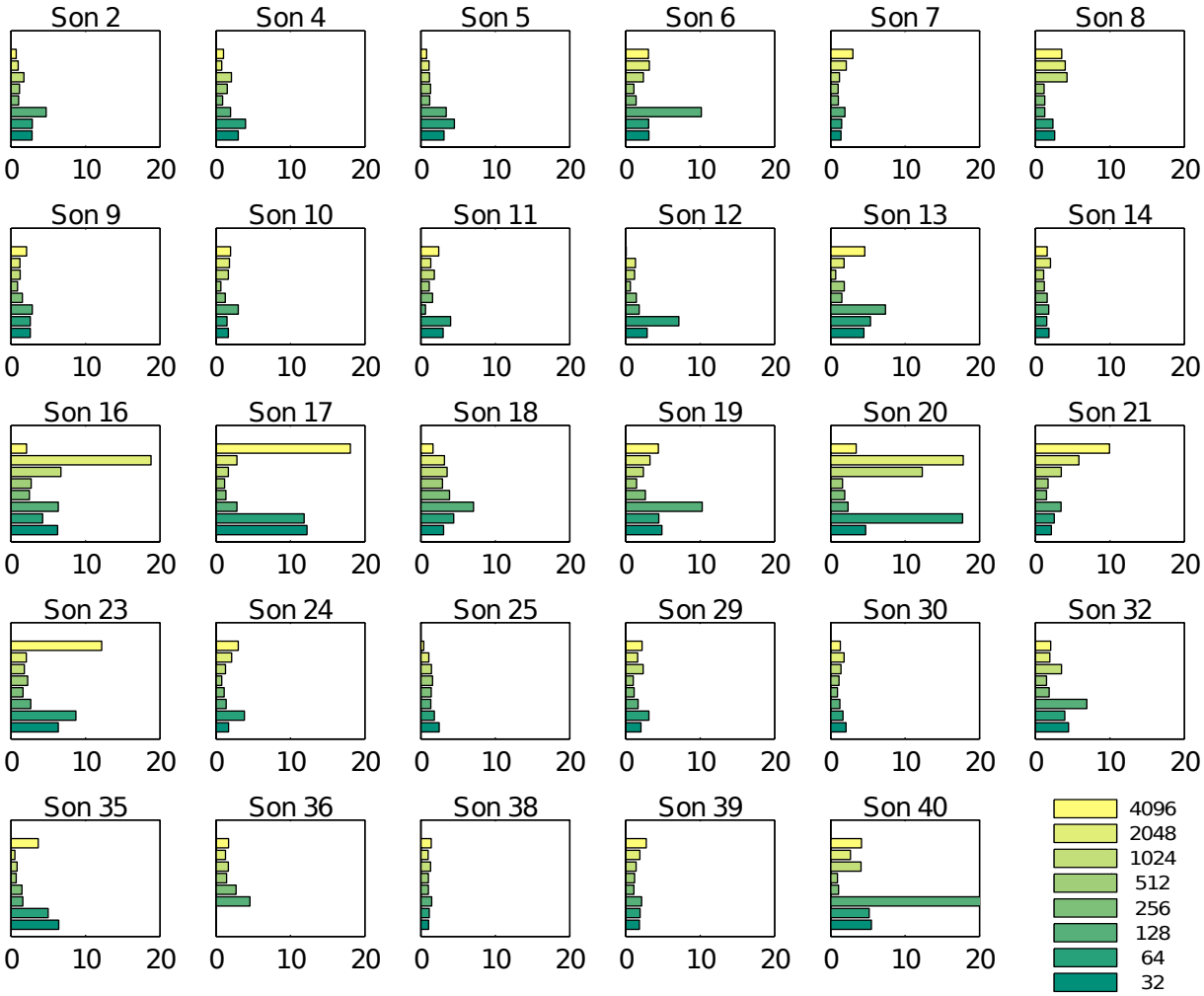


Figure 5.1 – Instationnarité (en abscisses) des 29 sons qui évoquent du mouvement et pour 8 fenêtres d'analyse. L'échelle d'instationnarité (de 0 à 20) est arbitraire.

5.2 Moteur de synthèse temps réel

Outre la validation de nos hypothèses sur les dimensions perceptives, une des questions importantes pour notre problème et qui n'a jusqu'à présent que très peu été discutée, est de savoir à quel point les évocations de mouvement sont liées aux phénomènes physiques et aux transformations sonores qui en découlent. Autrement dit, dans quelle mesure le mouvement évoqué peut être considéré comme une "métaphore" du mouvement physique. Cette question revient à chercher à savoir si les grands types de mouvement que l'on a identifiés peuvent être reproduits avec des transformations liées à la physique du mouvement et surtout quelles libertés il est possible de prendre par rapport à la physique. Cette section fournit des éléments de réponse à ces questions par la mise en œuvre d'une stratégie de synthèse basée sur la physique en même temps qu'elle répond aux autre contraintes énoncées précédemment.

Le moteur de synthèse fonctionne à partir d'un module qui génère des textures (signaux stationnaires *cf.* [Saint-Arnaud et Popat, 98] pour une définition précise) où différentes méthodes peuvent être combinées : synthèse additive, soustractive et granulaire. Des transformations sont ensuite appliquées sur ces textures et contrôlées par la même interface que celle utilisée dans le test 4 du chapitre précédent.

5.2.1 Stratégie de contrôle

La stratégie de synthèse proposée s'appuie sur 4 transformations relativement bien connues (*c.f.* [Chowning, 71] par exemple) : variation d'intensité en fonction de la distance source-auditeur, variation de hauteur due à l'effet Doppler, filtrage passe-bas due à l'absorption de l'air ainsi que réverbération. Il faut préciser que cette stratégie a été choisie parmi de nombreuses autres possibles et que d'autres stratégies sont discutées en conclusion de ce chapitre.

Pour chaque transformation, on se donne la possibilité d'augmenter ou de réduire son effet à l'aide de coefficients C_{transf} où $transf = \{intensite; doppler, absorption, reverb, elevation\}$. $C_{transf} = 1$ correspond à une transformation physiquement réaliste alors que des valeurs supérieures ou inférieures à 1 permettent respectivement d'amplifier ou de diminuer la transformation. Ces coefficients ont ensuite été calibrés par un test d'écoute pour obtenir le rendu le plus convaincant possible (*cf.* section 5.2.2).
La variation d'intensité est simulée en contrôlant un facteur de gain défini par

$$G_I = \frac{1}{C_{intensite} \times (DSA - 1) + 1} \tag{5.1}$$

où DSA est la distance source-auditeur.
La variation de hauteur due à l'effet Doppler est simulée en utilisant une ligne à retard, dont le temps de retard D_t est lié à la distance source-auditeur par la relation :

$$D_t = \frac{C_{doppler} \times DSA}{c} + D_{t0} \tag{5.2}$$

avec c vitesse du son dans l'air et D_{t0} le temps de retard minimum (pour une implémentation temps réel, on choisi la taille du vecteur de signal avec laquelle on travaille).
En plus de cette transformation, on considère une variation de hauteur pour matérialiser les déplacements sur l'axe vertical Z. Une source qui se déplace sur l'axe Z est simulée par une variation linéaire de hauteur. Cette variation est appliquée directement sur le module qui génère les textures. La hauteur H est donc définie par

$$H = C_{elevation} \times Z + H_0 \tag{5.3}$$

où H_0 est la hauteur de référence (en Hz).
L'atténuation des hautes fréquences due à l'absorption de l'air en fonction de la distance source-auditeur est simulée par un filtre passe-bas de type "shelf" dont la fréquence de coupure est fixée à 500 Hz et le facteur de gain calculé suivant la relation

$$G_f = \frac{1}{C_{absorption} \times DSA} \tag{5.4}$$

La réverbération utilisée est de type "feedback delay network" [Jot et Chaigne, 91] avec un temps de réverbération fixé à 2 secondes. On contrôle le rapport entre signal brut et signal réverbéré R_{wd} avec la relation

$$R_{wd} = \frac{1}{C_{reverb} \times DSA - 1} \tag{5.5}$$

Enfin, nous avons implémenté deux stratégies différentes pour le contrôle de la taille des trajectoires :
- La taille est contrôlée implicitement à travers la distance source-auditeur ; une grande taille donnera des variations plus importantes qu'une petite taille.
- Toutes les transformations sont calculées à partir d'une seule et même taille fixe, et on fait varier la hauteur et l'intensité moyenne du son en fonction de la taille (le son est plus faible et plus aigu pour des trajectoires de petites tailles).

La durée du son est synchronisée avec le mouvement de la sphère avec des rampes linéaires de 40ms appliquées en début et fin de son.

L'ensemble des transformations présentées ci-dessus a été mis en œuvre en temps réel avec le logiciel Max/Msp [3] en utilisant l'interface de dessins du test 3 présenté section 4.2 pour définir la distance source-auditeur (*cf.* figure 5.2).

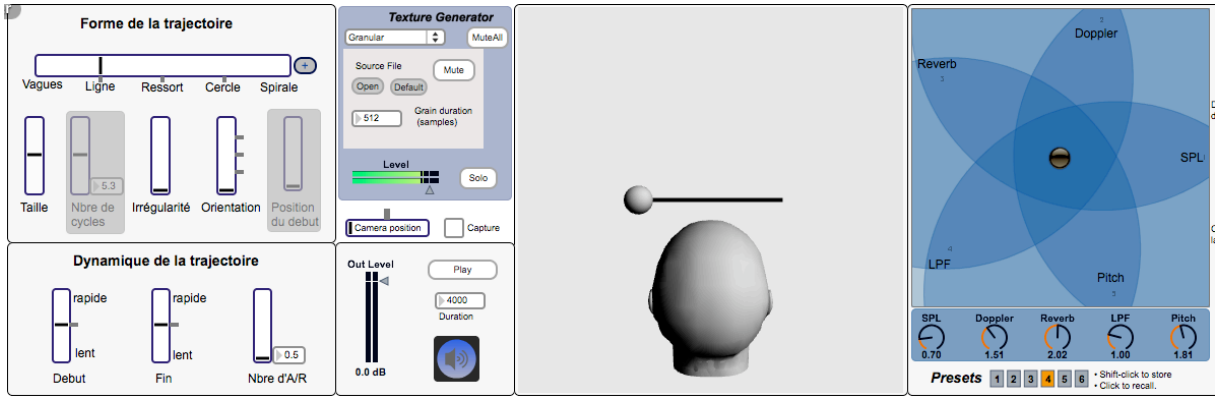


Figure 5.2 – Interface graphique du synthétiseur de mouvement reprenant les mêmes contrôles que l'interface d'évaluation des sons. Les contrôles de la partie droite de l'interface permettent de modifier l'importance relative attribuée à chaque transformation.

5.2.2 Calibration

Les coefficients utilisés pour amplifier (ou atténuer) les effets des 5 transformations ont été calibrés à l'aide d'un test perceptif effectué sur 5 sujets "experts" [4]. Pour ce faire, 10 trajectoires ont été définies à savoir les 8 *Formes* différentes et 2 supplémentaires avec différentes *Fréquences*. Il s'agissait de proposer pour chacune, le réglage des 5 paramètres qui fournit le résultat le plus convaincant. Pour faciliter la tâche, nous avons intégré à l'interface un contrôle graphique 2D qui permet de modifier les 5 paramètres simultanément (partie droite de la figure 5.2). Chacun des paramètres peut également être réglé indépendamment à l'aide d'un curseur rotatif. La moyenne des réglages proposés par chaque sujet et pour chaque trajectoire est finalement la valeur retenue pour le test de validation.
Les valeurs obtenues (et les grandeurs acoustiques correspondantes aux calibres retenus) sont :

- $C_{intensite} = 0.7$ (atténuation pour la distance max. D_{max} : 16 dB)
- $C_{doppler} = 1.5$ (temps de retard maximum : 50 ms)
- $C_{elevation} = 1.8$ (variation d'environ 1 octave à partir de $H_0 \simeq 150Hz$)
- $C_{absorption} = 1$ (atténuation de 20 dB)
- $C_{reverb} = 2$ ($R_{wd} = \text{inf}$ pour $D = 0.9D_{max}$)

Les résultats de cette étape de calibration montrent que les différentes transformations sont toutes susceptibles de fournir des sons qui évoquent du mouvement et ce, indépendamment les unes des autres. Pour certaines *Formes*, il existe des configurations de paramètres de calibration qui sont jugées plus convaincantes pour un sujet mais nous n'avons pas identifiée de configuration qui soit commune à tous les sujets. De ce point de vue il y a donc une part subjective dans ce jugement. Pour les valeurs retenues, la majorité des transformations sont amplifiées à part pour les variations d'intensité mais il faut préciser que d'autres transformations contribuent aux variations de sonie (principalement le filtrage passe bas et les variations de hauteur). Enfin la variation de hauteur suivant l'axe z qui est la seule transformation qui ne s'appuie pas sur des considérations physiques semble importante.

3. à part la réverbération qui a fait l'objet d'une implémentation spécifique en langage C réalisée par C.Verron
4. On entend par "expert" des sujets ayant des connaissances en synthèse sonore leur permettant de prévoir les conséquences sur le sonore en agissant sur les paramètres de calibration.

5.3 Test de validation

Il s'agit maintenant de valider les résultats obtenus dans le chapitre précédent avec le synthétiseur ainsi mis au point. Les hypothèses que nous souhaitons évaluer sont les suivantes :
- Les principales catégories de *Formes* se distinguent très clairement et à l'intérieur d'une catégorie, des formes visuellement différentes (par exemple cercle et spirale) seront jugées très similaires à l'écoute.
- Les principales *Directions* ("Monte", "Descend" et "Horizontal") se distinguent et on a tendance à juger plus proches deux trajectoires dont l'une est légèrement inclinée et une qui est verticale, que la même trajectoire inclinée avec une trajectoire horizontale, même si la trajectoire inclinée est visuellement beaucoup plus proche de l'horizontal.
- L'évaluation de la *Fréquence* d'oscillation devient moins précise quand elle augmente, et donc a même écart de fréquence, on jugera plus proches deux trajectoires qui oscillent rapidement que deux trajectoires qui oscillent lentement.

5.3.1 Stimuli

Les sons et les vidéos correspondantes utilisées dans ce test sont exclusivement issus du synthétiseur (*c.f.* section 5.2) et de la visualisation des trajectoires que contient son interface de contrôle. Tous les stimuli ont été générés sur la base d'une même texture obtenue par synthèse granulaire (durée des grains : 512 samples à 44,1KHz, possibilité de maximum 16 grains simultanés) sur la base d'un signal bruité comportant des composantes harmoniques en dessous de 2kHz. Comme on peut le voir sur la figure 5.3 la texture est caractérisée par de légères variations aléatoires de l'énergie au cours du temps. Comparé à un son de même densité spectrale de puissance mais parfaitement stationnaire, cette texture a l'intérêt de présenter de légères variations aléatoires qui rendent son écoute moins fastidieuse.

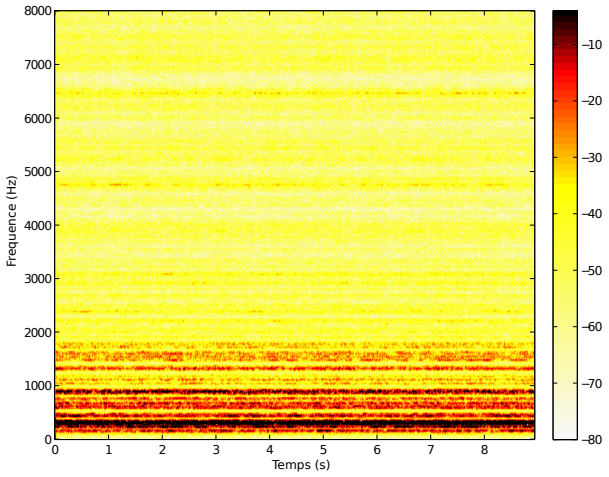


Figure 5.3 – Transformée de Gabor d'un extrait de la texture utilisée pour générer les stimuli.

Les hypothèses ont été testées séparément pour les trois dimensions (*Forme*, *Direction* et *Fréquence*). Les paires son/vidéo ont été générées avec les paramètres suivants :
- *Forme* : 8 formes élémentaires *Vague, Creux, Dôme, Rebond, Ligne, Ressort, Cercle* et *spirale*. Pour les formes "oscillantes", la fréquence à été fixée à 4 périodes, les trajectoires sont toutes horizontales (sauf *Rebond* qui est par définition vertical, position initiale en haut)
- *Direction* : Ligne droite avec 5 directions différentes (parfaitement horizontal, parfaitement vertical montant et descendant, déviant légèrement de l'horizontal montant et descendant)
- *Fréquence* : Vague horizontale avec 4 fréquences différentes (3, 4, 6 et 8 périodes).

Tous les stimuli ont été générés à partir de trajectoire de taille 0.5 (qui occupe la moitié de la fenêtre) avec pas d'irrégularité, une vitesse constante et pour une durée totale de 4 secondes. On rappelle que le synthétiseur procède à un fenêtrage avec des rampes linéaires de 4 *ms*.

5.3.2 Sujets

20 sujets d'une moyenne d'âge de 28 ans ont participé à l'expérience dont 8 femmes et 13 musiciens (au moins 10 ans de pratique d'un instrument). Comme pour les tests précédents, la majorité des sujets travaillaient dans la recherche en acoustique sur le campus dans lequel s'est déroulé cette recherche. Parmi les 20 sujets, 12 ont participé à au moins une des expériences précédentes.

5.3.3 Protocole

Le protocole qui a été mis en place est un test d'évaluation de la cohérence entre un son et d'une vidéo qui sont présentés simultanément. Les sujets devaient répondre au moyen d'un curseur linéaire gradué de 0 à 20 (0 pour "pas du tout cohérent" et 20 pour "parfaitement cohérent"). Les stimuli étaient croisés au sein de trois blocs, chaque bloc correspondant à une des dimensions testées. On a donc 8^2 paires pour les différentes *Formes*, 5^2 paires pour les différentes *Directions* et 4^2 paires pour les *Fréquences*. Les 105 paires étaient présentées aléatoirement dans une seule et même séance. L'interface graphique a été programmée avec Max/Msp, et se base simplement sur un lecteur de vidéo et un lecteur de son (*c.f.* figure 5.4). En plus de la position du curseur pour chaque paire les temps de réponse et le nombre de lectures sont enregistrés. En majorité, les sujets

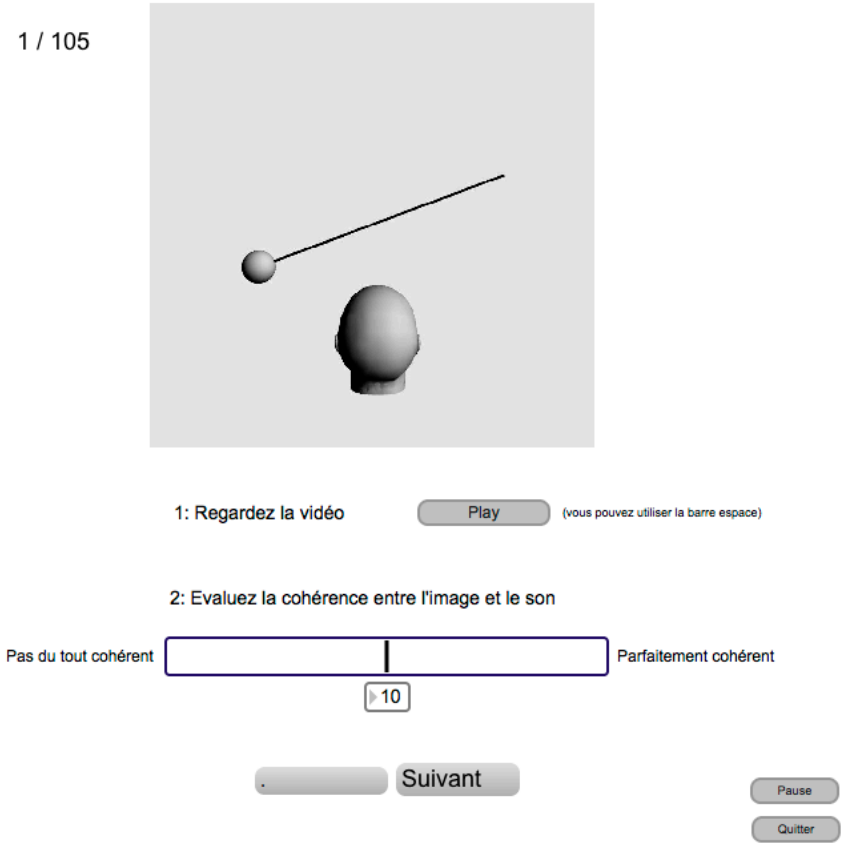


Figure 5.4 – Interface graphique du test de comparaison son/vidéo.

ont fait le test sur leur ordinateur personnel (et avec leur matériel audio). L'interface de test leur a

été distribuée et ils avaient pour consignes d'utiliser un casque, de se mettre dans une pièce calme, et de garder le niveau de restitution constant durant tout le test.

5.3.4 Résultats

Le test a duré en moyenne 15,3 minutes (écart type 2,6 min) et chaque paire a été écoutée en moyenne 1,25 fois.
Les analyses sont organisées de façon à répondre à deux questions. La première est de savoir si le synthétiseur génère des sons qui sont bien adaptés aux vidéos, autrement dit si les sujets ont jugé "proches" les associations de sons avec des vidéos qui ont été utilisées pour générer ces mêmes sons et "éloignés" des associations de sons avec des vidéos qui correspondent à d'autres sons dans le synthétiseurs. La deuxième question est d'évaluer si les hypothèses de départ qui ont été formulées à la suite du test de caractérisation graphique présenté au chapitre 4, sont (ou non) vérifiées.

Les données sont présentées sous forme de matrices de confusion de taille $N \times N$, où N correspond aux nombres de valeurs testées pour chaque dimension (soit 8 pour les *Formes*, 5 pour les *Directions* et 4 pour les *Nombre de cycles* ; *c.f.* figure 5.5).

La matrice de confusion est remplie entièrement (contrairement à ce qui est obtenu lors d'un test de comparaison par paires où la moitié de la matrice moins la diagonale est remplie) et elle est non-symétrique de diagonale non nulle (ni unitaire).

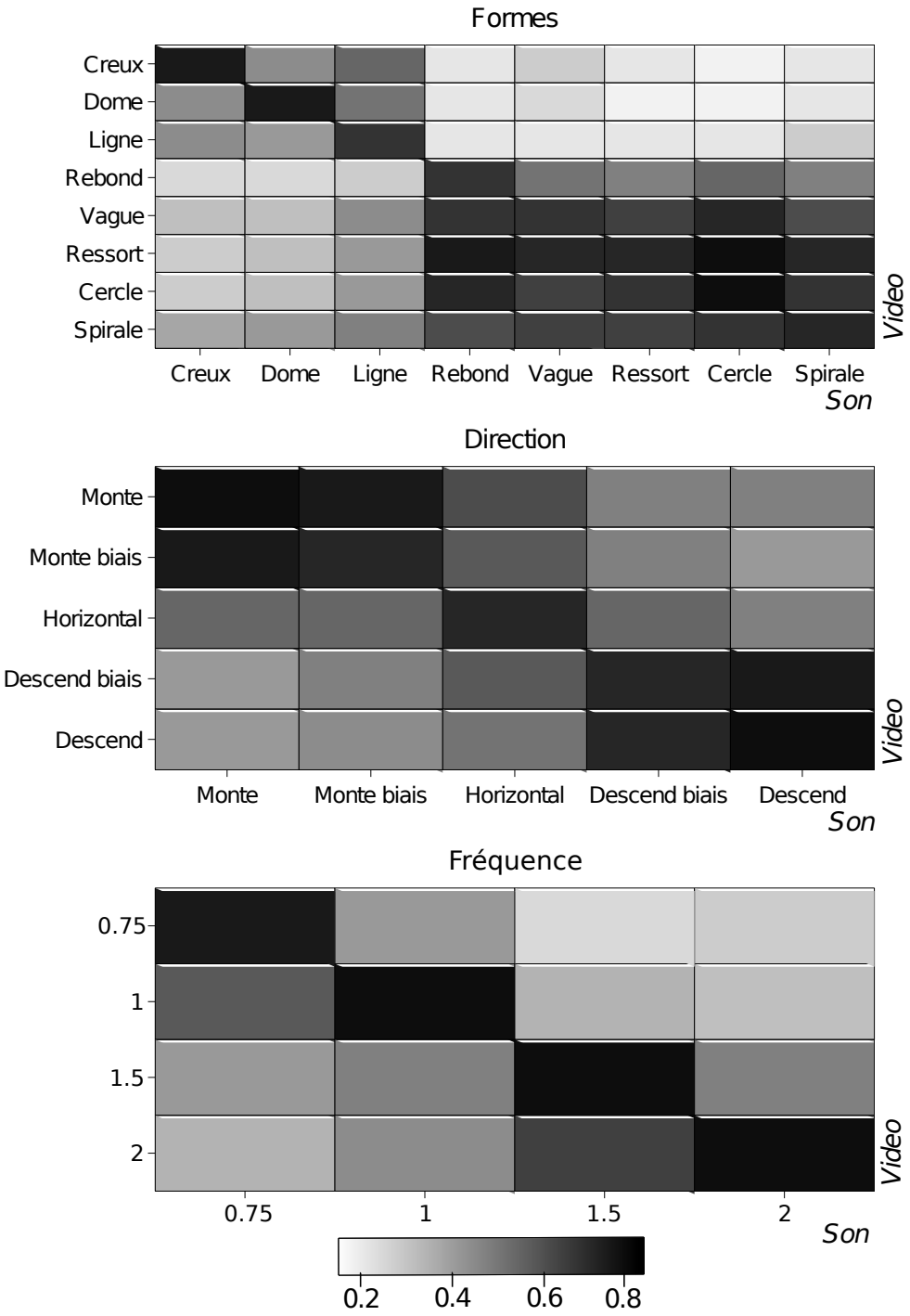


Figure 5.5 – Matrices de confusion (moyennées sur les sujets) représentant les réponses pour chacune des trois dimensions testées. Plus une case est foncée, plus le son et la vidéo ont été jugés “cohérents” l’un par rapport à l’autre. L’échelle de gris correspond à 1 pour 100% des sujets ayant trouvé l’association son/vidéo parfaitement cohérente et 0 pour 100% des sujets ayant trouvé l’association pas du tout cohérente.

Pertinence de la stratégie de synthèse proposée

Dans un premier temps, on peut extraire des données brutes (représentées par les matrices de confusion, figure 5.5) des informations qualitatives sur les comportements moyens. Force est de constater que les diagonales des matrices de confusion sont globalement plus sombre que le reste ce qui indique que les sujets ont jugé plus cohérentes les paires associées, notamment pour les oscillations où la distinction est très nette. Il est évident que les parties non diagonales sont non-nulles ce qui s'explique par les similitudes qui existent entre différentes valeurs de nos trois dimensions (notamment certaines *Formes* qui sont relativement proches entre elles comme *Cercle* et *Spirale*).
Nous avons donc procédé à une analyse de variance (test de kruskall-Wallis) en considérant la moyenne des réponses obtenues pour deux regroupements, l'un constitué des paires correspondant aux diagonales des matrices de confusion (sons et vidéos associés) et l'autre constitué des paires hors diagonales (sons et vidéos non-associés). Ces analyses se sont révélées significatives pour les trois variables (*Forme* :$\chi^2(63, 1216) = 826, p < 0.01$, *Direction* : $\chi^2(24, 475) = 173.44, p < 0.01$, *Fréquence* : $\chi^2(15, 304) = 182.39, p < 0.01$) ce qui indique que notre stratégie de synthèse fournit des sons cohérents avec la visualisation des trajectoires correspondantes.

Validation des hypothèses

Nous avons ensuite transformé les matrices de confusion en matrices de dissemblance. Pour cela, on considère la cohérence entre un son et toutes les vidéos avec lesquelles il a été comparé, comme ses coordonnées (on a donc des espaces à 8 dimensions pour les *Formes*, 5 dimensions pour les *Directions* et 4 dimensions pour les *Fréquences*). Ces "coordonnées" sont ensuite utilisées pour calculer des distances entre les sons avec une norme Euclidienne. Deux sons seront donc proches si leurs évaluations respectives vis à vis de chaque vidéo sont similaires (on ne se limite donc pas aux évaluations croisées entre les sons et les vidéos associées). Il est certain que ces "distances" sont à considérer avec précaution puisqu'elle ne sont pas mesurées directement et qu'une tâche de comparaison par paire n'utilisant que des sons, a toutes les chances de donner des résultats différents.
Pour chacune des trois dimensions, nous avons ensuite procédé à une analyse en cluster ascendante hiérarchique (avec distance du plus proche voisin). Les résultats de cette analyse sont présentés sur la figure 5.6. Pour les *Formes*, on retrouve bien la distinction nette entre celles qui sont dites "linéaires" et celles qui sont "oscillantes". En revanche, on ne distingue pas clairement, au sein des trajectoires oscillantes, les oscillations circulaires et les oscillations "régulières". Cette confusion s'explique probablement par les paramètres (durée, phase, fréquence) qui sont parfaitement égaux entre les trajectoires et qui conduisent à des sons très proches d'un point de vue signal.
Les *Directions* confirment les hypothèses dans le sens où les trajectoires faiblement montantes et faiblement descendantes sont très proches de la version parfaitement verticale correspondante. La trajectoire horizontale se trouve, quand à elle, totalement séparée des 4 autres.
Également en accord avec nos hypothèses, les trajectoires dont les fréquences sont les plus élevées sont plus proches que celles dont les fréquences sont les plus basses alors même que les rapports de fréquences sont les mêmes et que la différence est plus importante entre les deux fréquences les plus élevées. Il faut quand même préciser qu'au regard de la variabilité inter-sujets les différences de distance en considérant 3 et 4 cycles d'un côté et 6 et 8 cycles de l'autre ne sont pas significatives.

Remarques sur les interactions multimodales

La tâche que nous avons proposée dans ce test pose clairement le problème des interactions entre les deux modalités que sont le visuel et l'auditif. Même si l'étude de ces interactions dépasse largement le cadre de notre étude, on se doit au moins de rapporter les réflexions que nous ont engendrées ce test. D'abord il est clair que cette tâche met en jeu des situations qui sont incongrues

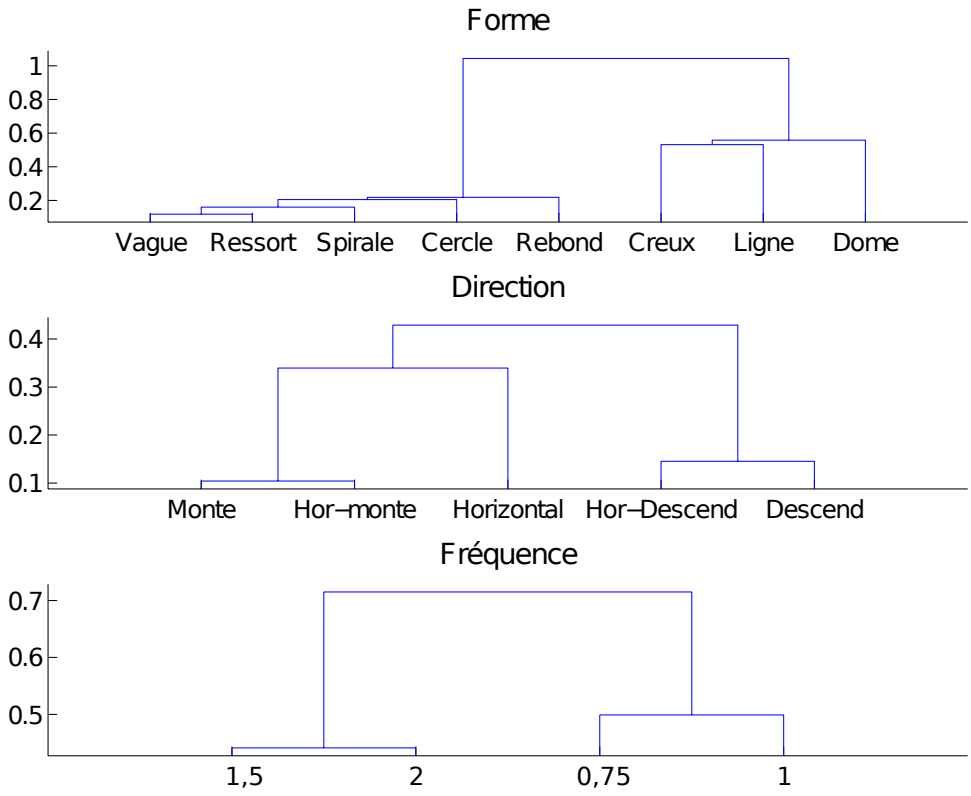


Figure 5.6 – Dendrogramme obtenu par analyse en cluster hiérarchique représentant les distances entre les différentes *Formes* (haut), *Directions* (milieu) et *Fréquences* (bas).

entre les deux modalités et que l'on n'a pas l'habitude de rencontrer dans la vie de tous les jours [5]. Ainsi on ressent une gêne importante quand un son et une vidéo non congruente sont présentés simultanément. De même, il semble que nous avons tendance à rechercher des relations entre les deux modalités même quand il y a clairement des incohérences, ce qui se traduit par des notes de congruence qui sont en moyenne relativement élevées (autour de 10 pour chacune des 3 dimensions). Enfin, l'observation des matrices de confusion (figure 5.5) montre qu'elles ne sont pas parfaitement symétriques ce qui va également dans le sens du problème de différences de traitement entre les deux modalités.

Pour quantifier ces différences, nous avons en plus calculé les matrices de distance entre les vidéos. Ainsi chaque vidéo est considérée comme un point dans un espace à N dimensions, chaque dimension étant un son avec lequel la vidéo à été comparée. Nous avons ensuite comparé les distances obtenues pour les sons précédemment calculées pour valider nos hypothèses avec celles obtenues pour les vidéos en utilisant une analyse de variance multivariée [6]. La figure 5.7 représente ces distances pour les deux modalités.

Ces analyses de variances se sont révélées significatives uniquement pour les *Formes* (*Forme* : Wilk's $\lambda = 0.1335$, $\chi^2(28) = 48.33$, $p < 0.01$, *Direction* : Wilk's $\lambda = 0.8253$, $\chi^2(10) = 4.3$, $p = 0.786$, *Fréquence* : Wilk's $\lambda = 0.478$, $\chi^2(6) = 10.163$, $p = 0.118$). On voit notamment dans ce cas (figure 5.7 en haut) que les sons sont globalement plus proches entre eux que les vidéos. Pour l'instant, il est difficile de conclure sur l'origine d'une telle différence et de plus amples analyses sont nécessaires pour déterminer si elles peuvent être dues à notre stratégie de synthèse ou à des différences fondamentales entre la perception des sons et la perception des images animées.

5. On pourrait qualifier ces situations de "non-écologique" après [Gibson, 78]
6. ou one way MANOVA pour Multivariate ANalysis Of VAriance

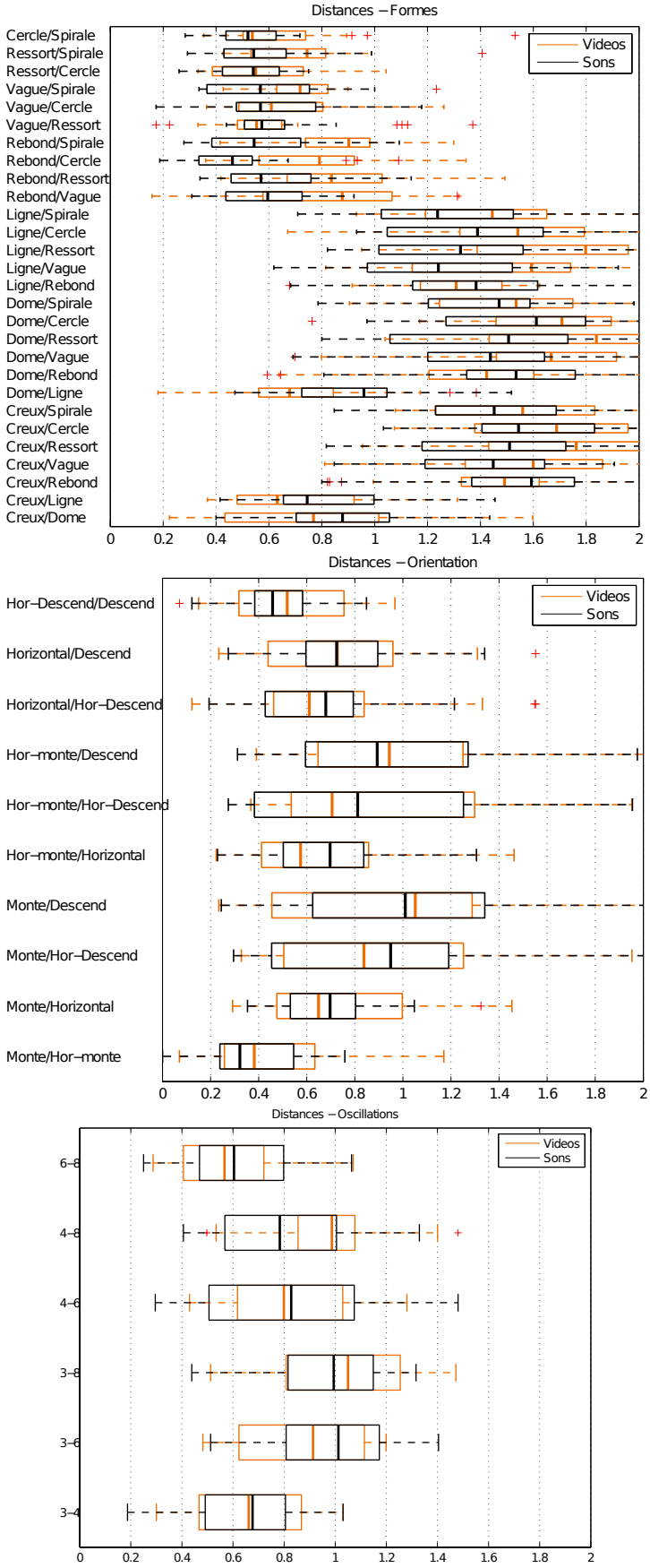


Figure 5.7 – Boîtes à moustache des distances entre les différentes *Formes*, *Directions* et *Fréquences*. Les distances sont obtenues en considérant la cohérence entre chaque son (en noir) avec toutes les vidéos et chaque vidéo (vert) par rapport à tous les sons.

5.4 Vers un contrôle générique du mouvement évoqué par les sons

Etant donnés les résultats obtenus dans toutes les études que nous avons menées, nous pouvons proposer un modèle générique pour le contrôle du mouvement évoqué par les sons.
Tout d'abord, nous avons identifié un ensemble de critères qui permettent de différencier ces mouvements et qui peuvent être considérés comme définissant une typologie. Cette typologie est basée sur différentes formes (Linéaire, oscillant), directions (monte, descend, horizontal), tailles (petit, moyen, grand) et dynamiques (vitesse constante, peu ou beaucoup de variations) qui définissent un ensemble de caractéristiques qui décrivent le mouvement de façon pertinente d'un point de vue perceptif.
Ensuite, plusieurs observations issues des expériences contribuent à l'avancée vers un contrôle générique du mouvement :

- Nous avons montré la pertinence des représentations graphiques pour décrire le mouvement. Par conséquent, notre contrôle se base sur n'importe quelle donnée graphique à condition qu'elle soit accompagnée d'une information temporelle (nous avons montré l'importance des aspects dynamiques).
- Des trajectoires complexes peuvent être décomposées et simplifiées en trajectoires équivalentes, c'est-à-dire qui seront perçues comme étant identiques
- Différentes stratégies de synthèse sont possibles, les variations des paramètres acoustiques associées aux caractéristiques des trajectoires peuvent être adaptées aux besoins des utilisateurs.

Le modèle d'analyse et de sonification de trajectoires que nous proposons est présenté sur la figure 5.8. Ce modèle peut être décomposé en trois étapes successives :

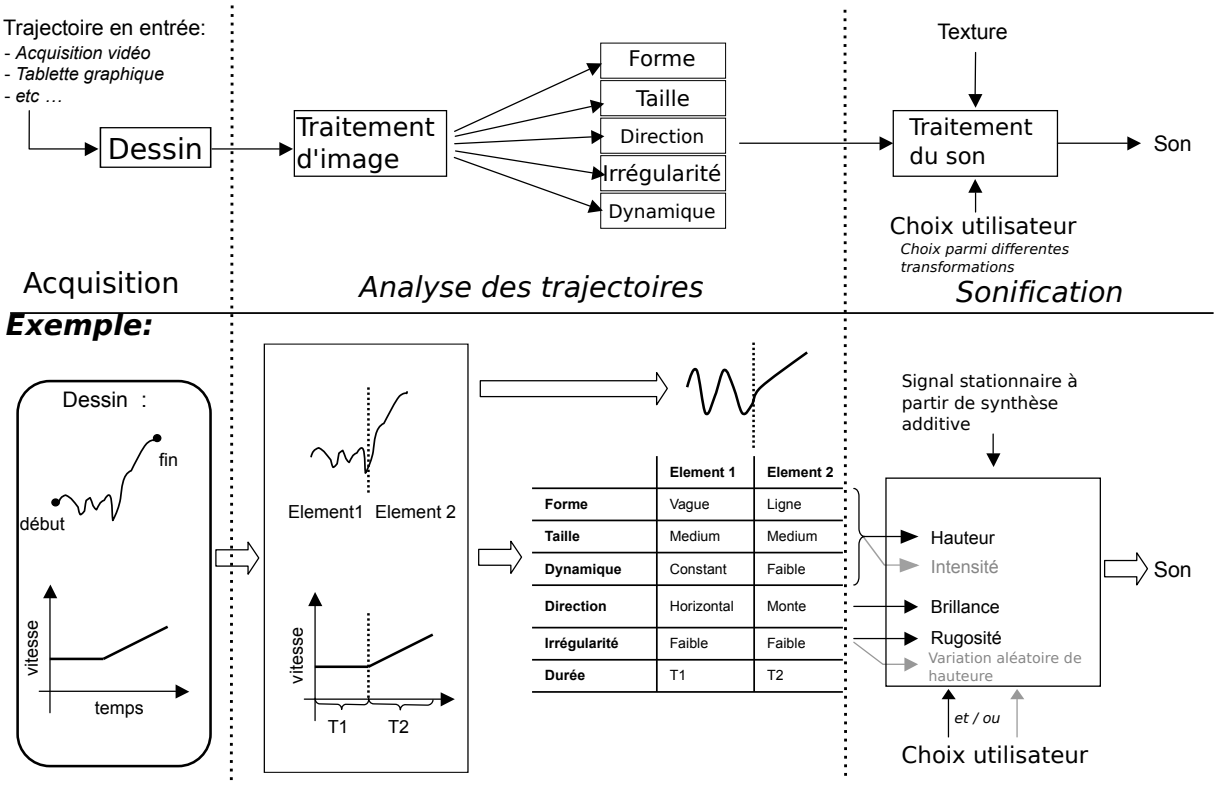


Figure 5.8 – Proposition de modèle d'analyse et de sonification de trajectoires.

1) L'acquisition des données qui peuvent être de natures très différentes (geste, image,...) et dont d'éventuelles transformations (reconnaissance de formes, filtrages,...) doivent les ramener à une trajectoire sous forme de courbe échantillonnée.

2) Les trajectoires sont analysées en décomposant et simplifiant les différents éléments qui les constituent. La décomposition consiste à identifier les parties qui correspondent aux formes élémentaires. Ce problème est tout à fait comparable à celui de la reconnaissance de caractères manuscrits pour lequel il existe des méthodes automatiques. Pour chaque segment, il s'agit également d'identifier l'ensemble des caractéristiques à savoir la taille, l'orientation, l'irrégularité et la dynamique. De même il existe des méthodes permettant d'extraire automatiquement ces caractéristiques (dont beaucoup sont décrites dans [Bishop, 06]). Il s'agira d'adapter ces méthodes en tenant compte des données perceptives obtenues dans nos expériences.
3) On procède enfin à la sonification en associant des transformations sonores aux différentes caractéristiques des trajectoires. Plusieurs variables pourront agir sur une même transformation (par exemple en modifiant la profondeur de modulation et la fréquence s'il s'agit d'une modulation d'amplitude) et une même variable pourra engendrer plusieurs modifications simultanées du signal. Il semble également intéressant de pouvoir laisser aux utilisateurs la possibilité de modifier la manière dont les caractéristiques des formes sont reliées aux modifications appliquées au signal pour adapter l'outil à leurs besoins.
Avant de pouvoir proposer un tel outil, il semble néanmoins nécessaire d'évaluer dans quelle mesure une transformation peut être reliée à une caractéristique du mouvement car il est évident que certaines relations sont incompatibles comme par exemple la direction et le taux de réverbération. Il sera également nécessaire de travailler sur la calibration des transformations pour que les limites et la loi d'évolution de la transformations soient pertinentes pour une utilisation donnée.

5.5 Conclusion

Dans ce chapitre nous avons tout d'abord présenté des résultats préliminaires concernant l'analyse des signaux en relation avec les dimensions perceptives. Bien que relativement succinctes, ces analyses ouvrent un certain nombre de pistes pour arriver à trouver des critères objectifs caractérisant le mouvement évoqué par les sons. Nous avons ensuite décrit une stratégie de synthèse qui nous a permis de générer des stimuli utilisés dans un test perceptif cherchant à vérifier les hypothèses formulées à la suite des études menées au chapitre précédent. La validation de ces hypothèses confirme l'intérêt de la démarche qui a été suivie et nous permet de poursuivre dans l'utilisation des résultats obtenus.
Dans le même temps, nous avons posé la question du lien entre le mouvement évoqué par les sons et les phénomènes physiques qui concernent le mouvement. Même si aucune preuve n'a été fournie, il semble que le mouvement évoqué est principalement lié aux lois d'évolution de caractéristiques des sons, que ces évolutions soient ou non des conséquences acoustiques du mouvement d'un objet. En revanche il est certain que les transformations que nous avons utilisées et qui s'appuient principalement sur des considérations physiques sont pertinentes pour évoquer du mouvement. Une première application a été de poser les bases d'un contrôle générique du mouvement permettant la sonification de dessins. Il est certain que ce travail n'en est qu'à ses prémices mais de nombreuses pistes de recherche ont été ouvertes et demandent à être poursuivies. Dans le chapitre suivant, il sera donc question de ces perspectives avec également un retour sur les idées et les questions qui ont servi de bases à ce qui a été réalisé au cours de cette thèse. Parmi ces idées de départ, on s'intéressera en particulier aux sons abstraits pour lesquels on a montré ici qu'ils permettent d'obtenir des résultats qui ont été confirmés avec des sons calibrés issus d'un synthétiseur.

Chapitre 6

Discussion et perspectives

6.1 Retour sur les sons abstraits 102
6.1.1 Utilisation des sons abstraits en psychiatrie 102
6.1.2 Le naturel 103
6.1.3 Sur l'utilisation de sons abstraits 104
6.2 Vers une méthodologie générale pour le contrôle intuitif de la synthèse 106
6.3 Applications et perspectives 107
6.3.1 Applications à court terme 107
6.3.2 Applications musicales 108
6.3.3 Perspectives 109

Les travaux qui ont été présentés dans les chapitres précédents ont ouvert de nombreuses pistes de réflexion notamment au travers des tests perceptifs. Ce chapitre a donc pour but de reprendre ces réflexions, notamment pour envisager des applications aux résultats obtenus.

6.1 Retour sur les sons abstraits

L'utilisation de sons abstraits, est un des points de départ de cette thèse et il s'agit ici de revenir sur l'intérêt qu'ils ont pu avoir dans le travail réalisé. Nous avons eu l'occasion de partager nos réflexions sur les sons abstraits avec une équipe de chercheurs travaillant dans des domaines très différents du nôtre et les résultats obtenus dans ce cadre contribuent à enrichir notre vision de la problématique. Nous commencerons donc par rapporter cet échange avant de conclure sur une utilisation plus globale des sons abstraits en perception sonore et en audio.

Le contenu de cette section a été développé dans [Merer et al., 10]

6.1.1 Utilisation des sons abstraits en psychiatrie

Les réflexions que nous avons eues autour des sons abstraits nous ont amené à collaborer avec une équipe de médecins autour d'un protocole nouveau concernant le diagnostic de la schizophrénie. Comme on l'a discuté précédemment, les sons abstraits peuvent être interprétés de façon très différente et parmi ces interprétations, se trouvent les notions d'originalité ou de bizarrerie. Hors la notion de bizarrerie est très importante dans les échelles d'évaluation des troubles psychiques telles que le DSM-IV [APA, 94]. Chez les patients atteints de schizophrénie, il est également souvent question d'audition notamment au travers d'hallucinations auditives (perception sans objet). Il se trouve que pratiquement aucune étude ne s'intéresse à la notion du bizarre dans la perception des sons [Cermolacce *et al.*, 10]. Les sons abstraits permettent donc d'étudier d'éventuelles différences entre sujets témoins et patients atteints de schizophrénie dans l'évaluation du caractère "bizarre" de sons.

Nous avons participé à la mise en place d'un test d'écoute dont une partie des sons est commune avec les sons utilisés durant les tests décrits au chapitre 4, le corpus étant complété par des

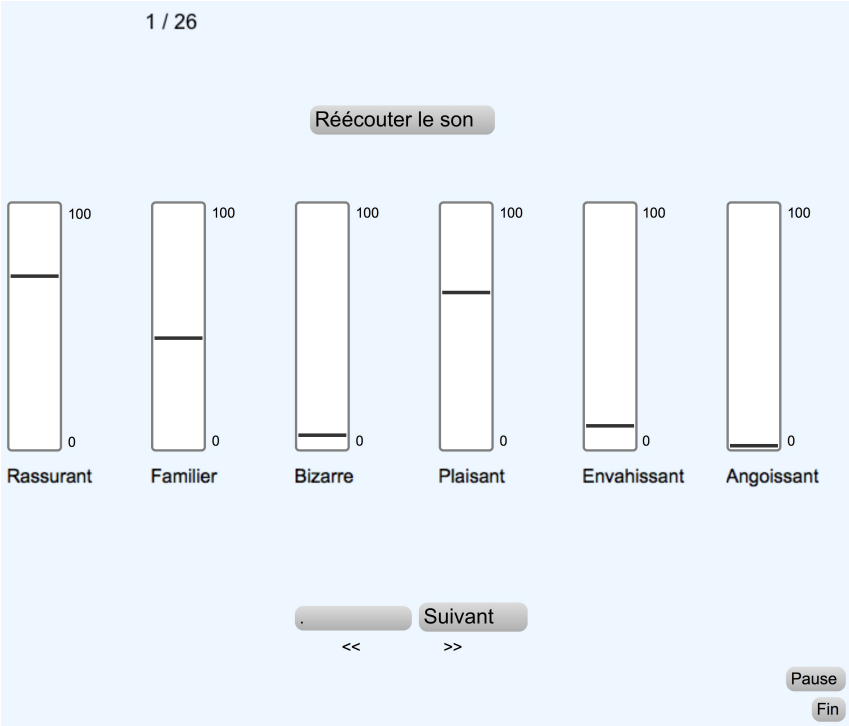


Figure 6.1 – Interface graphique du test d'évaluation de sons abstraits

sons d'environnement censés être jugé plus familiers que les sons abstraits. Le test consistait en l'évaluation de 26 sons sur 6 échelles verticales allant de 0 à 100 correspondant aux dimensions : "familier", "bizarre", "rassurant", "plaisant", "envahissant" et "angoissant" (*cf.* figure 6.1). 40 sujets ont passé ce test dont la moitié en tant que groupe contrôle ne présentant aucun trouble psychiatrique. Les résultats indiquent que les patients ont tendance à juger moins familiers, des sons qui sont jugés familiers par les témoins (et qui sont tous des sons d'environnement) alors que les sons jugés bizarres par les témoins (et qui sont tous des sons abstraits) sont jugés plus familiers mais tout aussi bizarres. De plus, les sons jugés familiers par le groupe contrôle sont jugés plus envahissant par les patients atteints de schizophrénie [Micoulaud-Franchi *et al.*, 11b].
Plusieurs autres résultats peuvent également être rapportés. Tout d'abord, il est clairement apparu que ce test n'a pas posé de problèmes aux sujets quelque soit le groupe considéré. D'ailleurs, les 6 dimensions étudiées se retrouvent dans des études relativement anciennes qui s'intéressaient à des descriptions sémantiques du timbre [Solomon, 57], [Bismarck, 74]. La notion de bizarrerie n'exclut pas les caractères plaisant et rassurant ce qui peut permettre dans un corpus de sons abstraits, d'équilibrer la valence émotionnelle. On trouve également des sons qui ont été jugés bizarre tout en étant familier, même si les deux dimensions sont anti-corrélées ($r = -0.7$).

Cette étude qui fait suite à une autre concernant des sons d'environnement [Micoulaud-Franchi *et al.*, 11a] montre clairement l'intérêt des sons abstraits et la possibilité de mettre en œuvre des approches similaires dans des domaines de recherche très différents.

6.1.2 Le naturel

On a vu dans les résultats du test de caractérisation du corpus (section 4.5.1) que la plupart des stimuli ont été jugé "synthétiques" par les sujets alors même qu'ils ont majoritairement été obtenus à partir d'enregistrements de sons "naturels". Dans ce même test, nous avons également mentionné les liens entre la notion de naturel/synthétique avec le caractère reconnaissable/non-reconnaissable de la source. La figure 6.2 montre que l'évaluation des sons sur ces deux dimensions est quasiment

identique. En synthèse et en particulier en synthèse de parole, la notion de naturel a été utilisée

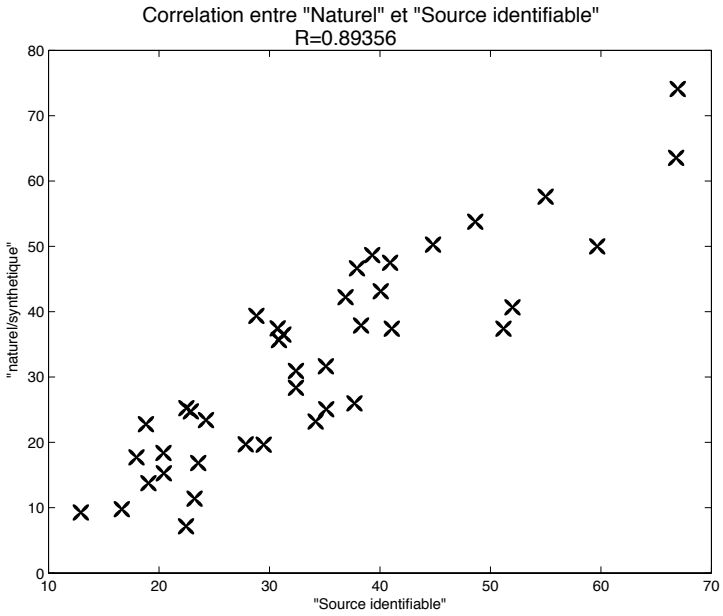


Figure 6.2 – Note moyenne obtenue aux questions Naturel/synthétique en fonction et Facilement reconnaissable/Difficilement reconnaissable pour chacun des 40 sons utilisés au chapitre 4

pour évaluer la qualité du modèle c'est-à-dire s'il est capable de donner l'illusion que le son est produit par un humain et non par un ordinateur (voir notamment [Nusbaum *et al.*, 97]). On est très loin du problème de la reconnaissance de sources, celles-ci étant de toute façon clairement identifiée. Par exemple, dans le domaine de la parole, même la moins performante des synthèses sera reconnue comme étant de la voix. On peut également citer l'exemple de [Farner *et al.*, 06] où les auteurs s'intéressaient à la clarinette et ont montré que le manque de naturel était relié à l'absence de variations de timbre. La notion de naturel est donc fortement dépendante du contexte dans lequel elle est utilisée et semble relativement multi-dimensionnelle. Concernant les sons abstraits, il semble que ce soit une notion qui exprime la difficulté (ou la facilité) avec laquelle on relie un son avec une cause physique. Le naturel semble donc être une notion importante qui ne peut être ignorée quand on s'intéresse au contrôle de la synthèse et c'est ce qui nous a poussé à collaborer à un projet s'intéressant au contrôle d'un modèle physique de clarinette [Guillemain et Merer, 10]. Les résultats de cette étude ont montrés que le contrôle et l'instrument avaient tous les deux une influence sur la perception du naturel.

6.1.3 Sur l'utilisation de sons abstraits

Dans les études que nous avons menées, les sons abstraits ont eu un rôle central car ils nous ont permis d'obtenir des corpus sans poser d'hypothèses de départ. Les corpus utilisés dans les tests étaient constitués de sons ayant des caractéristiques spectro-temporelles très variées et dont le "sens" que leur attribuent les auditeurs s'adaptent au contexte de l'étude. On a vu qu'un même corpus de sons abstraits pouvait être utilisé dans des études très différentes et que suivant la consigne, les sujets vont s'intéresser à des aspects très différents. On peut citer ici [Schön *et al.*, 10] où le même type de corpus constitué avec la même méthode de sélection (typologie de Schaeffer) a été utilisé dans un contexte encore différent : celui des neurosciences cognitives. Il s'agissait dans cette étude de comparer les processus cognitifs qui sont mis en jeu lors de l'écoute de sons par rapport à ceux qui sont mis en jeu avec des mots [1]. Ce qui est particulièrement intéressant pour notre discussion,

1. Les mécanismes d'attribution de sens sont très complexes et encore mal connus et il n'est pas envisageable ici d'expliquer en détail les problématiques étudiées ainsi que les méthodologies mises en œuvre. On renvoie donc ici le

c'est qu'il était nécessaire dans cette étude de faire en sorte que les sujets comparent effectivement la relation sémiotique entre les mots et les sons et non pas les mots avec un autre mot qui est automatiquement associé au son dont l'origine est évidente. Les sons abstraits étaient donc les seuls candidats possibles pour cette étude. Le fait que les sons abstraits limitent cet "étiquetage" n'empêche pas que l'on puisse leur associer des mots. En revanche, le mot qui leur est associé ne peut pas (ou peu) faire référence à une cause certaine et sera donc relativement variable d'un individu à l'autre.
Cette variabilité a d'ailleurs été décrite dans [Petitmengin *et al.*, 09] où les auteurs s'intéressent aux "histoires" que l'on se raconte lors de l'écoute de sons et notamment lors de l'écoute de sons abstraits. Les différents protagonistes rapportent beaucoup d'éléments qui ont traits à leur propre vécu. C'est d'ailleurs ce qui nous pousse à avancer l'hypothèse que les sons abstraits favorisent une écoute musicale [Gaver, 93a] c'est-à-dire qui se concentre plus sur les caractéristiques du signal (timbre, hauteur,...) qu'à une information sur la source ou l'environnement.

Si on s'intéresse aux différentes catégories de sons et notamment celles qui sont utilisées pour la recherche, on peut alors se poser la question de la place qu'occupent les sons abstraits. Le fait de s'intéresser à une catégorie particulière de sons est d'ailleurs un des aspects qui caractérise un domaine de recherche et les passerelles ne sont pas toujours évidentes entre les travaux qui s'intéressent à la parole, aux sons d'environnement ou à la musique. Il existe très peu d'études ayant pour but de proposer une catégorisation des sons, la plus utilisée étant celle de Gaver qui concerne principalement les sons d'environnement et qui se base sur la nature des sources acoustiques et les types d'interactions [Gaver, 93b]. Nous sommes à une époque où les haut-parleurs occupent une place de plus en plus importante dans la vie quotidienne et où les appareils que l'on manipule utilisent le son pour nous donner des informations sur leur fonctionnement. Les études perceptives réalisées en laboratoire utilisent d'ailleurs majoritairement des sons diffusés aux travers de haut-parleurs.
Nous sommes donc de plus en plus confrontés à l'écoute de sons sans lien direct avec la source qui est à l'origine du son. Il semblerait donc relativement logique de considérer une catégorisation des sons qui ne s'intéresse pas seulement aux sources mais également à la facilité avec laquelles celles-ci sont reconnues, ou plus généralement aux informations qu'on peut retirer des sons. Il semble d'ailleurs que, le fait de reconnaître plus ou moins la source influence la manière dont sont catégorisés les sons [Lemaitre *et al.*, 10]. [Ballas, 93] a décrit l'augmentation des temps de réponse dans l'identification de sources pour des sons enregistrés lorsque le nombre de causes possibles augmente au vu des caractéristiques acoustiques des sons. Même si on a parfois dans ce document parlé des sons abstraits comme étant une catégorie, il semble donc qu'il faille considérer des frontières continues entre les sons abstraits et les autres types de sons. Ainsi on parle de "degrés d'identification" ou de "facilité" à identifier une source au même titre qu'on parle du caractère "tonal" d'un son puisqu'il est possible de passer continûment de sons parfaitement harmoniques à des bruits aux caractéristiques les plus diverses [Jensen, 05].
Avant de conclure sur l'utilisation des sons abstraits, il est important de rapporter un dernier résultat obtenu dans nos expériences. Dans la caractérisation du corpus de sons abstraits (test 3 section 4.4.3) nous avions également interrogé les sujets sur d'éventuels mots associés aux sons. Les réponses obtenues (*cf.* tableaux en annexe C) montrent que plus un son est "reconnaissable", plus le nombre de sujets qui vont associer un mot est important. Il est également apparu que les mots associés ne correspondent pas à ce qui est réellement à l'origine du son mais plutôt à une source dont les caractéristiques acoustique sont proches. Par exemple le son 15 a entre autres obtenu des mots comme "chat", "scie circulaire", "tigre" ou "cymbale". On voit donc bien que les sujets identifient des caractéristiques du signal qui sont commune à ces différentes sources et c'est ce qui nous permet de dire que les sons abstraits aident à identifier des propriétés intrinsèques des signaux qui sont responsables d'évocations. Cette caractéristique est un des points clés dans les démarches que nous avons suivies vers des contrôles intuitifs pour le mouvement évoqué par les sons.

lecteur à l'article référencé.

De façon comparable aux images du même nom, les sons abstraits offrent de vastes possibilités d'interprétations aux auditeurs pour lesquels les facultés d'analyse du signal sonore sont fortement mises à contribution.

6.2 Vers une méthodologie générale pour le contrôle intuitif de la synthèse

Si on reprend la démarche globale que nous avons proposée en introduction de ce document (figure 1 p.9), on constate que le travail réalisé est relativement loin des objectifs fixés et que seule la première étape d'évaluation perceptive a été menée à son terme. On peut nuancer cette faiblesse par le fait que nous avons maintenant une vision globale de la méthode proposée et que nous connaissons mieux les contraintes qui doivent être remplies pour qu'elle aboutisse. En effet, il est apparu que l'évaluation perceptive est une étape fondamentale et que de ses résultats, dépendent tous ceux qui peuvent être obtenus par la suite. En ce sens il s'agit de l'étape la plus importante qui fournit à elle seule des sons et des informations sur ce qui est perçu par les auditeurs. En ce qui concerne l'étape d'analyse des signaux, il est certain qu'elle est totalement tributaire de la robustesse des informations perceptives sur lesquelles elle s'appuie ainsi que sur le nombre et la représentativité des sons qui sont analysés. En considérant les résultats obtenus avec l'approche classique qui se base sur le calcul de descripteurs de timbre il nous a semblé que des avancées notoires pourraient être faites en caractérisant les signaux de manière plus globale et en s'intéressant avant tout à l'évolution du signal plutôt qu'à des caractéristiques qui concernent le signal sur toute sa durée. C'est dans cette logique que sont menés des travaux utilisant des multiplicateurs de Gabor [Olivero *et al.*, 09, Olivero *et al.*, 10] que nous avons suivi de très près au cours de cette thèse. En résumé, il s'agit de caractériser dans le plan temps fréquence, les informations qui différencient deux signaux ce qui permettrait à terme d'identifier des "profils" caractéristiques d'un groupe de sons. Même si nous n'avons pas eu l'occasion de mettre en place de telles analyses, il est certain qu'il s'agit d'une des grandes perspectives ce travail.

L'étape d'analyse passe également par une mise en relation de la description des signaux avec les données perceptives. Même si les résultats obtenus à l'aide de méthodes d'apprentissage (par exemple les réseaux de neurones) peuvent être très convaincants, il semble que des méthodes moins "aveugles" pourraient permettre une meilleure compréhension du problème. Il faudrait par exemple pouvoir envisager d'expliquer certaines dimensions perceptives avec plusieurs descripteurs dont l'importance varie en fonction d'autres paramètres ce qui peut être réalisé en considérant des modèles Bayesiens [Karklin et Lewicki, 05]. Il semble d'ailleurs que ce type de modèle soit cohérent avec les processus cognitifs concernés [Griffiths *et al.*, 08].
Pour identifier les descripteurs pertinents, ceux-ci doivent nécessairement être présents parmi ceux qui sont donnés en entrée du modèle ce qui implique de définir à l'avance des descripteurs qui sont pertinents. A ce titre, des considérations perceptives peuvent être intégrées au calcul de descripteur comme par exemple des considérations sur la mémoire à court terme des auditeurs comme le propose [Jensen, 11]. Il faut également être sûr que les descripteurs sélectionnés décrivent bien les causes (souvent physiques) qui sont à l'origine de ce qui est perçu et s'ils ne sont pas plutôt la conséquence d'une cause différente. Enfin lorsque l'on s'intéresse à la synthèse, il s'agit de faire des analyses qui ne se contentent pas de pouvoir prédire le jugement d'auditeurs mais bien de pouvoir reproduire des informations perçues. Cette contrainte implique une analyse du signal qui n'est pas simplement descriptive mais aussi explicative. Un modèle idéal permettrait de comprendre quelles sont les caractéristiques des sons utilisées et surtout comment sont traitées ces caractéristiques par notre cerveau pour attribuer un jugement.

Le synthétiseur que nous avons conçu nous permet de formuler quelques hypothèses quant à la manière dont sont perçues les différentes transformations et nous invite à ajouter une étape dans la

démarche générale que nous avons suivie. Il semble que différentes transformations sont susceptibles de produire une impression de mouvement cohérente avec une même trajectoire. L'essentiel est donc de garantir une évolution temporelle qui soit cohérente entre celle du mouvement qui est décrit dans l'interface (par les paramètres de contrôle et la visualisation) et celle des caractéristiques du son. En revanche, les transformations sonores doivent être calibrées au sens où les variations perçues doivent répondre à un cahier des charges précis en termes de loi d'évolution. Cette calibration est donc une étape qui est nécessaire au même titre que l'évaluation finale du synthétiseur. La question de l'évaluation est également un problème complexe qui n'a reçu que très peu de réponses applicables à tous types de sons, même dans les cas où la synthèse vise à reproduire des sons existants [Miner *et al.*, 02].

Beaucoup d'études utilisant la synthèse ont contribué à comprendre ce qui, dans les signaux sonores, est important d'un point de vue perceptif notamment avec l'approche d'analyse par synthèse. Il semble par contre qu'il soit relativement difficile de concevoir des synthétiseurs à partir d'études perceptives et d'analyses de son. Néanmoins, la mise en œuvre de contrôles intuitifs passe nécessairement par un questionnement de ce qui est perçu ou plutôt ressenti par les auditeurs-utilisateurs. Le fait d'agir sur le dispositif qui génère le son (et donc ne n'être pas simplement auditeur) a d'ailleurs une influence qui a notre connaissance n'a pas été étudiée et qui le mériterait sûrement.

6.3 Applications et perspectives

6.3.1 Applications à court terme

Application à l'apprentissage de l'écriture

Les résultats obtenus au chapitre 4 posent les bases pour la mise en place d'outils permettant l'analyse de dessins en vue de leur sonification. Nous avons entre autres proposé un modèle générique de contrôle du mouvement (*cf.* section 5.4). Il est envisagé dans le cadre d'un projet qui vient de débuter (ANR MetaSon [2]), d'appliquer ce modèle dans un outil d'aide à l'apprentissage de l'écriture. Les enfants dysgraphiques arrivent en général assez bien à reproduire les lettres de l'alphabet et même les mots entiers et en tout cas le résultat est souvent assez proche de celui réalisé par des enfants normaux-scripteurs. Le problème réside principalement dans les aspects dynamiques de l'écriture, les trajectoires étant souvent très complexes avec des pauses, des discontinuités voire des retours en arrière inopportuns qui rendent l'écriture pénible et gène l'apprentissage. L'information nécessaire pour résoudre ces problèmes est donc principalement temporelle et ne peut pas être fournie avec un modèle visuel statique. L'idée de ce projet est donc d'essayer de fournir un outil permettant de donner un retour sonore qui serve de guide pour l'écriture. Même s'il est évident que l'outil de synthèse développé durant cette thèse n'est pas directement en mesure de répondre à cette problématique, les résultats qui ont servis à sa mise en place peuvent tout à fait être utilisés.

Applications "industrielles"

Comme discuté dans l'introduction de ce document, les applications de la synthèse sont majoritairement musicales mais il existe également des applications industrielles qui peuvent être concernées par les travaux réalisés ici. Dans le domaine de l'automobile, [Sciabica *et al.*, 10] s'intéressent aux bruits intérieurs d'un véhicule lorsque celui-ci accélère. Le problème est non pas de réduire le niveau sonore à l'intérieur de l'habitacle mais de comprendre ce qui est responsable de la sensation de puissance et de sportivité dans le son qui parvient au conducteur. Il s'agit donc, et la synthèse peut partiellement répondre à ce problème, de lier des caractéristiques des signaux

2. http://metason.cnrs-mrs.fr/

(qui sont nécessairement dynamiques) à une sensation. Ce problème est donc relativement proche de celui abordé durant la thèse et il y a de fortes chances que des caractéristiques des signaux qui seront identifiées pour notre problème s'appliquent également dans ce cas.
Toujours dans le domaine de l'automobile, l'arrivée prochaine de voitures électriques sur le marché pose la question du son qu'il va nécessairement falloir leur associer. En effet, il s'agit de véhicules quasiment silencieux et qui représentent un danger potentiel pour les piétons et particulièrement les aveugles [3]. D'autre part, les sons qui parviennent au conducteur lui apportent des informations importantes sur le fonctionnement de son véhicule, les supprimer est également un nouveau facteur de risque. Pour ces deux problèmes, il est tout à fait envisageable de générer des sons en lien avec des paramètres comme la vitesse ou l'accélération du véhicule et qui transmettent aussi bien au conducteur qu'aux piétons, une sensation de vitesse ou des informations plus précises sur la trajectoire du véhicule. C'est également un aspect auquel s'intéresse le projet ANR MetaSon (mentionné au paragraphe précédent) qui envisage de répondre à ce problème en utilisant la synthèse.

6.3.2 Applications musicales

Description et indexation pour les musiques électroacoustiques

A l'origine de cette thèse se trouvent des idées issues des problématiques rencontrées par les compositeurs de musiques électroacoustiques pour ce qui est de la description et la recherche d'éléments structurels de ces musiques (comme cela a été mentionné dans la discussion sur la définition de la problématique section 1.1 à propos des UST). [Bootz et Hautbois, 08] ont dors et déjà proposé un système de reconnaissance automatique des UST basé sur une description du signal appelés Motifs Temporels Paramétrés. De plus plusieurs travaux se sont intéressés à la description de musiques avec le problème de la segmentation au travers de la notion de "chunk" [Godøy, 09, Godøy *et al.*, 10], [Jensen et Kühl, 10] qui s'appuie en autres sur des considérations sur la mémoire et le geste.
La contribution qui serait envisageable ne concerne donc pas directement l'analyse de musique mais se situe à un niveau plus élémentaire, pour caractériser des sons isolés. Il existe dans ce domaine plusieurs travaux qui se basent sur les critères typologiques et morphologiques de Schaeffer [Ricard et Herrera, 04, Peeters et Deruty, 08, Bloit *et al.*, 09b]. Dans une logique similaire, il semble relativement intéressant pour les compositeurs de musiques électroacoustiques d'avoir une description des sons (qui sont susceptibles d'être utilisés dans des compositions) qui se base sur le sens qu'ils véhiculent. Cette notion de sens est d'ailleurs souvent mentionnée comme étant un élément manquant des classifications de Schaeffer. Il semble donc intéressant de mettre en place un outil de reconnaissance automatique du mouvement évoqué par les sons pour l'indexation des banques de sons des compositeurs. Bien entendu le mouvement évoqué n'est pas le seul élément à utiliser pour "annoter" les sons mais il n'en est pas moins un élément important.

Création et design sonore

Avec du recul, il semble que c'est dans le domaine de la création que les applications de l'outil de transformation/synthèse sont les plus évidentes. En permettant à l'utilisateur de définir une stratégie de transformation qui lui convient, et en offrant toute liberté sur le contenu spectral du son avant transformation, on dispose d'un outil tout à fait adapté à la création. Il serait néanmoins bienvenu de pouvoir utiliser en entrée de l'outil, n'importe quel type de son et non uniquement des textures (stationnaires). Il sera pour cela nécessaire, d'être en mesure d'analyser en amont le type de mouvement que le son est susceptible d'évoquer et d'adapter les possibilités de l'interface à cette analyse. Par exemple, si on souhaite travailler à partir d'un son qui est identifié comme évoquant un

3. Ce problème a fait l'objet de beaucoup de débats et a conduit les autorités de plusieurs pays à légiférer (On pourra consulter sur ce sujet les nombreux liens fournis sur la page wikipedia dédiée http://en.wikipedia.org/w/index.php?title=Electric_vehicle_warning_sounds&oldid=411158895)

mouvement circulaire il doit être possible d'enlever cette caractéristique aussi bien que la renforcer. Etant donné l'importance de la notion de mouvement évoqué et la grande avancé que représente la mise en place de contrôles intuitifs, on peut espérer que ce type d'outils constituera un nouvel argument en faveur de l'utilisation de la synthèse sonore dans ces domaines.

6.3.3 Perspectives

Perspective pour d'autres thématiques que le mouvement

Le test de caractérisation perceptive a été conçu spécifiquement pour la problématique du mouvement et ne peut donc pas directement être transposé à d'autres problématiques. On peut néanmoins recenser des éléments de notre démarche qui peuvent être utiles pour des dimensions perceptives autres que le mouvement :
- L'interface du test perceptif a été conçue pour être également l'interface de contrôle du synthétiseur. Pour d'autres problématiques, il est tout à fait envisageable de mettre en place un test perceptif qui utilise la même interface que le futur synthétiseur. Même si les deux interfaces ne sont pas parfaitement identiques, il semble important de chercher à concevoir une interface qui soit au plus près des besoins du futur synthétiseur.
- Les paramètres de l'interface sont relativement explicites puisqu'ils sont supportés à la fois par un mot pour les décrire et par une visualisation.
Sans forcément passer par le dessin ou des systèmes de visualisation (avec tous les problèmes d'interactions multimodales que cela engendre), il semble important de rendre chacun des paramètres le plus clair possible. Il semble également nécessaire de prévenir les éventuelles confusions sur le sens d'un paramètre avec les conséquences que cela peut avoir sur la variabilité dans les réponses. La conception d'une interface pour d'autres problématiques devra donc se faire avec ce souci de rendre les paramètres le plus explicite possible notamment en évitant l'utilisation de descriptions verbales.
- La sélections des sons en effectuant un pré test avec un grand corpus de sons permet d'optimiser le nombre de sons utilisés en regard des paramètres de l'interface (*i.e* des dimensions perceptives testées). Cette méthode est comparable à la mise en place de plans d'expériences.
Si on prend l'exemple du problème des émotions, on peut travailler à partir des 6 émotions de base (Joie, Tristesse, Dégoût, Peur, Colère, Surprise) définies dans la théorie des émotions [Ekman *et al.*, 82] qui sont "orthogonales" entre elles et permettent d'obtenir toutes les autres. On peut également envisager de représenter les émotions au travers de l'échelle SAM [4] [Bradley et Lang, 94] qui utilise des dessins de petits personnages dont l'expression du visage retranscrit l'état émotionnel. En utilisant le même corpus de son, la procédure qui en découle peut donc être très proche de celle que nous avons suivie pour la problématique du mouvement.

Mapping interactif

Une autre approche possible pour la mise en œuvre de mapping intuitifs et qui est complémentaire de celle que nous avons suivie est la mise en place d'un protocole permettant de réaliser automatiquement un mapping en fonction des réponses de l'utilisateur. Cette idée s'inspire des travaux réalisés par des chercheurs utilisant une approche évolutionniste pour la synthèse (voir l'étude bibliographique en section 2.3) et où l'évaluation de l'utilisateur est prise compte pour faire évoluer la population de chromosomes permettant de générer des sons. Il semble d'ailleurs que l'utilisation d'algorithmes génétiques soit incontournable dans une telle démarche.
A propos des méthodes permettant de réaliser un mapping pertinent d'un point de vue perceptif, [Gounaropoulos et Johnson, 06] décrivent deux méthodes possibles. La première est celle que nous avons suivie durant cette thèse et qui se base sur l'analyse de sons qui ont été évalués par des auditeurs. La seconde qu'il qualifie d'approche analytique, consiste à évaluer les changements perçus quand on modifie un paramètre de synthèse. En analysant chaque paramètre on est donc en mesure

4. Self Assessment Manikin scale

de reconstituer les paramètres qui sont susceptibles de reproduire une caractéristique perceptive donnée.
L'idée à développer consiste donc à proposer des modifications du signal différentes en fonction des réponses de l'utilisateur pour aboutir à la transformation optimale par rapport au rendu perceptif souhaité par l'utilisateur. En revanche il ne s'agit pas simplement de retrouver une combinaison "fixe" des paramètres de synthèse qui donne un rendu perceptif précis, mais bien de trouver une loi d'évolution des paramètres qui permet d'obtenir des sons dont une caractéristique perçue varie (linéairement) en accord avec les attentes de l'auditeur. Il s'agit donc bien d'une méthode qui permettrait d'obtenir un contrôle "intuitif" et qui plus est adapté aux besoins particuliers d'un utilisateur.

En pratique, il s'agirait de concevoir une interface constituée d'un unique curseur et d'un système d'évaluation par l'auditeur du résultat obtenu quand il agit sur le curseur. Au niveau algorithmique, il s'agit en premier lieu de définir un ensemble de transformations du signal que l'on souhaite pouvoir effectuer. Il s'agit ensuite de définir un ensemble de lois d'évolution des différents paramètres que prennent en entrée chacune des transformations précédemment définies. La difficulté se trouve principalement dans la capacité à proposer des combinaisons de transformations qui convergent vers ce que souhaite l'auditeur en prenant en compte son évaluation. L'évaluation doit porter non seulement sur la capacité de la stratégie à reproduire une caractéristique perceptive mais en plus doit définir les limites du contrôle et pouvoir donner une sensation d'évolution linéaire quand l'utilisateur fait varier linéairement le paramètre. Ces idées sont très proche de celles développées dans [Garcia, 01, Garcia, 02] à la différence près que dans ces études, l'auteur cherchait le meilleur algorithme permettant de synthétiser un son donné et non une méthode pour faire varier une caractéristique acoustique. Le calcul de la fonction d'adaptation se fait uniquement à partir du signal alors que dans la proposition qui est faite ici, il s'agit d'intégrer l'avis de l'utilisateur dans l'algorithme d'optimisation.
Cette approche pourrait donc être menée en s'appuyant sur ce que nous avons déjà réalisé dans cette thèse, notamment la stratégie de synthèse, ce qui permettrait d'apporter une autre vision et des éléments nouveaux au problème du contrôle.

Conclusion

Ce travail possède indéniablement un caractère exploratoire et met en jeu des problématiques qui font appel à des domaines de recherche très différents. Sans aller jusqu'à l'expression utilisée par [Risset, 10], qui parle de l'interdisciplinarité "qui consiste à s'asseoir entre deux chaises, ce qui est particulièrement inconfortable en France", il est certain que c'est un problème difficile à aborder. Comme on le voit sur la figure 7.1, le nombre de disciplines concernées par la thèse sont nombreuses.

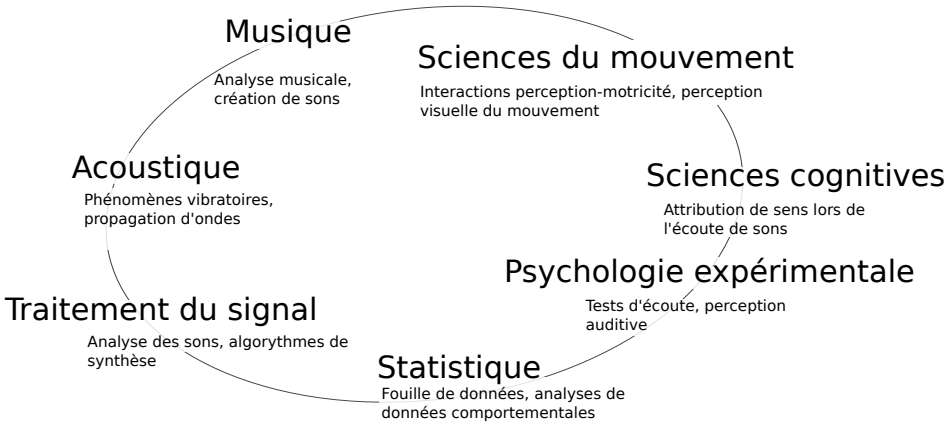


Figure 7.1 – Différents domaines de recherche qui ont été abordés durant la thèse

A l'origine de la thèse se trouvait des visées encore plus ambitieuses dans le sens où il était envisagé de mettre en place des protocoles basés sur les approches qui sont menées dans les neurosciences. Un travail préliminaire a d'ailleurs été mené sur les corrélats électrophysiologiques du mouvement pour voir s'il était possible d'envisager un protocole d'évaluation des sons qui se base sur une mesure objective de la perception (*cf.* section 1.5.2). En pratique l'utilisation de méthodes d'imagerie cérébrale n'était pas justifiée en regard de la quantité et la complexité des données obtenues en comportemental et c'est en partie pourquoi cette étude n'a pas été poursuivie.
La pluridisciplinarité impose nécessairement une formation (même succincte) dans des domaines qui sont abordés ce qui limite le temps passé à la recherche proprement dite. Relativement peu d'écrits existent pour "apprendre" la pluridisciplinarité et il est difficile de l'organiser à priori. J'ai par exemple découvert la difficulté qu'il y a, pour que des experts de différents domaines, trouvent tous un intérêt dans un projet et comment il est délicat, lorsqu'on a un besoin dans un domaine de compétence qui n'est pas le nôtre, d'avoir un échange qui soit réellement à double sens.
En pratique, l'essentiel des travaux réalisés l'ont été dans un domaine pour lequel ma formation ne me destinait pas ce qui demande un effort supplémentaire pour apporter une contribution à l'état de l'art.
Très peu d'études se sont intéressées à des questions similaires, celles qui s'en rapprochent le plus étant principalement issues du monde de la musique. Lors d'échanges avec d'autres autour des travaux réalisés dans le cadre de cette thèse, il est apparu que la problématique du mouvement est une problématique bien définie pour tout le monde mais avec une compréhension et des interprétations

de ce qui en découle au niveau du sonore qui sont très différentes suivant les interlocuteurs. Par exemple, pour certains, il n'est pas envisageable que le mouvement évoqué par les sons ne soit pas uniquement lié à la physique des sources qui a elle seule permet de prévoir ce qui sera perçu. Nous avons donc été confronté à de nombreuses idées reçues qui ont pu parfois créer des doutes sur la pertinence de l'approche qui a été suivie. Il a d'ailleurs été nécessaire dans la mise en place des protocoles de combattre nos propres idées reçues pour éviter de laisser de côté des éléments qui à première vue paraissaient évidents. Avec du recul, il est clair que les phénomènes étudiés sont complexes mais pour le moins passionnants et trouvent tous leur intérêt dans le contexte actuel d'utilisation croissante des sons pour transmettre des informations à des utilisateurs.
Malgré toutes ces difficultés, il a été possible de mener à bien plusieurs expériences qui apportent de nouveaux éléments pour comprendre la notion de mouvement évoqué par les sons. L'expérience de caractérisation de mouvement à travers le dessin montre d'ailleurs dans quelle limite ce qui est perçu comme du mouvement possède une dimension subjective. Il reste de nombreux verrous à faire sauter avant de pouvoir prétendre à une méthode qui permette de proposer une stratégie de synthèse et de contrôle à partir d'une description subjective du résultat attendu. En revanche, il semble que nombre de ces verrous ont été identifiés et demandent à être étudiés plus en détail.

La contribution apportée par cette thèse concerne donc principalement les aspects perceptifs du problème du contrôle et montre l'intérêt de la notion de mouvement évoqué par les sons comme dimension perceptive notamment dans le cas de sons abstraits. Au point de vue méthodologique, une avancée à été faite avec la mise en place d'un protocole adapté au problème du contrôle de la synthèse sonore. Une contribution a également été apportée quant à l'utilisation de sons abstraits, sons qui n'avaient jusqu'alors que très peu été étudiés en tant que tels et dont l'utilisation s'est avérée utile et prometteuse pour de futurs travaux.

Comme on l'a vu au chapitre précédent, les perspectives sont nombreuses et nous avons mis à jour plus de problèmes qu'apporté de réponses à la problématique de départ. Concernant le travail qui a été effectué, il s'est inscrit dans une durée forcément limitée et certains points restent inachevés offrant de nombreuses perspectives à court et long terme. Chacun des paramètres qui a été étudié dans le chapitre 4 (forme, taille, irrégularité, ...) devra faire l'objet d'études spécifiques et approfondies pour déterminer la précision et la limite de leur perception ainsi que les corrélats acoustiques. Un important travail devra être réalisé en traitement du signal, pour analyser précisément les sons et identifier leurs caractéristiques qui sont responsables des attributs perçus. Ce problème d'analyse est en fait l'étroite combinaison de deux problèmes l'un étant la description du signal l'autre étant la mise en relation de cette description avec les données perceptives.
Dans la suite logique des travaux qui ont abouti à la mise en place d'un formalisme pour le problème du mapping, il sera nécessaire de donner un cadre théorique aux aspects perceptifs du contrôle de la synthèse sonore. En d'autres termes, il sera nécessaire de trouver un modèle cognitif adapté à la boucle entre l'auditeur et le système de synthèse qu'il manipule.

Annexes

A Consignes et questionnaire des tests du chapitre 3

Consignes

Phase d'entraînement

Pour vous préparer au test que vous allez passer, vous allez entendre des sons enregistrés. L'objectif est d'arriver à fixer votre attention sur le type de mouvement auquel ils vous font penser (exemple : " *ça me fait penser à quelque chose qui tombe*") sans vous intéresser à la source. Il est très important d'essayer d'oublier l'objet ou les objets qui sont ou que vous croyez être à l'origine du son (exemple : " *ça me fait penser à un chat, à de l'eau qui coule*")

Phase 1

- Sur l'écran que vous voyez, chaque "S" correspond à un son que vous pouvez écouter autant de fois que vous voulez en cliquant dessus.
- Comme précédemment, vous ne chercherez pas à retrouver l'origine (objet/cause physique) du son
- Regroupez les sons qui vous font penser au même mouvement/déplacement.
- Vous pouvez faire autant de groupes que vous voulez et avec autant d'éléments par groupe que vous le souhaitez (chaque groupe peut contenir un nombre différent d'éléments).
- Pour des raisons pratiques, il est demandé de faire des groupes les plus compacts possible et bien répartis sur l'écran.

Phase 2

- Comme précédemment, il vous est demandé de ne pas chercher à reconnaître le ou les objet(s) qui sont à l'origine du son mais de vous focaliser sur l'impression de mouvement que vous donnent ces sons.
- Parmis ceux-ci, "N" seront repérés par une lettre de couleur bleu.
- Vous devez donc écouter ces sons et y associer les sons de la partie inférieure qui vous évoquent le même type de mouvement ou de déplacement.
- Certaines catégories peuvent rester vides et certains sons peuvent être considérés comme inclassables, ils seront laissés dans la partie inférieure.

Durant le test, n'hésitez pas à faire une pause si cela vous paraît nécessaire.

Questionnaire pour le test de classification

Sujet N° :

NOM/Prénom :
Contact (mail et/ou téléphone) :
Age :
Musicien :oui/non Niveau en musique :

Nombre de groupes réalisés :

Pour chaque groupe réalisé sur l'écran, donnez sa position et essayez de le décrire en un ou plusieurs mots(vous pouvez par exemple faire un dessin de l'écran).

Commentaire sur le test, les sons, les groupes que vous avez fait... :

B Consignes et questionnaires des tests du chapitre 4

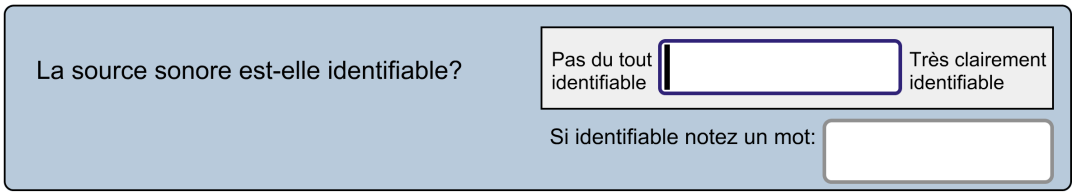
B 1 Test 3

Ces consignes sont présentée à l'écran avant l'entraînement, imprimées et posée sur la table durant le test et peuvent être affichées sur l'écran à tout moment (en version "résumée")

Le test auquel vous allez participer consiste à écouter des sons et répondre à huit questions:

Il vous faudra écouter le son et répondre à chacune de ces questions en déplaçant les curseurs, cochant les cases à cocher et en rentrant du texte dans les zones prévues à cet effet.

Question 1:

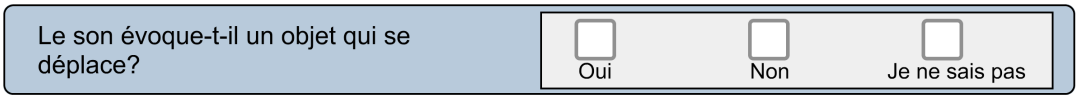


On entend par identifiable: Est-ce que vous pouvez associer facilement ce son à un objet ou une situation de la vie quotidienne.
Il est ensuite demandé de décrire brièvement ce que vous associez au son présenté.

Question 2:

On demande de juger si vous pensez que le son est un son naturel (enregistré) où si il a été synthétisé (généré par un procédé informatique).
Positionner le curseur au milieu signifie que le son epu aussi bien être naturel que synthétique.
Si vous ne savez pas, cochez la case en dessous.

Question 3:



Cocher la case qui convient (une seule réponse possible)

Question 4:

Le son évoque-t-il malgré tout un mouvement?	☐ Oui	☐ Non	☐ Je ne sais pas

• *Si le son n'évoque pas le déplacement d'un objet*, on veut savoir si le son vous évoque un mouvement abstrait ou qui n'est pas celui d'un objet réel.*

*Il n'est pas possible de donner une réponse à cette question si vous avez répondu Oui à la question 3.

Question 5:

Est-ce que l'objet produit du son indépendament de ce déplacement?	☐ Oui	☐ Non	☐ Je ne sais pas

• *En d'autres termes, il est demandé si selon vous, le son est produit par un ou plusieurs objet(s) qui interagissent avec l'environnement lors de leur déplacement.*
Par exemple une bille qui rebondit et roule sur le sol.

Il n'est pas possible de donner une réponse à cette question si vous avez répondu Non à la question 3.

Question 6:

Le son évoque-t-il un objet animé d'un mouvement interne?	☐ Oui	☐ Non	☐ Je ne sais pas

• *On entend par "mouvement interne" un mouvement propre à l'objet.*
Par exemple, une pendule ne se déplace pas, mais son balancier a bien un mouvement qui est à l'origine du son produit

Il n'est pas possible de donner une réponse à cette question si vous avez répondu Oui à la question 3.

Question 6 (bis):

Ce mouvement est-il:	Régulier	Aléatoire

• *Si vous avez répondu Oui à la question 6, on vous demande de préciser si ce mouvement interne est régulier ou non*

Question 7:

Vous auriez pu produire ce son? (par un geste)	Impossible [ ] Très probable

• *On cherche à savoir si c'est un son que vous auriez pu produire (par exemple en frottant un objet) et si vous vous imaginez en train de produire ce son en l'écoutant.*

Il n'est pas possible de donner une réponse à cette question si vous avez répondu Non à la question 3.

Question 8:

Ce son aurait pu induire une réaction de votre part?	Impossible [ ] Très probable

• On cherche à savoir si la trajectoire de l'objet aurait pu vous inciter à réagir (par exemple se déplacer pour éviter d'être percuté par l'objet) ou effectuer n'importe quelle autre action.

Auto-évaluation:

Evaluez la difficulté que vous avez eut à répondre aux questions:	Très difficile [ ] Très facile

• *Finalement, on demande de juger si vous avez beaucoup hésité pour répondre aux questions ou si au contraire vous avez répondu facilement et que vous n'avez pas de doutes sur vos réponses.*

Quelques conseils:

• *Essayez d'utiliser toute l'échelle de valeur*

• *Vous pouvez revenir en arrière pour vérifier une réponse que vous avez donné et réecouter un son.*

• *Essayez de donner une réponse la plus spontannée possible.*

• *Ne passez pas trop de temps sur une question (vous pourrez y revenir plus tard)*

• *Veillez à bien répondre à toutes les questions!*

Questionnaire en complément du **Test 3**

N° : Date :

- Renseignements personnels

Age :

Sexe : F / M

Niveau d'étude :

Avez-vous des problèmes auditifs ? oui - non

Si oui, précisez de quelle nature :

Vous jouez d'un instrument de musique ? Non - Débutant - Amateur - Professionnel

Si oui, lequel :

Est-ce que votre profession a un lien avec l'acoustique ou la synthèse sonore ? oui - non

Est-ce la première fois que vous participez à une expérience de ce type ? oui - non

- A propos du test

1) Pensez-vous que vos réponses ont évoluées au cours du test ? oui - non

2) Pensez-vous que nous vous avons fait évaluer plusieurs fois le même son ? oui - non

3) Avez-vous trouvé le test trop long ? oui - non

4) Avez-vous trouvé le test globalement difficile ? oui - non

Si oui pourquoi ?

5) Avez-vous trouvé une (ou plusieurs) question(s) particulièrement difficile(s) ? oui - non

Si oui précisez la(les) quelle(s) ?

6) Avez-vous eut envie de répondre à une question qui n'était pas posée ? oui - non

Si oui précisez :

Si vous souhaitez être tenu au courant des résultats de cette étude notez votre adresse mail :

.....@.....

Commentaires :

B 2 Test 4

Consignes du Test 4

Vous allez participer à une expérience visant à décrire l'information sur le déplacement d'objets que l'on peut percevoir en écoutant des sons.
Le test commence par une phase d'entraînement en deux parties où il est simplement demandé de reproduire des formes et déplacements pour vous familiariser avec l'interface.
Dans un second temps, vous aller avoir à juger 40 sons avec cette interface.

0.1 Entraînement 1 (Forme)

Les curseurs de la partie gauche permettent de modifier la forme qui est dessinée au centre de l'écran. Le test consiste à reproduire la forme qui vous est proposée sur la partie droite de l'écran à l'aide de ces curseurs.
Une fois que vous êtes satisfait de la forme obtenue, cliquez sur "*Suivant*" pour passer à un autre exemple.
Sur le bord de certains curseurs, il est possible de cliquer sur les petits traits noirs pour que le curseur se place exactement à l'endroit du trait (également valable pour les mots sous le curseur de forme).
Prenez le temps nécessaire pour explorer toutes les possibilités de l'interface.
Aucun son n'est donné à entendre, le but est simplement de se familiariser avec les paramètres de dessin.

0.2 Entrainement 2 (Position)

Les curseurs mis à votre disposition, permettent de changer la position de la boule au cour du temps.
Il s'agit cette fois de reproduire le déplacement de la boule présentée sur l'animation de la partie droite de l'écran.
Cliquez sur "*Play*" pour lancer l'animation, ainsi que sur "*Suivant*" pour changer d'exemple.
Cette fois encore aucun son n'est donné à entendre.

0.3 Test principal

- On vous demande avant tout de décrire globalement la trajectoire que vous avez perçu c'est à dire la trajectoire moyenne).
-Ne vous intéressez pas à la distance entre la source et l'observateur, l'interface ne permet pas de la décrire.
- Vous pouvez réécouter chaque son autant de fois que vous le souhaitez.
- L'auto-évaluation de vos réponses est aussi importante que le reste, elle vous permet de dire, le cas échéant si vous n'avez pas put répondre correctement.
- Vous pouvez revenir sur les sons que vous avez déjà renseignés pour éventuellement modifier votre réponse.
- Si vous voulez faire une pause, cliquez sur le bouton pause en bas à droite. - Prenez le temps de donner une réponse qui décrit correctement ce que vous entendez sans pour autant "bloquer" sur un son qui pose problème.
- Il n'y a pas de bonnes ou de mauvaise réponse, vous êtes seul juge de la qualité de votre réponse.
A la fin du test, vous aurez à remplir un rapide questionnaire où vous pourrez noter vos remarques.

Merci pour votre participation.

Questionnaire en complément du **Test 4**

N° : Date :

1) Pensez-vous que vos réponses ont évoluées au cours du test ? oui - non

2) Avez-vous trouvé le test trop long ? oui - non

3) Avez-vous trouvé le test difficile ? oui - non
Si oui pourquoi ?

4) Manquait-il des paramètres pour pouvoir répondre correctement ? oui - non
Si oui précisez :

5) Certains paramètres sont-ils trop limités (ex : Vous auriez souhaité faire des trajectoires plus grandes, avoir plus de bosses,...) ? oui - non
Si oui précisez :

6) Pensez-vous que le test 1 a influencé vos réponses à ce test ? oui - non
Si oui comment ?

7) Pensez-vous que nous vous avons fait évaluer plusieurs fois le même son ? oui - non

Commentaires :

C Mots utilisés pour décrire des sons abstraits dans le Test 3

L'ordre des mots suivant les lignes des tableaux n'est pas significatif

1	plaque dans une salle réverbérante	quelque chose à l'envers	instruments qui s'accordent	alarme	cloche	sirène de bateau	orgue								
2	de la voix bidouillée, ralentie, trafiquée														
3	route														
4	rebondissement sur grosse plaque de métal	bruit de tà´le	une plaque de métal qui tombe	chute d un bidon	métal	timbale	casserole	Plaque métal	tôle	explosion lointaine					
5	cousin des cornets de rara ou peau frottée	chaise ou table qu'on traine contre le sol	un frottement sur bois	frottement	bruit de chaise	Comme si on frottait un son de clarinette par terre	une chaise en bois sur du carrelage	instrument à vent	wagon	frottement	instrument	ponceuse	chaise	bois	klaxon
6	un rebond de diapason sur une corde en nylon	récipient se remplissant	remplir												
7	parasites sur un CD mal gravé	moteur à vapeur	un train au ralenti	machine à vapeur	broyeur	helico	hélice	voiture	moteur	Moteur	moteur	hélicoptère	machine ou mitrailleuse		
8	une mouche bionique	moteur	insecte	parasite radiophonique											
9	moteur	rasoir ou tondeuse électrique	appareil pour limer												
10	pluie d'astéroïde dans mon prochain film	oiseaux													
11	Un chien qui saute et aboie en même temps	meuble que l'on déplace	scie	un frottement	frottement	couper	feulement	cri animal							
12	fermeture	feu artifice	vaisseau spatial												
13	son trafiqué d'une boite d'allumettes qu'on secoue	pièce de monnaie	une pièce de monnaie qui tourne	couvercle et sac de sable											
14	une déchirure	orage	radio												
15	un chat virtuel pas content	un miaulement de chat	scie circulaire	chat	rugissement	tigre	Cymbale	panthère	chat	scie électrique	chat pas content				
16	des gouttes de mercure tombant dans un saladier en platine	rebonds de billes dans une coupelle	billes sur du verre (son un peu déformé)	billes tombant dans un bol e métal	billes	cassure d'un filament	rebondissement sur métal	orchestration sable et bille	tomber	pluie	bille				
17	une note de piano qui se casse en deux ou alors un coup de pistolet superposé à une note de piano dans les aigus	clavecin pour enfant	verre se brisant	marteau sur une pierre	machine à casser des pierres	animal	choc métallique								
18	structure métallique	sonnerie ?	larsen	roue métal	cloche	réveil	Verre cristal	disqueuse	perceuse	résonance					
19	bisous de grand mère en zoom avant sonore	star wars	Platine vinyle												
20	vinyle en reverse	fuites d'un liquide													
21	sifflet intergalactique	hélicoptère en plastique	insecte	oiseau	insecte										

22	guimbarde bien trafiquée augmentée de percussions	voix déformée ?	objet dans tuyau	musique	djreedoo	didge	voix humaine brouillée par électronique								
23	balle de tennis tombant dans un baril	membrane plastique	bidon	taper											
24	une croquette pour chien en train d'être mangée au ralenti	des clous dans une boite en plastique	galets dans un seau	un plastique froissé	lion	bassine de gravier	lion	Planche à roulettes	eau	papier					
25	plaque frottement	Un son de tymabale coupée dans son élan	souffle dans un tube	une voiture	plaque métal	ventilation	cloche								
26	un ballon de baudruche qui se dégonfle	frottement plastique		Braguette	violent suçon sonore										
27	très grande porte rouillée du manoir	La corne d'un bateau dans un port avec de la brume.	machine industrielle	camion	poids lourd	cloche	usine souterraine	sirène							
28	un tamagoschi qui essaye de parler	frottement sur une surface plastique rayée	crapoïde	fermeture éclair											
29	arrêt d'un hautparleur	zip vers le bas		sirène	jingle										
30	train qui freine	Scie électrique	réacteur	machine	machine	hélicoptère	machine industrielle								
31	mélange entre un \Ah\ criée et un zip remonté lentement suivi par un clic d'appareil photo	un peu appareil photo électronique	guiro	genre de crécelle	ressort qui se détend	grenouille	oiseau	machine à écrire bizarre	frottement	téléphone	grenouille				
32	résonance gong ou cloche	objet métallique vibrant	gong	un signal inversé	machine en métal	musique									
33	transformateur	sirène	résonance												
34	Un couvercle en fer sur lequel on appuie puis un son très clair	boite métallique	.caisse claire synthétique?	tà´le bombée se remettant en place	caisse claire	clapet	percussion	instrument	Caisse claire	caisse claire	casserole	caisse claire de batterie	caisse claire		
35	tige en plastique sur le bord d'une table	corde de guitare pincée	corde	guimbarde	corde	guitare	élastique tendu	corde	corde en tension	corde pincée	corde	Elastique	instrument	système à ressort	élastique
36	une sorte de tambour (type machine à laver, pas l'instrument)	objet sur une peau de tambour	roue	train dans un tunnel	train	machine a laver	moteur d'un monstre hybride irradié	ventilation dans tunnel							
37	clave pecu inconnue	bille dans quelque chose	un objet qui tombe	Poubelle											
38	un engin volant imaginaire au décollage	bruit aérodynamique	voiture de course	avion	réacteur d'avion	souffle d'un objet dans l'espace..	réacteur	vent		décollage	avion	Avion	décollage	vent	
39	j'entend une membrane et une corde. une sorte d'instrument de musique	berimbau ou canard très énervé!	berimbao			didjiridoo couplée à une guimbarde	Corde	instrument à corde primitif	berimbau						
40	Chasse d'eau	chasse d'eau	siphon												

D Liste des publications de l'auteur

Articles de revues à comité de lecture

[1] A. Merer, M. Aramaki, S. Ystad, and R. Kronland-Martinet. Perceptual characterization of motions evoked by sound for synthesis control. *Submitted for publication in IEEE transaction on audio speech and language processing*, 2011. 58, 86

[2] Jean-Arthur Micoulaud-Franchi, Mitsuko Aramaki, Adrien Merer, Michel Cermolacce, Solvi Ystad, Richard Kronland-Martinet, and Jean Vion-Dury. Categorization and timbre perception of environmental sounds in schizophrenia. *Psychiatry Research*, In Press, Corrected Proof:–, 2011. 103

[3] J. A. Micoulaud-Franchi, M. Aramaki, A. Merer, M. Cermolacce, S. Ystad, R. Kronland-Martinet, and J. Vion-Dury. Toward an Exploration of "Feeling of Strangeness" in Schizophrenia: Perspectives from "Acousmatic" and "Everyday" Listening. *Submitted for publication in Journal of Abnormal Psychology*, 2011. 103

Chapitres de livres

[4] A. Merer, S. Ystad, R. Kronland-Martinet, and M. Aramaki. Semiotics of sounds evoking motions: Categorization and acoustic features. In R. Kronland-Martinet, S. Ystad, and K. Jensen, editors, *CMMR 2007. Sense of Sounds*, pages 139–158. Springer, LNCS, 2008.
Version enrichie de [6]

[5] A. Merer, S. Ystad, R. Kronland-Martinet, and M. Aramaki, *Abstract sounds and their applications in audio and perception research*, CMMR 2010 Post Symposium Proceedings, Lecture Notes in Computer Science, vol. 6684, Springer Berlin / Heidelberg, 2011.
Version enrichie de [7]

Workshops et Conférences internationales

[6] A. Merer, M. Aramaki, S. Ystad, R. Kronland-Martinet, M. Besson, and J.L Velay. Perceptual categorization of moving sounds for synthesis applications. In *Proceedings the International Computer Music Conference*, pages 69–72, 2007. 123

[7] A. Merer, M. Aramaki, R. Kronland-Martinet, and S. Ystad. On the potentiality of abstract sounds in perception research. In *Proceedings of CMMR 2010, Málaga, Spain*, pages 207–219, 2010. 123

Conférence nationale

[8] P. Guillemain and A. Merer. Rôles du contrôle et du timbre dans la perception du naturel de sons de clarinette. Société Française d'Acoustique (SFA), *10ème Congrès Français d'Acoustique*, Lyon France, 2010. 104

Bibliographie

[AFNOR, 77] AFNOR, éditeur. *Recueil des normes françaises de l'acoustique, Tome 1 (vocabulaire), NF S 30-107.* Association Française de NORmalisation. (1977). 32

[Agaeva, 04] Agaeva, M. Velocity discrimination of auditory image moving in vertical plane. *Hearing Research*, 198(1-2) p.1 – 9 (2004). 23

[Aglioti et Pazzaglia, 10] Aglioti, S. et Pazzaglia, M. Representing actions through their sound. *Experimental Brain Research*, 206 p.141–151 (2010). 25

[Agon *et al.*, 00] Agon, C., Stroppa, M. et Assayag, G. High level musical control of sound synthesis in OpenMusic. *In Proceedings of the 2000 International Computer Music Conference*, p.332–335. Citeseer (2000). 29

[Al-Hindawe, 96] Al-Hindawe, J. Considerations when constructing a semantic differential scale. Rapport technique, La Trobe Papers in Linguistics (Volume 9) (1996). 29

[Alaerts *et al.*, 09] Alaerts, K., Swinnen, S. P. et Wenderoth, N. Interaction of sound and sight during action perception: Evidence for shared modality-dependent action representations. *Neuropsychologia*, 47(12) p.2593 – 2599 (2009). 61

[Alais et Burr, 04] Alais, D. et Burr, D. The ventriloquist effect results from near-optimal bimodal integration. *Current biology : CB*, 14(3) p.257–262 (2004). 61

[Andrade, 59] Andrade, E. Doppler and the doppler effect. *Endeavour*, 18 p.14–19 (1959). 20

[ANSI, 60] ANSI, éditeur. *USA Standard Acoustical Terminology.* American National Standards Institute, New York (1960). 32

[APA, 94] APA. *The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV).* American Psychiatric Association (1994). 102

[Aramaki *et al.*, 11] Aramaki, M., Besson, M., Kronland-Martinet, R. et Ystad, S. Controlling the perceived material in an impact sound synthesizer. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(2) p.301–314 (2011). 7, 33

[Arfib *et al.*, 02] Arfib, D., Couturier, J. M., Kessous, L. et Verfaille, V. Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. *Organised Sound*, 7(02) p.127–144 (2002). 29

[Aucouturier *et al.*, 07] Aucouturier, J., Defreville, B. et Pachet, F. The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music. *The Journal of the Acoustical Society of America*, 122 p.881 (2007). 22

[Bach *et al.*, 09] Bach, D. R., Neuhoff, J. G., Perrig, W. et Seifritz, E. Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology*, 74(1) p.28 – 33 (2009). 26

[Ballas, 93] Ballas, J. A. Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2) p.250 – 267 (1993). 6, 105

[Barthet, 08] Barthet, M. *De l'interprète à l'auditeur: une analyse acoustique et perceptive du timbre musical.* Thèse de doctorat, Université Aix-Marseille II (2008). 29, 50

[Beer et Röder, 04] Beer, A. L. et Röder, B. Attention to motion enhances processing of both visual and auditory stimuli: an event-related potential study. *Cognitive Brain Research*, 18(2) p.205 – 225 (2004). 24

[Bensa *et al.*, 04] Bensa, J., Dubois, D., Kronland-Martinet, R. et Ystad, S. Perceptive and cognitive evaluation of a piano synthesis model. *In* Springer, éditeur : *Computer music modeling and retrieval*, p.232–245 (2004). 5

[Berger, 64] Berger, K. W. Some factors in the recognition of timbre. *The Journal of the Acoustical Society of America*, 36(10) p.1888–1891 (1964). 21

[Bergman *et al.*, 09] Bergman, P., Skold, A., Vastfjall, D. et Fransson, N. Perceptual and emotional categorization of sound. *The Journal of the Acoustical Society of America*, 126(6) p.3156–3167 (2009). 41

[Bevilacqua *et al.*, 05] Bevilacqua, F., Müller, R. et Schnell, N. Mnm: a max/msp mapping toolbox. *In NIME '05: Proceedings of the 2005 conference on New interfaces for musical expression*, p.85–88, Singapore, Singapore. National University of Singapore (2005). 30

[Bezat, 07] Bezat, M. C. *Perception des bruits d'impact Application au bruit de fermeture de porte automobile*. Thèse de doctorat, Université de Provence - Aix-Marseille 1 (2007). 7

[Bishop, 06] Bishop, C. M. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer (2006). 101

[Bismarck, 74] Bismarck, G. Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica*, 30(3) p.146–159 (1974). 103

[Blake et Shiffrar, 07] Blake, R. et Shiffrar, M. Perception of human motion. *Annu. Rev. Psychol*, 58 p.47–73 (2007). 18

[Bloit *et al.*, 09a] Bloit, J., Rasamimanana, N. et Bevilacqua, F. Modeling and segmentation of audio descriptor profiles with segmental models. *Pattern Recognition Letters*, In Press, Corrected Proof p.– (2009). 22

[Bloit *et al.*, 09b] Bloit, J., Rasamimanana, N. et Bevilacqua, F. Towards morphological sound description using segmental models. *In DAFX*, Milan, Italie (2009). 108

[Blum et Langley, 97] Blum, A. et Langley, P. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1-2) p.245–271 (1997). 53, 54

[Bootz et Hautbois, 08] Bootz, P. et Hautbois, X. Les motifs temporels paramétrés. *In* Rix, E. et Formosa, M., éditeurs : *Vers une sémiotique générale du temps dans les arts, Actes du colloque "Les unités sémiotiques temporelles (UST), nouvel outil d'analyse musicale: Théories et applications Marseille décembre 2005*. Editions Delatour / Ircam (2008). 108

[Bossis, 06] Bossis, B. The Analysis of Electroacoustic Music: from sources to invariants. *Organised Sound*, 11(02) p.101–112 (2006). 36

[Bowler *et al.*, 90] Bowler, I., Purvis, A., Manning, P. et Bailey, N. On mapping n articulation onto m synthesiser-control parameters. *In Proc. of the 1990 International Computer Music Conference. San Francisco, Calif.: International Computer Music Association*, p.181–184 (1990). 30

[Bradley et Lang, 94] Bradley, M. M. et Lang, P. J. Measuring emotion: the self-assessment manikin and the semantic differential. *J Behav Ther Exp Psychiatry*, 25(1) p.49–59 (1994). 109

[Bregman, 90] Bregman, A. S. *Auditory Scene Analysis*. The MIT Press (1990). 67

[Brovelli *et al.*, 02] Brovelli, A., Battaglini, P., Naranjo, J. et Budai, R. Medium-range oscillatory network and the 20-Hz sensorimotor induced potential. *Neuroimage*, 16(1) p.130–141 (2002). 24

[Brown et Brewster, 03] Brown, L. et Brewster, S. Drawing by ear: Interpreting sonified line graphs. *In Proceedings of ICAD 2003*, p.152–156 (2003). 60

[Cadoz *et al.*, 82] Cadoz, C., Luciani, A., Florens, J. et Dars-Berberyan, T. The control channels of instrumental playing in computer music - real time in computer music, incidence of basic models. *In Proceedings of the 1982 International Computer Music Conference (ICMC), Venise*, p.73–91 (1982). 23

[Caetano *et al.*, 05] Caetano, M., Manzolli, J. et Von Zuben, F. Interactive control of evolution applied to sound synthesis. *Proceedings of the 18th International Florida Artificial Intelligence Research Society (FLAIRS), Clearwater Beach, Florida, EUA*, p.51–56 (2005). 31

[Caramiaux *et al.*, 09] Caramiaux, B., Bevilacqua, F. et Schnell, N. Towards a gesture-sound cross-modal analysis. *In Gesture in Embodied Communication and Human-Computer Interaction, Lecture Notes in Computer Science (LNCS)*. Springer Verlag (2009). 20

[Carlile et Best, 02] Carlile, S. et Best, V. Discrimination of sound source velocity in human listeners. *The Journal of the Acoustical Society of America*, 111(2) p.1026–1035 (2002). 23

[Cermolacce *et al.*, 10] Cermolacce, M., Sass, L. et Parnas, J. What is Bizarre in Bizarre Delusions ? A Critical Review. *Schizophrenia Bulletin*, 36(4) p.667–679 (2010). 102

[Cheng et Wakefield, 01] Cheng, C. et Wakefield, G. Moving sound source synthesis for binaural electroacoustic music using interpolated head-related transfer functions (HRTFs). *Computer Music Journal*, 25(4) p.57–80 (2001). 22

[Choi *et al.*, 95] Choi, I., Bargar, R. et Goudeseune, C. A manifold interface for a high dimensional control space. *In Proceedings of the 1995 International Computer Music Conference*, p.385–392 (1995). 29, 30

[Chowning, 71] Chowning, J. The simulation of moving sound sources. *Journal of the Audio Engineering Society*, p.J. Audio Eng. Soc. 19 (1), 2–6. (1971). 21, 22, 90

[Cossette *et al.*, 00] Cossette, I., Sliwinski, P. et Macklem, P. Respiratory parameters during professional flute playing. *Respiration physiology*, 121(1) p.33–44 (2000). 19

[Cross, 68] Cross, L. Electronic music, 1948-1953. *Perspectives of New Music* (1968). 34

[Dack, 02] Dack, J. Abstract and concrete. *Journal of Electroacoustic Music*, 14 (2002). 35

[Dahlstedt, 01] Dahlstedt, P. Creating and exploring huge parameter spaces: Interactive evolution as a tool for sound generation. *In Proceedings of the 2001 International Computer Music Conference*, p.235–242 (2001). 31

[Desainte Catherine et Marchand, 99] Desainte Catherine, M. et Marchand, S. Structured Additive Synthesis: Towards a Model of Sound Timbre and Electroacoustic Music Forms. *In Proceedings of the International Computer Music Conference (ICMC99)*, p.260–263, Chine (1999). 32

[Dixon, 06] Dixon, S. Onset detection revisited. *In Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, p.133–137 (2006). 22, 51

[Dolson, 86] Dolson, M. The phase vocoder: A tutorial. *Computer Music Journal*, p.14–27 (1986). 42

[Dubois, 00] Dubois, D. Categories as acts of meaning: The case of categories in olfaction and audition. *Cognitive Science Quarterly*, 1(1) p.35–68 (2000). 40

[Dubois *et al.*, 06] Dubois, D., Guastavino, C. et Raimbault, M. A cognitive approach to urban soundscapes: Using verbal data to access everyday life auditory categories. *Acta Acustica united with Acustica*, 92(6) p.865–874(10) (2006). 37

[Eimert, 57] Eimert, H. What is electronic music. *Die Reihe*, 1 (1957). 34

[Eitan et Granot, 06] Eitan, Z. et Granot, R. Y. How music moves: musical parameters and listeners images of motion. *music perception*, 23(3) p.221–247 (2006). 17

[Ekman *et al.*, 82] Ekman, P., Friesen, W. et Ellsworth, P. *Emotion in the human face*. Cambridge University Press New York (1982). 109

[Evans, 98] Evans, M. J. Synthesising moving sounds. *In* Brewster, S., éditeur : *Proceedings of the 5th International Conference on Auditory Display (ICAD98)*, University of Glasgow, U.K. British Computer Society (1998). 22

[Farbood *et al.*, 07] Farbood, M., Kaufman, H. et Jennings, K. Composing with hyperscore: An intuitive interface for visualizing musical structure. *In Proc. of ICMC* (2007). 59

[Farner *et al.*, 06] Farner, S., Kronland-Martinet, R., Voinier, T. et Ystad, S. Timbre variations as an attribute of naturalness in clarinet play. *In* Kronland-Martinet, R., Voinier, T. et Ystad, S., éditeurs : *Computer Music Modeling and Retrieval*, volume 3902 de *Lecture Notes in Computer Science*, p.45–53. Springer Berlin / Heidelberg (2006). 104

[Farrer *et al.*, 03] Farrer, C., Franck, N., Paillard, J. et Jeannerod, M. The role of proprioception in action recognition. *Consciousness and Cognition*, 12(4) p.609–619 (2003). 25

[Fastl, 01] Fastl, H. Neutralizing the meaning of sound for sound quality evaluations. *In Proc. Int. Congress on Acoustics ICA 2001, Rome, Italy*, volume 4 CD-ROM (2001). 36

[Fayyad et Irani, 93] Fayyad, U. et Irani, K. Multi-interval discretization of continuous-valued attributes for classification learning. *In Proceedings of the International Joint Conference on Uncertainty in AI*, p.1022–1027 (1993). 54

[Feldman *et al.*, 92] Feldman, J., Epstein, D. et Richards, W. Force dynamics of tempo change in music. *Music Perception*, p.185–203 (1992). 17

[Fremiot *et al.*, 96] Fremiot, M., Mandelbrojt, J., Formosa, M., Delalande, G., Pedler, E., P.Malbosc et Gobin, P. *Les Unités Sémiotiques Temporelles: éléments nouveaux d'analyse musicale*. diffusion ESKA. MIM Laboratoire Musique et Informatique de Marseille (1996), documents musurgia édition. 13

[Friberg et Sundberg, 99] Friberg, A. et Sundberg, J. Does music performance allude to locomotion ? a model of final ritardandi derived from measurements of stopping runners. *The Journal of the Acoustical Society of America*, 105(3) p.1469–1484 (1999). 16

[Friberg *et al.*, 00] Friberg, A., Sundberg, J. et Frydén, L. Music from Motion: Sound Level Envelopes of Tones Expressing Human Locomotion. *Journal of New Music Research*, 29(3) p.199 – 210 (2000). 16

[Garcia, 01] Garcia, R. A. Growing sound synthesizers using evolutionary methods. *In In Proceedings ALMMA 2001: Artificial Life Models for Musical Applications Workshop*, p.99–107 (2001). 31, 110

[Garcia, 02] Garcia, R. A. Automatic design of sound synthesis techniques by means of genetic programming. *In Audio Engineering Society Convention 113* (2002). 31, 110

[Gauchi, 05] Gauchi, J.-P. Plans d'expériences optimaux: un exposé didactique. *Revue MODULAD*, 33 p.139 – 162 (2005). 39

[Gaver, 93a] Gaver, W. W. How do we hear in the world ? explorations of ecological acoustics. *Ecological Psychology*, 5(4) p.285 – 313 (1993). 35, 105

[Gaver, 93b] Gaver, W. W. What in the world do we hear ? an ecological approach to auditory source perception. *Ecological Psychology*, 5(1) p.1–29 (1993). 33, 105

[Getzmann et Lewald, 09] Getzmann, S. et Lewald, J. Effects of natural versus artificial spatial cues on electrophysiological correlates of auditory motion. *Hearing Research*, In Press, Corrected Proof p.– (2009). 24

[Gibet, 87] Gibet, S. *Codage, représentation et traitement du geste instrumental: application à la synthèse de sons musicaux par simulation de mécanismes instrumentaux*. Thèse de doctorat, Institut de Mathématiques Appliquées de Grenoble (1987). 23

[Gibson, 78] Gibson, J. J. The ecological approach to the visual perception of pictures. *Leonardo*, 11(3) p.227–235 (1978). 35, 98

[Glasberg et Moore, 02] Glasberg, B. et Moore, B. A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50(5) p.331–342 (2002). 67

[Gobin *et al.*, 04] Gobin, P., Kronland-Martinet, R., Lagesse, G., Voinier, T. et Ystad, S. Designing musical interfaces with composition in mind. *Computer Music Modeling and Retrieval*, p.33–61 (2004). 30

[Godøy, 09] Godøy, R. Chunking sound for musical analysis. *In* Ystad, S., Kronland-Martinet, R. et Jensen, K., éditeurs : *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music*, volume 5493 de *Lecture Notes in Computer Science*, p.67–80. Springer Berlin / Heidelberg (2009). 108

[Godøy, 06] Godøy, R. I. Gestural-sonorous objects: embodied extensions of schaeffer's conceptual apparatus. *Organised Sound*, 11(02) p.149–157 (2006). 20

[Godøy *et al.*, 10] Godøy, R. I., Jensenius, A. R. et Nymoen, K. Chunking in music by coarticulation. *Acta Acustica united with Acustica*, 96 p.690–700(11) (July/August 2010). 108

[Goldstone et Kersten, 03] Goldstone, R. et Kersten, A. Concepts and categorization. *In* Healy, A. F. et Proctor, R. W., éditeurs : *Comprehensive handbook of psychology*, volume 4, p.599–621. New Jersey: Wiley (2003). 41

[Gounaropoulos et Johnson, 06] Gounaropoulos, A. et Johnson, C. Synthesising timbres and timbre-changes from adjectives/adverbs. *Applications of Evolutionary Computing*, p.664–675 (2006). 32, 33, 109

[Gouyon et Dixon, 05] Gouyon, F. et Dixon, S. A review of automatic rhythm description systems. *Computer Music Journal*, 29(1) p.34–54 (2005). 52

[Gracco et Abbs, 86] Gracco, V. L. et Abbs, J. H. Variant and invariant characteristics of speech movements. *Experimental Brain Research*, 65 p.156–166 (1986). 7

[Grantham, 89] Grantham, D. Motion aftereffects with horizontally moving sound sources in the free field. *Perception & psychophysics*, 45(2) p.129 (1989). 25

[Grantham, 86] Grantham, D. W. Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *The Journal of the Acoustical Society of America*, 79(6) p.1939–1949 (1986). 23

[Grantham, 98] Grantham, D. W. Auditory motion aftereffects in the horizontal plane: The effects of spectral region, spatial sector, and spatial richness. *Acta Acustica united with Acustica*, 84 p.337–347(11) (March/April 1998). 25

[Grey et Gordon, 78] Grey, J. M. et Gordon, J. W. Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, 63(5) p.1493–1500 (1978). 21

[Griffiths *et al.*, 00] Griffiths, T., Green, G., Rees, A. et Rees, G. Human brain areas involved in the analysis of auditory movement. *Human Brain Mapping*, 9(2) p.72–80 (2000). 24

[Griffiths *et al.*, 08] Griffiths, T., Kemp, C. et Tenenbaum, J. Bayesian models of cognition. *Cambridge handbook of computational cognitive modeling*, p.59–100 (2008). 106

[Griffiths *et al.*, 98] Griffiths, T., Rees, G., Rees, A., Green, G., Witton, C., Rowe, D., Büchel, C., Turner, R. et Frackowiak, R. Right parietal cortex is involved in the perception of sound movement in humans. *nature neuroscience*, 1(1) p.74–79 (1998). 24

[Groux et Verschure, 08] Groux, S. L. et Verschure, P. F. M. J. Perceptsynth: Mapping perceptual musical features to sound synthesis parameters. *In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, USA (2008). 32

[Guastavino, 07] Guastavino, C. Categorization of environmental sounds. *Can J Exp Psychol*, 61(1) p.54–63 (2007). 37, 41

[Guillemain et Merer, 10] Guillemain, P. et Merer, A. Rôles du contrôle et du timbre dans la perception du naturel de sons de clarinette. *In* Société Française d'Acoustique SFA, éditeur : *10ème Congrès Français d'Acoustique 10ème Congrès Français d'Acoustique*, Lyon France (2010). 104

[Guyon et Elisseeff, 03] Guyon, I. et Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3 p.1157–1182 (2003). 54

[Gygi *et al.*, 04] Gygi, B., Kidd, G. et Watson, C. Spectral-temporal factors in the identification of environmental sounds. *Journal of the acoustical society of america*, 115(3) p.1252–1265 (2004). 6

[Gygi *et al.*, 07] Gygi, B., Kidd, G. et Watson, C. Similarity and categorization of environmental sounds. *Perception & psychophysics*, 69(6) p.839 (2007). 41

[Harley, 94] Harley, M. A. Spatial sound movement in the instrumental music of Iannis Xenakis. *Journal of New Music Research*, 23(3) p.291 – 314 (1994). 16

[Harris et Sergeant, 71] Harris, J. D. et Sergeant, R. L. Monaural/Binaural Minimum Audible Angles for a Moving Sound Source. *J Speech Hear Res*, 14(3) p.618–629 (1971). 23

[Hauk *et al.*, 06] Hauk, O., Shtyrov, Y. et Pulvermüller, F. The sound of actions as reflected by mismatch negativity: rapid activation of cortical sensory–motor networks by sounds associated with finger and tongue movements. *European Journal of Neuroscience*, 23(3) p.811–821 (2006). 24

[Helmholtz, 75] Helmholtz, H. L. F. *Sensation of tone as a physiological basis for the theory of music*. Longmans green and co (1875). 16, 21

[Herrera *et al.*, 02] Herrera, P., Yeterian, A. et Gouyon, F. Automatic classification of drum sounds: a comparison of feature selection methods and classification techniques. *In Proceedings of Second International Conference on Music and Artificial Intelligence*, Edinburgh, Scotland (2002). 53

[Hoffman et Cook, 06] Hoffman, M. et Cook, P. R. Feature-based synthesis: Mapping acoustic and perceptual features onto synthesis parameters. *In Proceedings of the 2006 International Computer Music Conference (ICMC), New Orleans* (2006). 33

[Hoffman et Cook, 07] Hoffman, M. et Cook, P. R. Real-time feature-based synthesis for live musical performance. *In NIME '07: Proceedings of the 7th international conference on New interfaces for musical expression*, p.309–312 (2007). 33

[Honing, 03] Honing, H. The final ritard: On music, motion, and kinematic models. *Computer Music Journal*, 27(3) p.66–72 (2003). 16, 17

[Horner *et al.*, 93] Horner, A., Beauchamp, J. et Haken, L. Machine tongues XVI: Genetic algorithms and their application to FM matching synthesis. *Computer Music Journal*, 17(4) p.17–29 (1993). 30

[Houben *et al.*, 04] Houben, M., Kohlrausch, A. et Hermes, D. Perception of the size and speed of rolling balls by sound. *Speech communication*, 43(4) p.331–345 (2004). 24

[Houix, 03] Houix, O. *Catégorisation auditive de sources sonores*. Thèse de doctorat, Université du Maine (2003). 7

[Howard *et al.*, 07] Howard, D. M., Disley, A. C. et D., H. A. Timbral adjectives for the control of a music synthesizer. *In Proceedings of the International Congres on Acoustics ICA, Madrid Spain* (2007). 33

[Hunt *et al.*, 00] Hunt, A., Wanderley, M. et Kirk, R. Towards a model for instrumental mapping in expert musical interaction. *In Proceedings of the 2000 International Computer Music Conference*, p.209–212 (2000). 29

[Hunt et Wanderley, 02] Hunt, A. et Wanderley, M. M. Mapping performer parameters to synthesis engines. *Organised Sound*, 7(02) p.97–108 (2002). 30

[Hunt *et al.*, 02] Hunt, A., Wanderley, M. M. et Paradis, M. The importance of parameter mapping in electronic instrument design. *In NIME '02: Proceedings of the 2002 conference on New interfaces for musical expression*, p.1–6, Singapore, Singapore. National University of Singapore (2002). 5

[Iwaya et Suzuki, 07] Iwaya, Y. et Suzuki, Y. Rendering moving sound with the doppler effect in sound space. *Applied Acoustics*, 68(8) p.916–922 (2007). 22

[Jehan et Schoner, 01a] Jehan, T. et Schoner, B. An audio-driven perceptually meaningful timbre synthesizer. *In Proceedings of the 2001 International Computer Music Conference* (2001). 32

[Jehan et Schoner, 01b] Jehan, T. et Schoner, B. An audio-driven, spectral analysis-based, perceptual synthesis engine. *In Audio Engineering Society Convention 110* (2001). 32

[Jekosch, 05] Jekosch, U. *Assigning Meaning to Sounds - Semiotics in the Context of Product-Sound Design*, chapitre 8, p.193–221. J. Blauert (2005). 6

[Jenison *et al.*, 98] Jenison, R., Neelon, M., Reale, R. et Brugge, J. Synthesis of virtual motion in 3d auditory space. *In Proceedings of the 20th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society* (1998). 22

[Jensen, 99] Jensen, K. *Timbre models of musical sounds*. Thèse de doctorat, University of Copenhaguen, Copenhague, Denmark (1999). 32

[Jensen, 05] Jensen, K. Atomic noise. *Organised Sound*, 10(01) p.75–81 (2005). 105

[Jensen, 11] Jensen, K. On the Use of Memory Models in Audio Features. *In Proceedings of FSRM/CMMR 2011* (2011). 106

[Jensen et Kühl, 10] Jensen, K. et Kühl, O. Towards a model of musical chunks. *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music*, p.81–92 (2010). 108

[Johnson, 99] Johnson, C. G. Exploring the sound-space of synthesis algorithms using interactive genetic algorithms. *In* Patrizio, A., Wiggins, G. et Pain, H., éditeurs : *Proceedings of the AISB'99 Symposium on Musical Creativity*, p.20–27, Brighton. Society for the Study of Artificial Intelligence and Simulation of Behaviour (1999). 31

[Johnson et Gounaropoulos, 06] Johnson, C. G. et Gounaropoulos, A. Timbre interfaces using adjectives and adverbs. *In NIME*, p.101–102 (2006). 31

[Johnson et Larson, 03] Johnson, M. L. et Larson, S. 'something in the way she moves'–metaphors of musical motion. *Metaphor and Symbol*, 18(2) p.63 – 84 (2003). 17

[Jot et Chaigne, 91] Jot, J. M. et Chaigne, A. Digital delay networks for designing artificial reverberators. *In Proceedings of the 90th Convention of the Audio Engineering Society* (1991). 91

[Kaczmarek, 05] Kaczmarek, T. Auditory perception of sound source velocity. *The Journal of the Acoustical Society of America*, 117(5) p.3149–3156 (2005). 23

[Karklin et Lewicki, 05] Karklin, Y. et Lewicki, M. S. A hierarchical bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Computation*, 17(2) p.397–423 (2005). 106

[Kawai *et al.*, 04] Kawai, K., Kojima, T., K.Hirate et Yasuoka, M. Personal evaluation structure of environmental sounds: experiments of subjective evaluation using subjects' own terms. *Journal of sound and vibration*, 277 p.523–533 (2004). 41

[Kim *et al.*, 05] Kim, H.-G., Moreau, N. et Sikora, T. *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. John Wiley & Sons (2005). 22

[Kimura *et al.*, 05] Kimura, T., Kakehi, K., Takeda, K. et Itakura, F. Spatial coding based on the extraction of moving sound sources in wavefield synthesis. *In ICASSP* (2005). 22

[Kitagawa et Ichihara, 02] Kitagawa, N. et Ichihara, S. Hearing visual motion in depth. *Nature*, 416(6877) p.172–174 (2002). 61

[Krishna et Sreenivas, 04] Krishna, A. et Sreenivas, T. Music instrument recognition: from isolated notes to solo phrases. *In Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 4, p.iv–265. IEEE (2004). 21

[Kronland-Martinet et Voinier, 08] Kronland-Martinet, R. et Voinier, T. Real-time perceptual simulation of moving sources: Application to the leslie cabinet and 3d sound immersion. *EURASIP Journal on Audio, Speech, and Music Processing*, 2008 (2008). 22

[Kronman et Sundberg, 87] Kronman, U. et Sundberg, J. Is the musical ritard an allusion to physical motion. *Action and perception in rhythm and music*, 55 p.57–68 (1987). 17

[Lackner et Shenker, 85] Lackner, J. et Shenker, B. Proprioceptive influences on auditory and visual spatial localization. *Journal of Neuroscience*, 5(3) p.579 (1985). 25

[Laurent *et al.*, 98] Laurent, H., Doncarli, C. et CNRS, N. Stationarity index for abrupt changes detection in thetime-frequency plane. *IEEE Signal processing letters*, 5(2) p.43–45 (1998). 22

[Lavandier et Defreville, 06] Lavandier, C. et Defreville, B. The contribution of sound source characteristics in the assessment of urban soundscapes. *Acta Acustica united with Acustica*, 92 p.912–921(10) (2006). 38

[Lemaitre *et al.*, 10] Lemaitre, G., Houix, O., Misdariis, N. et Susini, P. Listener expertise and sound identification influence the categorization of environmental sounds. *Journal of Experimental Psychology: Applied*, 16(1) p.16 – 32 (2010). 105

[Leman et Cirotteau, 05] Leman, M. et Cirotteau, D., éditeurs. *Sound to Sense, Sense to Sound: A State-of-the-Art*. Version 0.10. Information Society Technologie, Disponible librement sur le site http://www.s2s2.org/docman/ (2005). 8

[Lesbros, 96] Lesbros, V. From images to sounds, a dual representation. *Computer Music Journal*, p.59–69 (1996). 59

[Lesbros, 99] Lesbros, V. Phonograms, elastic couplings, and trajectories. *Computer Music Journal*, 23(2) p.70–79 (1999). 59

[Li *et al.*, 91] Li, X., Logan, R. et Pastore, R. Perception of acoustic source characteristics: Walking sounds. *The Journal of the Acoustical Society of America*, 90 p.30–36 (1991). 24

[Liberman et Mattingly, 85] Liberman, A. M. et Mattingly, I. G. The motor theory of speech perception revised. *Cognition*, 21(1) p.1 – 36 (1985). 7

[Lighthill, 52] Lighthill, M. J. On Sound Generated Aerodynamically. I. General Theory. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 211(1107) p.564–587 (1952). 21

[Lighthill, 54] Lighthill, M. J. On Sound Generated Aerodynamically. II. Turbulence as a Source of Sound. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 222(1148) p.1–32 (1954). 21

[Lindblom, 87] Lindblom, B. Absolute constancy and adaptive variability: Two themes in the quest for phonetic invariance. *In Proceedings of the XIth International Congress of Phonetic Sciences*, volume 3, p.1–18 (1987). 7

[Lindblom et Sundberg, 71] Lindblom, B. E. F. et Sundberg, J. E. F. Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, 50(4B) p.1166–1179 (1971). 20

[Lutfi, 07] Lutfi, R. Human sound source identification. *Auditory Perception of Sound Sources*, p.13–42 (2007). 6

[Lutfi et Wang, 99] Lutfi, R. A. et Wang, W. Correlational analysis of acoustic cues for the discriminationn of auditory motion. *Journal of the acoustical society of america*, 106(2) p.919–928 (1999). 23, 49

[MacDougall et Moore, 05] MacDougall, H. G. et Moore, S. T. Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *J Appl Physiol*, 99(3) p.1164–1173 (2005). 17

[Malloch *et al.*, 08] Malloch, J., Sinclair, S. et Wanderley, M. A network-based framework for collaborative development and performance of digital musical instruments. *In* Kronland-Martinet, R., Ystad, S. et Jensen, K., éditeurs : *Computer Music Modeling and Retrieval. Sense of Sounds*, volume 4969 de *Lecture Notes in Computer Science*, p.401–425. Springer Berlin / Heidelberg (2008). 30

[Mandelis, 03] Mandelis, J. Genophone: Evolving sounds and integral performance parameter mappings. *In Proceedings of the 2003 international conference on Applications of evolutionary computing*, p.535–546. Springer-Verlag (2003). 31

[Manzolli *et al.*, 01] Manzolli, J., Maia, Jr., A., Fornari, J. et Damiani, F. The evolutionary sound synthesis method. *In Proceedings of the ninth ACM international conference on Multimedia*, MULTIMEDIA '01, p.585–587, New York, NY, USA. ACM (2001). 31

[Martin et Flandrin, 85] Martin, W. et Flandrin, P. Detection of changes of signal structure by using the Wigner-Ville spectrum. *Signal processing*, 8(2) p.215–233 (1985). 22

[Mathews, 63] Mathews, M. . V. The digital computer as a musical instrument. *Science*, 142(3592) (1963). 5

[Mauguière et Fischer, 07] Mauguière, F. et Fischer, C. Potentiels évoqués en neurologie: réponses normales. *EMC Neurologie, Elsevier Masson SAS*, 17 (2007). 24

[McAdams, 93] McAdams, S. Recognition of sound sources and events. *In* McAdams, S. et Bigand, E., éditeurs : *Thinking in Sound: The cognitive psychology of human audition*, p.146–198. Oxford University Press (1993). 6

[McAdams *et al.*, 95] McAdams, S., S.Winsberg, S.Donnadieu, Soete, G. D. et J.Krimphoff. Perceptual scaling of synthetized musical timbres:common dimensions, specificities and latent subject classes. *Psychological Reshearch*, 58 p.177–192 (1995). 21, 32, 38

[McBeath et Neuhoff, 02] McBeath, M. et Neuhoff, J. The Doppler effect is not what you think it is: Dramatic pitch change due to dynamic intensity change. *Psychonomic bulletin & review*, 9(2) p.306 (2002). 23

[McDermott *et al.*, 05] McDermott, J., Griffith, N. J. L. et O'Neill, M. Toward user-directed evolution of sound synthesis parameters. *Applications on Evolutionary Computing*, p.517–526 (2005). 31

[McDermott *et al.*, 07] McDermott, J., Griffith, N. J. L. et O'Neill, M. Evolutionary guis for sound synthesis. *Applications of Evolutinary Computing*, p.547–556 (2007). 31

[McDermott *et al.*, 08] McDermott, J., Griffith, N. J. L. et O'Neill, M. *The Art of Artificial Evolution*, chapitre Evolutionary Computation Applied to Sound Synthesis, p.81–101. Natural Computing Series. Springer Berlin / Heidelberg (2008). 31

[Merer *et al.*, 10] Merer, A., Aramaki, M., Kronland-Martinet, R. et Ystad, S. On the potentiality of abstract sounds in perception research. *In Proceedings of CMMR 2010, Málaga, Spain*, p.207–219 (2010). 102

[Merer *et al.*, 11] Merer, A., Aramaki, M., Ystad, S. et Kronland-Martinet, R. Perceptual characterization of motions evoked by sound for synthesis control. *Submitted for publication in IEEE transaction on audio speech and language processing* (2011). 58, 86

[Merer *et al.*, 07] Merer, A., Aramaki, M., Ystad, S., Kronland-Martinet, R., Besson, M. et Velay, J. Percepual categorization of moving sounds for synthesis applications. *In Proceedings the International Computer Music Conference*, p.69–72 (2007).

[Merer *et al.*, 08] Merer, A., Ystad, S., Kronland-Martinet, R. et Aramaki, M. Semiotics of sounds evoking motions: Categorization and acoustic features. *In* Kronland-Martinet, R., Ystad, S. et Jensen, K., éditeurs : *CMMR 2007. Sense of Sounds*, p.139–158. Springer, LNCS (2008). 37

[Meyer-Eppler, 58] Meyer-Eppler, W. Statistic and psychologic problems of sound. *Die Reihe*, 1 p.55–61 (1958). 34

[Micoulaud-Franchi *et al.*, 11a] Micoulaud-Franchi, J.-A., Aramaki, M., Merer, A., Cermolacce, M., Ystad, S., Kronland-Martinet, R. et Vion-Dury, J. Categorization and timbre perception of environmental sounds in schizophrenia. *Psychiatry Research*, In Press, Corrected Proof p.– (2011). 103

[Micoulaud-Franchi *et al.*, 11b] Micoulaud-Franchi, J. A., Aramaki, M., Merer, A., Cermolacce, M., Ystad, S., Kronland-Martinet, R. et Vion-Dury, J. Toward an exploration of "feeling of strangeness" in schizophrenia: Perspectives from "acousmatic" and "everyday" listening. *Submitted for publication in Journal of Abnormal Psychology* (2011). 103

[Minard *et al.*, 10] Minard, A., Misdariis, N., Houix, O. et Susini, P. Catégorisation de sons environnementaux sur la base de profils morphologiques. *In* Société Française d'Acoustique SFA, éditeur : *10ème Congrès Français d'Acoustique 10ème Congrès Français d'Acoustique*, p.–, Lyon France (2010). 60

[Miner *et al.*, 02] Miner, N. E., Goldsmith, T. E. et Caudell, T. P. Perceptual validation experiments for evaluating the quality of wavelet-synthesized sounds. *Presence: Teleoperators and Virtual Environments*, 11(5) p.508–524 (2002). 107

[Miranda *et al.*, 00] Miranda, E. R., Correa, J. et Wrightt, J. Categorising complex dynamic sounds. *Organised Sound*, 5(02) p.95–102 (2000). 41

[Moore, 83] Moore, F. A general model for spatial processing of sounds. *Computer Music Journal*, 7(3) p.6–15 (1983). 21

[Moore, 89] Moore, F. Spatialization of sounds over loudspeakers. *In* Mathews, M. et Pierce, J., éditeurs : *Current directions in computer music research*, p.103. MIT Press (1989). 22

[Mulder *et al.*, 97] Mulder, A., Fels, S. et Mase, K. Mapping virtual object manipulation to sound variation. *IPSJ Sig Notes*, 97(122) p.63–68 (1997). 30

[Müller, 07] Müller, M. *Information retrieval for music and motion*. Springer-Verlag New York Inc (2007). 19

[Neelon et Jenison, 03] Neelon, M. F. et Jenison, R. L. The effect of trajectory on the auditory motion aftereffect. *Hearing Research*, 180(1-2) p.57 – 66 (2003). 25

[Neff, 85] Neff, D. L. Stimulus parameters governing confusion effects in forward masking. *The Journal of the Acoustical Society of America*, 78(6) p.1966–1976 (1985). 67

[Neuhoff, 98] Neuhoff, J. Perceptual bias for rising tones. *Nature*, 395(6698) p.123–123 (1998). 25, 68

[Neuhoff, 01] Neuhoff, J. An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, 13(2) p.87–110 (2001). 25

[Neuhoff *et al.*, 09] Neuhoff, J., Planisek, R. et Seifritz, E. Adaptive sex differences in auditory motion perception: looming sounds are special. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1) p.10 (2009). 26

[Nusbaum *et al.*, 97] Nusbaum, H., Francis, A. et Henly, A. Measuring the naturalness of synthetic speech. *International Journal of Speech Technology*, 2 p.7–19 (1997). 104

[O'Brien *et al.*, 01] O'Brien, J., Cook, P. et Essl, G. Synthesizing sounds from physically based motion. *In Proceedings of the 28th annual conference on Computer graphics and interactive techniques SIGGRAPH*, p.529–536 (2001). 23

[Olivero *et al.*, 09] Olivero, A., Daudet, L., Kronland Martinet, R. et Torrésani, B. Analyse et catégorisation de sons par multiplicateurs temps-fréquence. *In XXIIe colloque GRETSI XXIIe colloque GRETSI (Dijon)*, Dijon France (2009). 106

[Olivero *et al.*, 10] Olivero, A., Torrésani, B. et Kronland Martinet, R. A new method for Gabor multipliers estimation : application to sound morphing. *In Proceedings EUSIPCO European Signal Processing Conference (EUSIPCO 2010)*, p.507–511, Aalborg Danemark (2010). 106

[Paavilainen *et al.*, 99] Paavilainen, P., Jaramillo, M., Näätänen, R. et Winkler, I. Neuronal populations in the human brain extracting invariant relationships from acoustic variance. *Neuroscience Letters*, 265(3) p.179 – 182 (1999). 6

[Pachet et Roy, 07] Pachet, F. et Roy, P. Exploring billions of audio features. *In* Eurasip, éditeur : *Proceedings of CBMI 07* (2007). 22

[Paine, 04] Paine, D. G. Gesture and musical interaction: interactive engagement through dynamic morphology. *In NIME '04: Proceedings of the 2004 conference on New interfaces for musical expression*, p.80–86, Singapore. National University of Singapore (2004). 30

[Palmer *et al.*, 07] Palmer, C., Carter, C., Koopmans, E. et Loehr, J. Movement, planning, and music: Motion coordinates of skilled performance. *Proceedings of ICoMCS December*, p.119 (2007). 17

[Parizet *et al.*, 05] Parizet, E., Hamzaoui, N. et Sabatie, G. Comparison of some listening test methods: A case study. *Acta Acustica united with Acustica*, 91 p.356–364(9) (March/April 2005). 40

[Pavani *et al.*, 02] Pavani, F., Macaluso, E., Warren, J. D., Driver, J. et Griffiths, T. D. A common cortical substrate activated by horizontal and vertical sound movement in the human brain. *Current Biology*, 12(18) p.1584–1590 (2002). 24

[Peeters et Deruty, 08] Peeters, G. et Deruty, E. Automatic morphological description of sounds. *In Acoustics 08*, Paris, France (2008). 108

[Pellegrino *et al.*, 92] Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. et Rizzolatti, G. Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91 p.176–180 (1992). 24, 25

[Petitmengin *et al.*, 09] Petitmengin, C., Bitbol, M., Nissou, J., Pachoud, B., Curallucci, H., Cermolacce, M. et Vion-Dury, J. Listening from within. *Journal of Consciousness Studies*, 16 p.363–404(42) (2009). 105

[Ramstein, 91] Ramstein, C. *Analyse, représentation et traitement du geste instrumental: application aux instruments à clavier*. Thèse de doctorat, Institut national polytechnique de Grenoble (1991). 19

[Rasamimanana, 08] Rasamimanana, N. *Geste instrumental du violoniste en situation de jeu : analyse et modélisation*. Thèse de doctorat, Université Pierre et Marie Curie (Paris 6), Paris (2008). 18

[Rath *et al.*, 03] Rath, M., Avanzini, F., Bernardini, N., Borin, G., Fontana, F., Ottaviani, L. et Rocchesso, D. An introductory catalog of computer-synthesized contact sounds, in real-time. *In Proc. Colloquium of Musical Informatics*, p.103–108 (2003). 23

[Rath et Fontana, 03] Rath, M. et Fontana, F. *High-level models: bouncing, breaking, rolling, crumpling, pouring*, chapitre 9. D. Rocchesso and F. Fontana (2003). 21

[Repp, 93] Repp, B. H. Musical motion: Some historical and contemporary perspectives. Status report: Sr-114, Haskins Laboratories (1993). 59, 60

[Ricard et Herrera, 04] Ricard, J. et Herrera, P. Morphological sound description computational model and usability evaluation. *In AES 116th Convention* (2004). 108

[Risset, 10] Risset, J.-C. A propos d'interdisciplinarité - informatique musicale : synthèse, traitement, perception ; musicologie et S.T.I.C. ; oeuvre musicale et mixité. *In Journées d'Informatique Musicale*, Rennes France (2010). Invited keynote speech OR 19. 111

[Risset et Wessel, 99] Risset, J. C. et Wessel, D. L. Exploration of timbre by analysis and synthesis. *In* Deutsch, D., éditeur : *The psychology of music*, Series in Cognition and Perception, p.113–169. Academic Press (1999). 6, 31, 86

[Rogosky et Golstone, 05] Rogosky, B. et Golstone, R. Functional features in language and space: Insights from perception, categorization and development. *In* Carlson, L. A. et van der Zee, E., éditeurs : *Adaptation of Perceptual and Semantic Features*, p.257–273. Oxford University Press, Oxford, England (2005). 41

[Rosch, 78] Rosch, E. Principles of categorization. *In* Rosch, E. et Lloyd, B. B., éditeurs : *Cognition and Categorization*, p.27–48. John Wiley & Sons Inc (1978). 41

[Rosenblum *et al.*, 87] Rosenblum, L. D., Carello, C. et Pastore, R. E. Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception*, 16(2) p.175–186 (1987). 23

[Rousset, 08] Rousset, J. Proposition graphique pour la notation des unités sémiotiques temporelles. *In* Rix, E. et Formosa, M., éditeurs : *Vers une sémiotique générale du temps dans les arts, Actes du colloque "Les unités sémiotiques temporelles (UST), nouvel outil d'analyse musicale: Théories et applications Marseille décembre 2005*. Editions Delatour / Ircam (2008). 59, 61

[Saint-Arnaud et Popat, 98] Saint-Arnaud, N. et Popat, K. Analysis and synthesis of sound textures. *In* Rosenthal, D. F. et Okuno, H. G., éditeurs : *Computational auditory scene analysis, Proceedings of the Ijcai-95 Workshop*, chapitre 19, p.308. L. Erlbaum Associates Inc. (1998). 90

[Sakamoto *et al.*, 97] Sakamoto, H., Hayashi, F. et S.Sugiura. Psycho-physiological responses by listening to some sounds from our dayly life. *Journal of sound and vibration*, 205(4) p.499–503 (1997). 25

[Salmelin *et al.*, 95] Salmelin, R., Hamaalainen, M., Kajola, M. et Hari, R. Functional segregation of movement-related rhythmic activity in the human brain. *Neuroimage*, 2(4) p.237–243 (1995). 24

[Schaeffer, 66] Schaeffer, P. *Traité des objets musicaux*. Editions du seuil (1966). 35, 41

[Schindler, 84] Schindler, K. Dynamic timbre control for real-time digital synthesis. *Computer Music Journal*, 8(1) p.28–42 (1984). 32

[Schön *et al.*, 10] Schön, D., Ystad, S., Kronland-Martinet, R. et Besson, M. The evocative power of sounds: Conceptual priming between words and nonverbal sounds. *Journal of Cognitive Neuroscience*, 22(5) p.1026–1035 (2010). 13, 104

[Schubert et Wolfe, 06] Schubert, E. et Wolfe, J. Does timbral brightness scale with frequency and spectral centroid ? *Acta acustica united with acustica*, 92(5) p.820–825 (2006). 51

[Sciabica *et al.*, 10] Sciabica, J.-F., Bezat, M.-C., Roussarie, V., Kronland-Martinet, R. et Ystad, S. Towards timbre modeling of sounds inside accelerating cars. *In* Ystad, S., Aramaki, M., Kronland-Martinet, R. et Jensen, K., éditeurs : *Auditory Display*, volume 5954 de *Lecture Notes in Computer Science*, p.377–391. Springer Berlin / Heidelberg (2010). 107

[Sekuler *et al.*, 97] Sekuler, R., Sekuler, A. B. et Lau, R. Sound alters visual motion perception. *Nature*, 385(6614) p.308–308 (1997). 61

[Serra, 07] Serra, X. State of the art and future directions in musical sound synthesis. *In Proceedings of the IEEE 9th Workshop on Multimedia Signal Processing*, p.9–12 (2007). 5

[Shams *et al.*, 02] Shams, L., Kamitani, Y. et Shimojo, S. Visual illusion induced by sound. *Cognitive Brain Research*, 14(1) p.147–152 (2002). 61

[Shove et Repp, 95] Shove, P. et Repp, B. Musical motion and performance: Theoretical and empirical perspectives. *In* Rink, J., éditeur : *The practice of performance*, chapitre 3, p.55–83. Cambridge University Press (1995). 16, 17

[Smalley, 97] Smalley, D. Spectromorphology: explaining sound-shapes. *Organised Sound*, 2(02) p.107–126 (1997). 19

[Solomon, 57] Solomon, L. Semantic approach to the perception of complex sounds. *Journal of the acoustical society of america*, 30 p.421–425 (1957). 103

[Steiner, 06] Steiner, H.-C. Towards a catalog and software library of mapping methods. *In NIME '06: Proceedings of the 2006 conference on New interfaces for musical expression*, p.106–109, Paris, France. IRCAM — Centre Pompidou (2006). 30

[Stevens, 00] Stevens, K. *Acoustic Phonetics*. The MIT Press (2000). 7

[Stoelinga *et al.*, 03] Stoelinga, C., Hermes, D., Hirschberg, A. et Houtsma, A. Temporal aspects of rolling sounds: A smooth ball approaching the edge of a plate. *Acta Acustica united with Acustica*, 89 p.809–817 (September/October 2003). 21

[Stroppa, 84] Stroppa, M. The analysis of electronic music. *Contemporary Music Review*, 1(1) p.175–180 (1984). 36

[Tajadura-Jimenez *et al.*, 10] Tajadura-Jimenez, A., Väljamäe, A., Asutay, E. et Västfjäll, D. Embodied auditory perception: The emotional impact of approaching and receding sound sources. *Emotion*, 10(2) p.216 – 229 (2010). 26

[Takagi, 01] Takagi, H. Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE*, 89(9) p.1275–1296 (2001). 30

[Takala *et al.*, 93] Takala, T., Hahn, J., Gritz, L., Geigel, J. et Lee, J. Using physically-based models and genetic algorithms for functional composition of sound signals, synchronized to animated motion. *In Proceedings of the International Computer Music Conference*, p.180–180 (1993). 31

[Tardieu *et al.*, 08] Tardieu, J., Susini, P., Poisson, F., Lazareff, P. et McAdams, S. Perceptual study of soundscapes in train stations. *Applied Acoustics*, 69(12) p.1224–1239 (2008). 38

[Thaut et Abiru, 10] Thaut, M. et Abiru, M. Rhythmic Auditory Stimulation in Rehabilitation of Movement Disorders: A Review of Current Research. *Music Perception*, 27(4) p.263–269 (2010). 18

[Todd *et al.*, 02] Todd, N., Lee, C. et O'Boyle, D. A sensorimotor theory of temporal tracking and beat induction. *Psychological Research*, 66 p.26–39 (2002). 17

[Todd, 92] Todd, N. P. M. The dynamics of dynamics: A model of musical expression. *The Journal of the Acoustical Society of America*, 91(6) p.3540–3550 (1992). 17

[Todd, 93] Todd, N. P. M. Multi-scale analysis of expressive signals: Recovery of structure and motion. *In Proceedings of the Stockholm Music Acoustics Conference SMAC93*, p.146–149 (1993). 16

[Todd, 95] Todd, N. P. M. The kinematics of musical expression. *The Journal of the Acoustical Society of America*, 97(3) p.1940–1949 (1995). 16, 17

[Traube et Depalle, 04] Traube, C. et Depalle, P. Phonetic gestures underlying guitar timbre description. *In Proceedings of 8th International Conference on Music Perception and Cognition*, p.658–661 (2004). 20

[Tsakostas, 07] Tsakostas, Christos ; Floros, A. Real-time spatial representation of moving sound sources. *In Audio Engineering Society Convention 123* (2007). 22

[Valle et Lombardo, 03] Valle, A. et Lombardo, V. A two-level method to control granular synthesis. *In XIV CIM-Proceedings of the XIV Colloquium on Musical Informatics 2003*, p.136–140. Citeseer (2003). 30

[van den Doel *et al.*, 01] van den Doel, K., Kry, P. G. et Pai, D. K. Foley automatic: Physically-based sound effects for interactive simulation and animation. *In Computer Graphics (ACM SIGGRAPH 01 Conference Proceedings*, p.537–544 (2001). 23

[Van Nort, 09] Van Nort, D. Instrumental listening: sonic gesture as design principle. *Organised Sound*, 14(02) p.177–187 (2009). 20

[Van Nort et Wanderley, 06] Van Nort, D. et Wanderley, M. The LoM Mapping Toolbox for Max/MSP/Jitter. *In Proc. of the 2006 International Computer Music Conference (ICMC)*, p.397–400 (2006). 30

[Verron *et al.*, 10] Verron, C., Aramaki, M., Kronland-Martinet, R. et Pallone, G. A 3-D immersive synthesizer for environmental sounds. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(6) p.1550–1561 (2010). 33

[Vertegaal et Bonis, 94] Vertegaal, R. et Bonis, E. Isee: An intuitive sound editing environment. *Computer Music Journal*, 18(2) p.21–29 (1994). 29

[Vines *et al.*, 04] Vines, B. W., Wanderley, M. M., Krumhansl, C. L., Nuzzo, R. L. et Levitin, D. J. Performance gestures of musicians: What structural and emotional information do they convey ? *In Gesture-Based Communication in Human-Computer Interaction*, volume 2915 de *Lecture Notes in Computer Science*, p.468–478. Springer Berlin / Heidelberg (2004). 20

[Viviani et Cenzato, 85] Viviani, P. et Cenzato, M. Segmentation and coupling in complex movements. *Journal of Experimental Psychology: Human Perception and Performance*, 11(6) p.828 – 845 (1985). 18

[Viviani et Stucchi, 92] Viviani, P. et Stucchi, N. Biological movements look uniform: Evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3) p.603 – 623 (1992). 18

[Vogel, 99] Vogel, C. *Etude sémiotique et acoustique de l'identification des signaux sonores d'avertissement en contexte urbain*. Thèse de doctorat, Université Paris 6 (1999). 6

[Wallach, 40] Wallach, H. The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4) p.339 – 368 (1940). 25

[Wanderley, 99] Wanderley, M. Non-obvious Performer Gestures in Instrumental Music. *In Gesture-based communication in human computer interaction: International Gesture Workshop, GW'99, Gif-sur-Yvette, France*, p.37–48. Springer Verlag (1999). 19

[Wanderley, 02] Wanderley, M. (guest editor) Special issue on mapping in computer music. *Organised Sound*, 7(02) (2002). 30

[Wanderley et Battier, 00] Wanderley, M. et Battier, M., éditeurs. *Trends in gestural control of music*. IRCAM — Centre Pompidou (2000). 30

[Wanderley et Depalle, 04] Wanderley, M. et Depalle, P. Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4) p.632–644 (2004). 30

[Warren *et al.*, 02] Warren, J. D., Zielinski, B. A., Green, G., Rauschecker, J., et T.D.Griffiths. Perception of sound-source motion by the human brain. *Neuron*, 34 p.139–148 (2002). 24

[Wessel et Wright, 02] Wessel, D. et Wright, M. Problems and prospects for intimate musical control of computers. *Computer Music Journal*, 26(3) p.11–22 (2002). 29

[Wessel, 79] Wessel, D. L. Timbre space as a musical control structure. *Computer Music Journal*, 3(2) p.45–52 (1979). 32

[Wexler et Klam, 01] Wexler, M. et Klam, F. Movement prediction and movement production. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1) p.48–58 (2001). 18

[Wightman et Kistler, 99] Wightman, F. L. et Kistler, D. J. Resolution of front–back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America*, 105(5) p.2841–2853 (1999). 25

[Winkler, 95] Winkler, T. Making motion musical: Gesture mapping strategies for interactive computer music. *In Proceedings of the 1995 international Computer music conference*, p.261–264 (1995). 30

[Wu et Li, 08] Wu, X. et Li, Z. A study of image-based music composition. *In 2008 IEEE International Conference on Multimedia and Expo*, p.1345–1348 (2008). 59

[Xiao *et al.*, 07a] Xiao, J., Borgnat, P. et Flandrin, P. Sur un test temps-fréquence de stationnarité. *In Colloque GRETSI-2007 21e Colloque sur le Traitement du Signal et des Images. GRETSI-2007*, p.957–960, Troyes France. GRETSI & UTT, GRETSI (2007). 22, 89

[Xiao *et al.*, 07b] Xiao, J., Borgnat, P. et Flandrin, P. Testing Stationarity with Time-Frequency Surrogates. *In EUSIPCO-2007 15th European Signal Processing Conference EUSIPCO-2007*, Poznan Pologne. EURASIP, EURASIP (2007). 22, 89

[Xiao et Flandrin, 07] Xiao, J. et Flandrin, P. Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement. *Signal Processing, IEEE Transactions on*, 55(6) p.2851–2860 (2007). 89

[Yadegari *et al.*, 02] Yadegari, S., Moore, F., Castle, H., Burr, A. et Apel, T. Real-time implementation of a general model for spatial processing of sounds. *In Proceedings of the 2002 International Computer Music Conference*, p.244–247 (2002). 22

[Ystad *et al.*, 05] Ystad, S., Kronland-Martinet, R., Schön, D. et Besson, M. Vers une approche acoustique et cognitive de la sémiotique des objets sonores. *In Unités sémiotiques temporelles: théorie et applications*. Cité de la musique Marseille (2005). 8

[Zattra, 05] Zattra, L. Analysis and analyses of electroacoustic music. *In Sound and Music Computing (SMC05), Salerno, Italy* (2005). 36

[Zwaan *et al.*, 04] Zwaan, R. A., Madden, C. J., Yaxley, R. H. et Aveyard, M. E. Moving words: dynamic representations in language comprehension. *Cognitive Science*, 28(4) p.611 – 619 (2004). 61

[Zwicker et Fastl, 90] Zwicker, E. et Fastl, H. *Psychoacoustics, facts and models*. Springer Verlag (1990). 51, 67

Total: 263 références

Nota: Dans la version pdf de ce document, les titres qui apparaissent en bleu sont des liens hypertexte qui permettent d'accéder aux versions en ligne des documents cités.

Titre Caractérisation acoustique et perceptive du mouvement évoqué par les sons pour le contrôle de la synthèse

Résumé Un des grands enjeux actuels dans le domaine de la synthèse sonore consiste à proposer des stratégies de contrôle intuitif accessible à un grand nombre d'utilisateurs. Parmi les attributs fondamentaux du son, ceux qui sont liés au mouvement restent encore largement inexplorés et, de fait, difficilement contrôlables. Cette thèse s'intéresse à la perception des caractéristiques des sons monophoniques qui varient au cours du temps et qui sont à l'origine de la notion perceptive de "mouvement évoqué", ainsi qu'à la mise au point d'une stratégie de contrôle intuitif. Pour cela, nous avons mis au point une série d'expériences basées sur des protocoles de catégorisation et d'évaluation sonore par le biais de dessins via une interface graphique paramétrée. Afin de favoriser l'écoute sur les attributs intrinsèques du son, nous avons privilégié l'utilisation de sons dits "abstraits" pour lesquels la source physique n'est pas facilement identifiable. Les résultats de ces expériences nous ont permis de définir un ensemble de dimensions perceptives importantes pour la caractérisation du mouvement évoqué, telles que la forme de la trajectoire, l'orientation ou bien la taille. Ces dimensions ont été par la suite validées grâce à un protocole expérimental exploitant la synthèse sonore. Le travail réalisé ouvre de nombreuses perspectives, notamment dans les domaines de la réalité virtuelle, du "design sonore" et de la musique. La méthodologie proposée vise à être suffisamment générale pour être transposable à l'étude d'autres attributs perceptifs liés au ressenti (émotions, qualité . . .).

Mots-cles mouvement, contrôle de synthétiseur, sons abstraits, perception, analyse-synthèse

---

Title Perceptual and acoustic characterization of movement evoked by sound for synthesis control

Abstract An important issue within sound synthesis, is to propose intuitive control strategies that are accessible to most potential users. Among fundamental sound attributes, the ones linked to motion are almost unexplored and thus cannot be easily controlled. This Ph.D. concerns the perception of time varying characteristics of monophonic sounds, which leads to the notion of motion evoked by sounds and the conception of intuitive control strategies for such motions. To address those questions, we developed several experiments based on categorization tasks and evaluations through drawings using a parameterized graphical interface. In order to help listeners focus on intrinsic sound attributes, we mainly used so-called "abstract" sounds for which the physical sound source is not easily recognizable. The results of those experiments enabled us to identify attributes that characterize evoked motion from a perceptual point of view, such as shape, direction and size of the trajectory. Those dimensions were further validated in an experimental protocol using sound synthesis. This work opens many perspectives, in particular within virtual reality, sound design and musical domains. The proposed methodology aimed at being sufficiently generic to be transposed to other perceptual attributes such as emotions and quality.

Keywords moving sounds, synthesizer control, abstract sounds, perception, analysis-synthesis

---

Laboratoire CNRS - Laboratoire de Mécanique et d'Acoustique (UPR7051)
31 chemin Joseph-Aiguier 13402 Marseille cedex 20