



HAL
open science

CARACTÉRISATION DE LA QUALITÉ SONORE DE L'ENVIRONNEMENT URBAIN : UNE APPROCHE PHYSIQUE ET PERCEPTIVE BASÉE SUR L'IDENTIFICATION DES SOURCES SONORES

Boris Defreville

► **To cite this version:**

Boris Defreville. CARACTÉRISATION DE LA QUALITÉ SONORE DE L'ENVIRONNEMENT URBAIN : UNE APPROCHE PHYSIQUE ET PERCEPTIVE BASÉE SUR L'IDENTIFICATION DES SOURCES SONORES. Acoustique [physics.class-ph]. Université de Cergy Pontoise, 2005. Français. NNT: . tel-00011453

HAL Id: tel-00011453

<https://theses.hal.science/tel-00011453>

Submitted on 24 Jan 2006

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.



ECOLE DOCTORALE SCIENCES ET INGENIERIE

De l'université de Cergy-Pontoise

THÈSE

Présentée pour obtenir le grade de docteur de l'université de Cergy-Pontoise

Spécialité : Acoustique

**CARACTÉRISATION DE LA QUALITÉ SONORE
DE L'ENVIRONNEMENT URBAIN :
UNE APPROCHE PHYSIQUE ET PERCEPTIVE BASÉE
SUR L'IDENTIFICATION DES SOURCES SONORES.**

par

Boris DEFREVILLE

Laboratoire Mobilités, Réseaux, Territoires et Environnements

Soutenue le 15 décembre 2005 devant le jury composé de :

M. Michel BÉRENGIER Rapporteur

M. Georges CANÉVET Rapporteur

Mme Catherine LAVANDIER Co-directeur de thèse

M. Philippe GUIGNOUARD Examineur

M. François PACHET Examineur

M. Gérard IGNAZI Directeur de thèse

REMERCIEMENTS

Des remerciements non exhaustifs dans une liste qui n'en est pas moins.

Merci

À Catherine LAVANDIER, Maître de conférences de l'Université de Cergy-Pontoise, qui m'a parfaitement guidé tout au long de ce travail de thèse,

À Philippe GUIGNOUARD, Docteur en acoustique et Directeur du LASA, qui m'a permis de réaliser des travaux de recherches en parallèle à des activités spécifiques d'un bureau d'étude.

À Gérard IGNAZI, Professeur des Universités à l'université de Cergy-Pontoise, pour s'être intéressé à mon travail et avoir accepté d'être mon directeur de thèse.

À Michel BÉRENGIER, Directeur de recherche au LCPC de Nantes et à Georges CANÉVET, Directeur de recherche au LM de Marseille, pour s'être intéressés à mon travail et pour avoir accepté d'en être les rapporteurs.

À François PACHET, Docteur en informatique et ingénieur de recherche à SONY CSL, pour notre collaboration fructueuse et sa participation au jury.

Et bien sûr à toute l'équipe du LASA : Aline LIBERT, Cécile GOUTALIER, Claire CHAUFOR, Elise ARSLANAGIC, Sandrine LIAIGRE, Cyrille AUBERT, David ADAM de VILLIERS, Guillaume DUVAL, Julien RICORDEAU, Kiem-y LAM, Samuel TOCHON-DANGUY et Thomas BOUQUET, sans oublier Arnaud CAN et Christophe ROSIN.

À Jérôme MONCEAUX et Daniel SALAS pour leur génie informatique, à David MAILLARD pour le *mix*, à Anne BENOLIEL pour la charte graphique, à Benoît BARBOT pour nos travaux communs, à Corsin VOGEL pour ses remarques et à l'ensemble des protagonistes des expériences d'évaluations sonores.

TITRE : Caractérisation de la qualité sonore de l'environnement urbain : une approche physique et perceptive basée sur l'identification des sources sonores

RESUME LONG

Le bruit dans nos villes est perçu comme une question de qualité de vie. Il est généralement évalué par la mesure de son niveau sonore. Si cette mesure est représentative d'environnements bruyants caractérisés par des flots continus de véhicules, elle s'avère insuffisante pour caractériser la « couleur » d'une ambiance sonore où différentes sources coexistent. L'amélioration de la qualité sonore de l'environnement urbain passe par une meilleure compréhension de l'évaluation subjective que nous avons des sources de bruit (caractérisation). D'autre part, elle nécessite l'extraction d'un indicateur de désagrément calculable à partir de mesures acoustiques de l'environnement sonore, ceci afin de réaliser des diagnostics de situations actuelles (constatation) ou futures (prédiction).

C'est ainsi que la présente étude se propose d'examiner la qualité d'une ambiance sonore urbaine du point de vue de sa composition, c'est-à-dire des sources sonores occurrentes, et d'en déterminer un modèle prédictif.

S'inspirant de méthodes utilisées par la psychologie expérimentale, la première partie de l'étude s'attache à mieux comprendre comment est évalué un environnement urbain de quartier.

Une première expérience met en évidence, à travers une décomposition structurale d'une ambiance sonore en éléments de bases que sont les sources sonores, le fait que celles-ci ne sont pas toutes perçues de la même manière et qu'en conséquence leurs évaluations métrologiques doivent être adaptées. Par exemple, dans une rue, les sources telles que les voitures sont évaluées sur des notions plutôt spatiales alors que les autres véhicules (de deux ou quatre roues) sont évaluées sur des notions plutôt temporelles.

Une deuxième expérience s'intéresse à la perception de la force sonore et du désagrément sonore. Ce dernier, bien que fortement lié au niveau sonore, dépend également de la nature des sources émergentes.

Une troisième expérience tend à montrer que l'évaluation globale de caractères liés aux sources sur une séquence de plusieurs minutes s'appuie sur les événements marquants de cette séquence. Les évaluations du niveau sonore et du désagrément, quant à eux, sont plutôt proches d'une moyenne.

Enfin, une dernière expérience, menée cette fois in situ, vient corroborer les observations faites lors des trois tests précédents. Ce test sur site montre, de plus, que la force sonore n'est pas bien adaptée pour évaluer l'évolution de la qualité sonore d'ambiance de type parc ou marché. De ce fait, ce sont les sources sonores perçues qui contribuent à l'évaluation de son caractère hédonique.

Ces expériences permettent de proposer un indicateur lié au désagrément sonore pour une rue de quartier ; en fonction du lieu, cet indicateur utilise la force sonore de la séquence, mais il prend toujours en compte les caractéristiques des différentes sources émergentes.

La deuxième partie de l'étude propose deux méthodes pour la mesure automatique de cet indicateur à travers l'identification des sources. La première, simple, mesure la qualité d'ajustement entre les événements sonores et des gabarits spectraux types. La seconde utilise un outil de recherche automatique de traits caractéristiques d'événements sonores par combinaisons d'opérateurs de traitement du signal.

Un mécanisme d'identification des sources a pour vocation, à travers le calcul de l'indicateur, de s'approcher d'une prédiction en temps réel du désagrément sonore. Il représente, in fine, une aide à la gestion d'un environnement sonore de quartier.

TITRE : Caractérisation de la qualité sonore de l'environnement urbain : une approche physique et perceptive basée sur l'identification des sources sonores

RESUME COURT

Le bruit dans nos villes est perçu comme une question de qualité de vie. Il est généralement évalué par la mesure de son niveau sonore. Si cette mesure est représentative d'environnements bruyants caractérisés par des flots continus de véhicules, elle s'avère insuffisante pour caractériser la « couleur » d'une ambiance sonore où différentes sources coexistent. C'est ainsi que la présente étude se propose d'examiner la qualité d'ambiances sonores urbaine du point de vue de leur composition, c'est-à-dire des sources sonores occurrentes., la finalité étant la mise au point d'une méthode d'évaluation automatique d'un environnement sonore urbain.

La première partie met en évidence le fait que celles-ci ne sont pas toutes perçues de la même manière et qu'en conséquence leurs évaluations métrologiques doivent être adaptées. Ce travail permet de proposer un indicateur lié au désagrément sonore ; en fonction du lieu, cet indicateur utilise la force sonore de la séquence, mais il prend toujours en compte les caractéristiques des différentes sources émergentes.

La deuxième partie de l'étude propose deux méthodes pour la mesure automatique de cet indicateur à travers l'identification des sources. Ces outils représentent, in fine, une aide à la gestion d'un environnement sonore de quartier.

MOTS-CLES :

Acoustique, Psychoacoustique, Qualité sonore, Ambiance sonore, Sources sonores, Force sonore, Perception auditive, Représentations mentales, Tests auditifs, Environnement sonore, Environnement urbain, Identification automatique, Taxinomie, Bases de données, Ville, Parc, Rue de quartier, Marché.

**TITRE en anglais : Characterisation of urban sound quality :
a perceptive and an acoustical approach based on source
identification**

ABSTRACT

Noise in cities is perceived as a question of quality of life. It is generally evaluated by the measurement of his sound level. If this measure is representative of noisy environments which are characterised by continuous flow of vehicles, it is insufficient to characterise the “colour” of a urban soundscape where different sources coexist.

The first part reveals that sound sources are not perceived the same manner and their metrological evaluation should be adapted consequently. This present work proposes an indicator linked to the unpleasantness of sound. Depending on the place it describes, this indicator sometimes uses the loudness of the sequence but always takes account characteristics of emergent sound sources.

The second part of the study proposes two methods for the automatic calculation of this indicator thanks to the identification of sound sources. This tool represents, in fine, a help to the management of an urban soundscape.

KEY-WORDS :

Acoutics, Psychoacoustics, Hearing tests, Sound quality, Loudness, Urban soundscape, Sound sources, Sound perception, Mental representations, Sound environment, Urban environment, Automatical identification, Taxinomie, Data bases, City, Parc, Market.

TABLE DES MATIÈRES

REMERCIEMENTS	1
RESUME LONG	5
RESUME COURT	6
ABSTRACT	7
TABLE DES MATIÈRES	9
INTRODUCTION GENERALE	17
I. Problématique	17
II. Organisation du mémoire	21
CHAPITRE 1. DU SON AU PAYSAGE SONORE, DU SIGNAL ACOUSTIQUE A L'IMAGE AUDITIVE	25
I. Introduction à l'état de l'art	25
II. Rôle et théories de la perception auditive	26
2.1. Approche « psychologique »	26
2.2. Approche « évolutionniste ».....	27
2.3. Théorie cognitive	30
2.4. Théorie écologique	33
2.5. L'écoute sonore.....	34
III.Approche psychophysique de l'audition – Théorie et méthodes	36
3.1. La psychoacoustique comme discipline de la psychophysique	36
3.2. Les critères psychoacoustiques comme descripteurs de sensations auditives	37
3.2.1. La sonie	37
3.2.2. Calcul de la sonie	39
3.2.3. L'acuité	40
3.2.4. La force de fluctuation et rugosité	40
3.3. Les critères psychoacoustiques et le bruit	41
3.4. Le timbre et les analyses multidimensionnelles	42
IV. Le paysage sonore	44
4.1. La Qualité Sonore.....	44
4.1.1. Les objets	44
4.1.2. Les bruits.....	46
4.1.3. Les sons courants et sons naturels	46

4.2. Paysage sonore in-situ.....	48
4.3. Les séquences sonores complexes : ambiances sonores urbaines.....	51
V. Gêne ou désagrément sonore ?	56
VI. Conclusions.....	58
6.1. Synthèse.....	58
6.2. Espace descriptif des Objets Sonores.....	60

CHAPITRE 2. CARACTERISATION SUBJECTIVE DE L'ENVIRONNEMENT SONORE URBAIN..... 63

I. Méthode expérimentale.....	64
1.1. Les enregistrements.....	64
1.2. Les stimuli sonores.....	65
1.2.1. Vingt séquences de 15 s.....	65
1.2.2. Séquence de 5 minutes.....	67
1.3. Les sujets des tests.....	68
1.4. Calculs des critères physiques.....	68
II. Caractérisation des séquences de 15 secondes.....	70
2.1. Test n° 1 : Prégnance, présence et proximité des sources sonores.....	70
2.1.1. Questionnaire et procédure.....	70
2.1.2. Normalisation des données, moyennes et corrélations.....	72
a). Normalisation.....	72
b) Moyennes des réponses.....	73
c) Corrélations.....	76
2.1.3. Analyses factorielles.....	78
a) Description des valeurs.....	78
b) Descriptions des analyses.....	78
2.1.3.1. Analyse en Composantes Principales.....	81
b) Procédure.....	81
c) Interprétations.....	82
A- ACP pour les sources prises séparément.....	82
a) ACP sur les voitures.....	83
b) ACP sur les Cyclomoteurs.....	84
c) ACP sur les bus.....	85
d) ACP sur les motos.....	86
e) Conclusions sur les véhicules.....	86
f) ACP sur les oiseaux.....	87
g) ACP sur les voix d'adultes.....	88
h) ACP sur les voix d'enfants.....	88
B- ACP pour les sources groupées.....	89
a) ACP sur les voix.....	89
b) ACP sur les voix et sur les oiseaux.....	90
c) ACP sur tous les véhicules.....	91
d) ACP sur toutes les sources.....	92
2.1.3.2. Analyse factorielle des correspondances.....	93
a) Transformation des données.....	93
b) Individus et variables.....	94
c) Procédure.....	95
d) Interprétations.....	95
A- AFC sur les véhicules.....	95
B- AFC sur les voix et les oiseaux.....	100
C- AFC sur toutes les sources.....	103
2.1.4. Discussion.....	103

2.1.5 Synthèse des expériences sur la prégnance,.....	104
la présence et la proximité des sources sonores	104
2.2. Tests n° 2 et n° 3 : Force et désagrément sonore sur les séquences de 15 s	106
2.2.1. Questionnaire et procédure.....	106
2.2.2. Normalisation des données.....	107
2.2.3. Relations entre force sonore estimée et désagrément sonore.....	108
2.2.4. Relation entre la force sonore et les critères physiques	109
2.2.5. Relation entre le désagrément et la force sonore subjective	111
2.3. Force et désagrément sonore pour le test réalisé in-situ.....	112
2.3.1. Description du test	112
2.3.2. Force et désagrément sonore	115
2.3.2.1. Réponses des sujets.....	115
2.3.2.2. Force sonore estimée et désagrément sonore	117
III. Etablissement d'un modèle de désagrément.....	119
3.1. Introduction.....	119
3.2. Les modèles et leur validité	120
3.2.1. Le modèle de régression linéaire	120
3.2.2. Méthode d'estimation des modèles : la méthode des moindres carrés.....	120
3.2.3. Coefficient de détermination (R^2) et R^2 ajusté (R^2_{aj})	121
3.2.4. Coefficient de corrélation	121
3.2.5. Test de validité des modèles.....	122
3.2.6. Intervalle de confiance sur les paramètres du modèle (tests d'hypothèses)	122
3.2.7. Analyse graphique des résidus.	122
3.2.8. Test de Durbin-Watson.....	123
3.3. Le désagrément sonore	123
3.3.1. Données des modèles	123
3.3.1.1. Variables de départ :.....	123
3.3.1.2. Variables objectives supplémentaires	125
3.3.2. Technique.....	125
3.3.3. Modèles avec la force sonore comme variable.....	126
3.3.3.1. Modèle subjectif	126
3.3.3.2. Modèle objectif.....	128
3.3.3.3. Interprétations	130
3.3.3.4. Critères psychoacoustiques	130
3.3.4. Modèles sans sonie	133
3.3.4.1. Modèle subjectif	135
3.3.4.2. Modèle objectif.....	136
3.4. Domaines de validité des modèles objectifs – intervalles de confiance	138
3.5. Synthèse sur les modèles de régression	140
IV. Caractérisation de la séquence de 5 min	144
4.1. Test n° 4 : Prégnance, présence et proximité pour la séquence globale de 5 min	144
4.1.1. Questionnaire et procédure.....	144
4.1.2. Normalisation des données	145
4.1.3. Présentation des Résultats	146
4.1.4. Analyses statistiques	148
4.1.5. Effet de récence	151
4.2. Tests n° 5 et n° 6 : Sonie et désagrément sonore pour la séquence globale de 5 min.....	154

4.2.1. Questionnaire et procédure.....	154
4.2.2. Analyses statistiques	155
4.2.2.1. Résultats et Moyennes	155
4.2.2.2. Comparaisons avec les jugements des vingt stimuli	157
4.2.3. Effet de récence :	158
4.3. Jugement global d'une séquence réalisée in-situ.....	160
V. Conclusions du chapitre 2.....	162
CHAPITRE 3. ETUDES POUR L'IDENTIFICATION AUTOMATIQUE DES SOURCES SONORES.....	165
<hr/>	
I. Pourquoi l'identification ?.....	165
II. L'identification des sources ou événements sonores	168
2.1. Théorie	168
2.1.1. Les étapes de l'identification	168
2.1.2. Les facteurs cognitifs.....	170
2.1.3. La classification des sons.....	172
2.2. Applications aux sons de l'environnement urbain.....	175
2.2.1. Des méthodes aux applications	175
2.2.2. La classification supervisée :	176
III. Activité des Sources Sonores – A2S.....	177
3.1. Principe	177
3.2. Routines	179
3.2.1. Ressemblance spectrale.....	179
3.2.2. Détection de pics	181
3.2.2.1. Principe.....	181
3.2.2.2. Emergence à deux bandes.....	182
3.2.2.3. Emergence à une bande	183
3.3. Détermination des gabarits spectraux.....	184
3.3.1. Introduction.....	184
a) Démarche empirique.....	184
b) Démarche statistique	185
3.3.2. Démarche empirique.....	185
3.3.2.1. Constitution d'une base de données.....	185
3.3.2.2. Catégories	185
3.3.2.3. Utilisations des sonagrammes.....	186
3.3.2.4. Gabarits	187
3.3.3. Démarche statistique.....	190
3.3.3.1. Ecart type des attributs.....	192
3.3.3.2. Moyennes des attributs.....	196
3.4. Résultats des identifications	197
3.4.1. Oiseaux	197
3.4.2. Voix.....	197
3.4.3. Véhicules	199
3.5. Activité des sources sonores	201
3.6. Conclusion	202
IV. Identification par utilisation d'opérateurs de traitement du signal	203
4.1. Description d'EDS	203

4.2. Mode opératoire.....	204
4.2.1. Constitution d'une base de données.....	204
4.2.2 Recherche des fonctions.....	204
4.2.3. Assemblage des opérateurs	205
4.2.4. Pertinence d'une fonction	207
4.2.5. Création d'un modèle par association de fonctions	207
4.3. Constitution d'une base de données	210
4.3.1. Métrologie et mesure du signal.....	210
4.3.2. Détermination de la durée des séquences.....	210
4.3.3. Normalisation des séquences.....	211
4.4. Classification véhicules/non véhicules.....	211
4.4.1. Pourquoi cette distinction ?.....	211
4.4.2. Méthode	211
4.4.3. Résultats	212
4.4.4. Conséquences pour la suite des recherches.....	216
4.5. Classification cyclomoteurs – autres véhicules.....	217
4.5.1. Base de données cyclomoteurs/véhicules	217
4.5.2. Fonctions	217
4.5.3. Modèles	220
4.5.4. Optimisation d'un modèle.....	221
4.5.5. Vérification du modèle sur des séquences tests.....	221
4.6. Bus	223
4.6.1. Bases de données	223
4.6.2. Fonctions	223
a) Base de départ B1.....	223
b) Base B2.....	224
c) Base B3	225
d) Base B4.....	226
4.6.3. Modèle.....	226
4.6.4. Modèle d'arbre décisionnel logique	228
4.7. Autres sources	229
4.7.1. Automobiles.....	229
4.7.2. Motos.....	230
4.7.3. Oiseaux	231
4.7.4. Voix.....	233
4.8. Ebauche d'un organigramme des détections	234
CONCLUSION GENERALE & PERSPECTIVES	239
<hr/>	
I. Evaluation d'une ambiance urbaine	239
II. Détection automatique des sources	241
BIBLIOGRAPHIE	243
ANNEXES	265
<hr/>	

Le souvenir d'un mot se distingue du souvenir d'un son quelconque, naturel ou musical, en ce qu'au premier correspond toujours un modèle ou un schéma extérieur, fixé soit dans les habitudes phonétiques du groupe (c'est-à-dire sur un support organique), soit sous forme imprimée (c'est-à-dire sur une surface matérielle), alors que la plupart des hommes, lorsqu'ils entendent des sons qui ne sont pas des mots, ne peuvent guère les comparer à des modèles qui seraient purement auditifs, parce que ceux-ci leur manquent.

Certes, lorsque, dans mon cabinet de travail, je lève la tête pour écouter un moment les bruits du dehors et du dedans, je puis dire : ceci, c'est le bruit d'une pelle à charbon dans le corridor, cela, c'est le pas d'un cheval dans la rue, c'est le cri d'un enfant, etc. Mais, on le voit, ce n'est pas autour d'une représentation typique auditive que se groupent d'ordinaire les sons ou les bruits d'une même catégorie : quand je veux reconnaître ces bruits, je songe aux objets ou aux êtres qui, à ma connaissance, en produisent d'analogues, c'est-à-dire que je me reporte à des notions qui ne sont pas essentiellement d'ordre sonore. C'est le son qui fait penser à l'objet, parce qu'on reconnaît l'objet à travers le son ; mais l'objet lui-même (c'est-à-dire le modèle auquel on se reporte) évoquerait rarement tout seul le son. Quand on entend un cliquetis de chaînes, ou bien un bruit de brides froissées, de chevaux au galop, un claquement de fouet, on pense à des prisonniers, à une course de chars. Que ces spectacles nous apparaissent sur l'écran d'un cinématographe, sans qu'un orchestre invisible les accompagne en imitant les sons, nous n'évoquerons pas nous-mêmes les sons, et les figures qui s'agitent dans le silence nous feront beaucoup moins illusion.

Maurice Halbwachs (1939),

« La mémoire collective chez les musiciens ».

Extrait de la *Revue philosophique*, mars-avril 1939, pages 136 à 165.

INTRODUCTION GENERALE

I. Problématique

Le processus d'exode rural commencé il y a maintenant plusieurs décennies, la densification et l'expansion des espaces urbains, l'émergence des sons mécaniques en nombre sans arrêt croissant : voilà autant de processus qui font que les sociétés occidentales deviennent de plus en plus sensibles vis-à-vis de leur environnement et de leur cadre de vie. Quand, en 1867, Adolphe Radeau écrivait que « le son est un télégraphe invisible dont se sert la nature » [RADAU-67], sans doute ne s'imaginait-il pas que les sons « naturels » allaient bientôt, dans notre environnement quotidien, être substitués par les sons mécaniques autrement plus masquants et bruyants. Le très complet rapport de l'AFSSE de mai 2004, intitulé « Impacts sanitaires du bruit », indique que les français attribuent une valeur importante et croissante au droit à la tranquillité sonore dans leur cadre de vie [AFSSE-04]. Sans être associé à un problème de société ou d'environnement, comme peut l'être la qualité de l'air, le bruit est perçu comme une question locale de qualité de vie. De ce fait, le droit à la tranquillité sonore dans le cadre de vie est une préoccupation importante et grandissante dans l'opinion des gens.

Le premier grand pas en termes de réglementation des nuisances sonores a été la loi n° 92-1444 du 31 décembre 1992 relative à la lutte contre le bruit [LOI BRUIT-92]. En général, ses décrets d'applications partent du principe qu'il faut limiter les bruits émis par des sources particulières (voisinage, machine...) par rapport au niveau de bruit résiduel. La mesure est physique : le niveau sonore en décibel A, et le niveau d'émergence autorisé par rapport au niveau résiduel est de cinq décibels A en période diurne et trois décibels A en période nocturne. Cette réglementation sur les émergences est complétée par la fixation de valeurs limites de niveaux de bruit, déjà en vigueur dans de nombreux pays européens (Pays-Bas, Suisse, Allemagne, Italie...). Ces valeurs limites s'appuient sur deux principaux critères liés au bruit que sont la gêne psychologique et la perturbation du sommeil. Cette réglementation vise à protéger les populations riveraines lors de la création nouvelle d'infrastructures de transports, à traiter les bâtiments neufs, à construire aux abords d'infrastructures existantes ou encore à résorber les points noirs existants. La particularité de cette réglementation est l'introduction, dans le seuil limite d'exposition quotidienne au bruit, d'un malus en fonction du type de source sonore responsable de la perturbation. Ces correctifs

traduisent le fait qu'à niveau sonore égal, un bruit de trafic aérien est qualifié de plus désagréable qu'un bruit de trafic routier, qui lui-même est jugé plus désagréable qu'un bruit de trafic ferroviaire.

La deuxième avancée importante en termes de bruit dans l'environnement est la directive européenne de 2002 relative à l'évaluation et la gestion du bruit dans l'environnement [DIRECTIVE-02]. Sa particularité est d'introduire au niveau européen (et pour la première fois dans la réglementation française [ORDONNANCE-04]) la nécessité d'informer le public sur son environnement sonore sous forme de cartographies. A ce titre, la directive propose l'indice L_{den} . Cet indice énergétique (qui est la moyenne énergétique des niveaux avec des pondérations suivant les périodes de soirée et de nuit) caractérise le bruit des grandes infrastructures de transport ou le bruit dans les grandes agglomérations. Dans ce cas, la notion de gêne à long terme est assez bien décrite par un indice basé sur le caractère bruyant des agglomérations ou des grands axes routiers. De plus, les indices énergétiques considérés sont mesurables assez facilement à l'aide d'un sonomètre intégrateur et leur prédiction est également modélisable de façon satisfaisante par des logiciels. Pour ces situations, la réduction des nuisances sonores passe par la localisation des points critiques et par la diminution des niveaux sonores.

A l'échelle d'un quartier, plus proche de l'échelle d'usage de la ville par les riverains, beaucoup de rues ne sont pas soumises aux grands flux de circulation. De ce fait les niveaux moyens journaliers y sont acceptables. Par contre, ce qui caractérise ces espaces urbains, ce sont les événements sonores qui s'y produisent et qui leur donnent une « couleur ». Dans un travail de thèse, Valérie Maffiolo a demandé à des riverains de répondre librement à la question « qu'est-ce que le Paris sonore pour vous ? ». Leurs réponses étaient centrées autour des lieux, des activités et des sources qui caractérisent ces lieux. De cette manière, V. Maffiolo a mis en évidence « l'influence de la signification des sources sur le jugement agréable/désagréable des ambiances sonores urbaines » [MAFFIOLO-99]. Dès lors, la caractérisation objective de l'environnement sonore par un niveau moyen sur une période journalière risque d'être limitée pour traduire le ressenti des riverains. Pour illustrer ce propos, nous pouvons citer une recherche sur les « zones 30 » [ROZEC-99], récemment aménagées dans nos villes dans le but d'améliorer le confort sonore de ces désormais « quartiers tranquilles ». Il apparaît qu'en général, dans ces espaces, le niveau sonore moyen journalier ne diminue pas, bien que l'ambiance soit ressentie comme plus agréable. En effet, bien que le bruit routier soit désormais marginal dans ces quartiers, il l'est au profit de bruits d'activités humaines, qui deviennent prépondérantes, démontrant ainsi que la mesure énergétique n'est pas toujours la plus adaptée pour interpréter le ressenti des riverains.

Il semble donc que, pour évaluer le confort sonore du quartier, c'est sur la composition de l'ambiance sonore que les gens se basent. Dans cette voie, des études ont montré que l'évaluation des ambiances sonores est sous la dépendance non seulement des individus et des contextes d'écoutes [VIOLLON-99, DUBOIS-98], donnant signification¹ et identité aux lieux, mais aussi des activités, des pratiques et des usages dans lesquels ils sont impliqués [MZALI-00]. L'identification de la source ou des sources comme « objet bruit » est ainsi primordiale dans l'évaluation de la qualité d'une ambiance sonore [LAVANDIER et al -03] car elle lui donne un sens. Par conséquent, la caractérisation d'une ambiance sonore passe par la mesure de nouveaux indicateurs acoustiques fondés sur l'évaluation subjective des différents bruits.

Pour l'instant, l'acousticien dispose, pour répondre à la réglementation, d'un outil de diagnostic : le sonomètre. Celui-ci mesure le niveau de pression acoustique pondéré A L_{Aeq} . La mesure se limite donc la plupart du temps à un niveau énergétique pondéré par la réponse fréquentielle de l'oreille à 40 phons. L'information relative à la composition de l'environnement sonore n'est pas traduite par cette mesure. S'il réalise en complément une mesure du signal audio, l'acousticien peut ensuite allouer aux sources entendues leur niveau sonore correspondant. Néanmoins, cette tâche est fastidieuse car elle nécessite, en plus de l'enregistrement du signal audio, sa réécoute et le codage des sources sur la courbe du L_{Aeq} . Ainsi, sans instrument capable de « reconnaître » la composition d'un environnement, il est évident que l'acousticien aura des difficultés à qualifier une ambiance sonore autrement que par son niveau.

C'est dans la perspective de mieux comprendre les réactions de non-satisfaction vis-à-vis d'un environnement sonore de quartier que le bureau d'étude acoustique LASA et l'université de Cergy-Pontoise, spécialiste dans la description d'environnements, unissent leur savoir-faire. Le but de ce travail est double.

Il vise tout d'abord à mieux comprendre la perception qu'ont les populations de leur environnement sonore lorsqu'elles sont en ville. Les sources sonores entendues et leur signification étant au centre d'une évaluation qualitative, elles sont directement intégrées comme variables dans notre étude.

¹ Dans ce mémoire, le mot *signification* désigne les propriétés de l'objet qui engagent l'intentionnalité du sujet et de ses actes.

A partir d'une meilleure compréhension de l'intégration des « objets bruit » dans notre perception, ainsi que de l'étude croisée entre leurs propriétés physiques et leurs propriétés perceptives, ce travail tend ensuite à proposer des pistes pour élaborer un nouvel indice qui soit plus proche d'un ressenti que la simple mesure du niveau sonore.

Cet indice, descriptif du désagrément sonore, a deux vocations principales qui représentent à terme des besoins concrets de notre société actuelle : être utilisé par les intervenants du milieu urbain et être intégré dans des cartographies mises à la disposition du public. Les cartographies permettent non seulement d'anticiper un environnement sonore, qu'il soit calme ou bruyant, mais représentent également un outil de communication utile aux pouvoirs publics ainsi qu'aux collectivités. Bien que les plans d'actions prévus par la directive doivent être avant tout conçus afin de « prévenir et de réduire, si nécessaire, le bruit dans l'environnement et de protéger les zones calmes contre l'augmentation du bruit » [DELCAMPE-02], tôt ou tard il faudra définir aux zones autres que bruyantes des critères afférents plus précis que le niveau de bruit reçu : type de sources, influences de la morphologie urbaine sur les émissions (ronds-points, feux rouges...). L'implication du public, informé à travers les cartographies, est également prévue tant dans la proposition que dans l'élaboration des plans d'actions. Un manque de représentativité de leur environnement proche se traduit par un risque de limitation de leur champ d'action.

Dernier point : comme nous l'avons fait remarquer plus haut, sans outil spécifique capable d'analyser la composition d'une ambiance sonore, il est ardu d'en extraire les sources occurrentes. Notre travail se propose donc d'explorer des voies permettant d'identifier de façon automatique les sources de l'environnement sonore urbain.

Puisqu'il traite de la description et de la perception de l'environnement sonore qui sont par définition complexes et variées (expressions numériques temporelles sujettes à des fluctuations très grandes et à des taux de reproductibilité faibles, impossibilité de fixer avec rigueur les aspects d'une situation), ce travail s'inscrit dans le domaine des sciences de l'imprécis. Dans l'ouvrage intitulé « Les sciences de l'imprécis », Abraham Moles le définit en ces termes : « Contrairement à l'étude rigoureuse de phénomènes régis par des lois fonctionnelles rigides et des démonstrations rigoureuses, il est caractérisé par des concepts vagues, l'incertitude de ses déductions, la faible portée des séquences de raisonnement qu'il peut construire (faible propagation de la causalité) et des liaisons entre les éléments successifs d'une chaîne de pensée de faibles corrélations » [MOLES-95]. Le but d'une étude faisant parti du domaine des sciences imprécises, puisque c'est leur nature, est d'extraire des concepts et de s'en servir pour parvenir à une *prédictivité* - même si

c'est une *prédictivité* réduite – et surtout de proposer ou de suggérer des opérations à effectuer sur le monde extérieur. Dans ce sens, nous ne pourrions en effet pas prédire avec exactitude à partir d'une mesure si une ambiance sonore est qualifiée d'agréable ni trouver la solution définitive à son amélioration. Néanmoins nous pourrions, en analysant les sources présentes, prédire les tendances hédoniques et entrevoir ainsi des actions en vue d'améliorer la perception de la qualité sonore. A terme, une telle étude permettra de réaliser des cartographies à l'échelle de l'utilisateur qui, correctement effectuées et bien exploitées, représenteront un outil précieux pour informer les citoyens et construire des plans d'actions pour améliorer leur cadre de vie.

II. Organisation du mémoire

Traiter des sons et de leur perception nécessite la description du phénomène acoustique ainsi que la connaissance de certaines bases théoriques de la perception, de méthodes d'évaluation de la perception et des outils d'interprétation des jugements auditifs. Par conséquent, il est important d'exposer une revue sommaire des bases théoriques pluridisciplinaires sur lesquelles le travail présenté s'appuie ou dont il s'inspire, ainsi que de présenter un état de l'art non exhaustif des connaissances sur la perception sonore en rapport plus ou moins étroit avec notre sujet. C'est l'objet du premier chapitre.

Le deuxième chapitre est axé sur la partie perceptive de ce travail. Il tourne autour des expériences réalisées lors des trois années d'études. A partir d'écoutes de vingt stimuli d'une durée de quinze secondes chacun représentant une ambiance sonore urbaine et mélangeant des sources telles que des véhicules, des voix et des chants d'oiseaux, des auditeurs sont invités à estimer sur des échelles psychométriques des jugements relatifs à un espace tridimensionnel des sources sonores. Les trois dimensions qui caractérisent les « objets sonores » sont leur prégnance, leur présence et leur proximité. Une deuxième série d'expériences (portant sur les mêmes séquences sonores) est axée sur la perception de l'intensité et du désagrément sonore. Une troisième expérience porte sur une séquence globale qui est une recombinaison continue des vingt stimuli. La prégnance, la présence, la proximité, l'intensité et le désagrément sont alors évalués globalement.

Des analyses statistiques multidimensionnelles (analyse en composantes principales et analyse factorielle des correspondances), faites à partir des réponses des sujets, nous permettent de comprendre les sensibilités perceptives des auditeurs. Par ce biais, nous avons mis en évidence

par exemple que les auditeurs perçoivent les voitures d'une manière différente de celle des autres sources (cyclomoteurs, motos, bus...). Ainsi, l'extraction de descripteurs physiques propres aux différentes sources est entreprise. Nous calculons une série de critères relatifs à l'intensité et à la durée d'apparition des sources sonores. Ces critères sont extraits des courbes de L_{Aeq} codées manuellement en fonction de l'apparition de chaque catégorie de source sonore. Nous mettons par la suite en relation les descriptions objectives (critères physiques) et subjectives (descriptions extraites des tests psychoacoustiques) par des analyses de corrélations, le but étant de trouver les mesures physiques pouvant « expliquer » les dimensions perceptives présentées, selon le type de source sonore entendue. La réduction des dimensions par les analyses factorielles et leur corrélation avec des mesures physiques nous permettent alors d'établir des modèles de désagrément sous forme de régression linéaire multiple. La principale caractéristique des modèles subjectifs et objectifs ainsi obtenus est d'intégrer la perception de sources sonores en complément du niveau afin de décrire le désagrément. Ces critères perceptifs sont d'ordre temporel. En parallèle, nous présentons une étude reprenant le même mode opératoire, mais cette fois-ci réalisée in situ. Cette étude, qui s'est déroulée au mois d'avril 2005 dans Paris, nous permet de mesurer des rapprochements entre la perception de séquences de courtes et la perception de séquences de moyennes durées. De plus, il est possible de confronter des données recueillies en laboratoire et in situ. Enfin, nous avons la possibilité de comparer des évaluations d'ambiances différentes (rue circulée, parc et marché).

Le troisième chapitre traite de l'identification automatique des sources sonores de l'environnement urbain. L'identification est importante si l'on veut automatiser la mesure du désagrément prenant en compte les sources sonores. Deux techniques sont exposées. La première s'appuie sur une simple description spectrale et temporelle des sources. A partir de spectres en bande de tiers d'octaves propres aux sources sonores, associés à leur temps d'apparition, nous allons voir qu'il est possible, dans une certaine mesure, d'analyser en temps réel l'activité des sources [DEFREVILLE-03]. La seconde technique s'appuie sur un système innovant d'extraction automatique de descripteurs musicaux, à partir du signal acoustique. EDS (Extractor Discovery System) est un outil informatique développé très récemment par SONY CSL afin d'obtenir une meilleure représentation de titres musicaux par leur contenu et de permettre l'extraction automatique de descripteurs musicaux pertinents dans le contexte de la norme Mpeg7 (il sera ainsi possible de choisir un morceau dans une large base de données musicales sur des descripteurs tels que l'énergie de la chanson ou son tempo, ou bien sur la présence ou non d'une voix ou d'un instrument en particulier...) [ZILS-04].

Cet outil est un moyen d'appliquer de manière aisée un grand nombre d'opérateurs de traitement du signal existants aux sons de l'environnement. Ainsi nous obtenons des fonctions capables de fournir des descriptions très précises des variations spectro-temporelles des sources sonores étudiées.

D'une façon plus générale, ce travail montre que l'identification de sources particulières ne peut pas se faire sans la prise en compte de l'ensemble des sources existantes dans l'environnement sonore urbain. Nous nous rapprochons dès lors d'un domaine qui est la taxonomie¹ des sons.

Le mémoire s'achève sur une conclusion générale dont la fonction est de synthétiser les résultats expérimentaux et d'ouvrir sur des perspectives dont ces derniers peuvent faire les objets.

¹ Taxonomie ou taxinomie : études théorique des bases, lois, règles, principes d'une classification. Définition du *Le Petit Robert*, édition 2004.

CHAPITRE 1. DU SON AU PAYSAGE SONORE, du signal acoustique à l'image auditive

I. Introduction à l'état de l'art

Le son, défini en tant que sensation auditive engendrée par une onde acoustique [N.F.S.30-001], semble être partout dans notre culture [FERRINGTON-03]. Ne pas l'entendre nous rend mal à l'aise, fait ressurgir des émotions [WRIGHTSON-99], mais, à forte intensité, il peut également être douloureux. Un son peut se présenter sous différentes formes : de la musique, du bruit, une voix... suivant la façon dont nous le percevons, l'interprétons, l'écoutons [MOLES-81]. Les sons constituent des repères spatiaux et temporels, créent des émotions et peuvent agir sur notre comportement immédiat. L'origine d'un son provient d'une mise en déplacement brusque de particules entraînant une déformation du milieu. Cette déformation se propage de proche en proche, par l'action des molécules qui se déplacent à une vitesse donnée de leur position d'équilibre pour y revenir immédiatement, après une phase d'élongation, entraînant du même coup le mouvement des molécules voisines et ainsi de suite. Ce mouvement provoque une déformation du milieu caractérisée par de très faibles variations relatives de pression et de masse volumique. A travers un processus adiabatique, le déplacement des particules crée une onde acoustique caractérisée par son équation de propagation exprimée généralement en fonction de sa vitesse et de sa pression.

Dès qu'un milieu est formé de particules dites élastiques, une onde acoustique peut être créée ou transmise par déformation de ce milieu suivant les propres caractéristiques physiques de celui-ci. Le rythme des oscillations des particules autour de leur position médiane est traduit physiquement par la fréquence du son produit ou par leur longueur d'onde. La plupart des sons de notre environnement sont complexes, c'est-à-dire qu'ils sont composés de plusieurs fréquences à la fois, parfois ordonnées selon des règles harmoniques, et leur intensité est variable dans le temps. La directivité, la propagation et l'absorption sont, selon les fréquences, également sujettes à des variations suivant le milieu où le son se propage et les obstacles qu'il peut rencontrer.

L'onde acoustique est donc « colorée » (absorbée, isolée, diffractée) par les propriétés physiques du ou des éléments (densité, forme, température, humidité...) dans lequel elle se propage, rencontre ou se crée et porte en quelque sorte leur signature. Barry Truax exprime cette propriété en ces termes : « ...the sound arriving at the ear is the analogue of the current state of the physical environment, because as the wave travels, it is charged by each interaction with the environment »¹ [TRUAX-84]. Cette signature est donc utile au système auditif qui en tire des informations sur son environnement.

Ces illustrations montrent qu'un son peut être également défini comme une indication physique d'une perturbation brusque en termes temporels et minime en termes d'intensité d'un état d'équilibre particulière (sens physique), permettant au système auditif de nous renseigner sur notre environnement (sens perceptif) comme par exemple la nature ou la cause de la perturbation ou sur le milieu ambiant.

Pour notre travail sur l'environnement urbain, il est important de savoir quelles utilités peuvent avoir ces sons pour l'être humain et de comprendre la réaction de l'homme qui est conditionnée par la façon dont il perçoit son environnement sonore.

II. Rôle et théories de la perception auditive

2.1. Approche « psychologique »

Il est de tradition de distinguer sensation et perception. La perception étant définie comme une somme de sensations elles-mêmes définies comme événements psychiques élémentaires (conscients ou inconscients) résultant d'une modification de l'environnement et impliquant un traitement minimum d'information par le système nerveux central à la suite d'une stimulation d'un organe sensoriel [BAGOT-99]. En audition, cette distinction entre sensation et perception n'est plus aussi robuste, comme le rappelle Rainer Guský [GUSKY-97]. En effet, la frontière entre les deux termes n'est pas évidente selon les cas : des sensations définies comme « simples » posent parfois plus de problèmes à l'auditeur que des perceptions dites « complexes ». Par exemple, le seuil de niveau sonore masqué est souvent appelé niveau de sensation. Cependant la réponse d'un sujet fait non seulement appel à sa sensibilité auditive mais également à un

¹ «Le son qui arrive à nos oreilles est une analogie du présent état de notre environnement physique. En effet, comme l'onde voyage, elle est chargée des interactions de l'environnement.»

traitement cognitif tel que l'évaluation des « coûts et bénéfiques » de répondre que le stimulus a été détecté ou pas. Il existe des procédures pour évaluer la stratégie de réponse du sujet (Théorie de la Détection du Signal). Autre exemple : nous parlons souvent de sensation de désagrément ou de sensation de gêne sonore alors qu'elle est clairement le résultat d'une évaluation affective du son entendu. Pour notre travail, nous continuerons à adopter une approche simple visant à définir la perception comme étant une somme de sensations.

Le débat entre sensation et perception nous renvoie à la question du rôle que le système perceptif doit accomplir. En général, avant d'être un mode de connaissance des choses, la perception est l'activité vitale de tout organisme en contact avec son milieu. Selon Bergson [BERGSON-00], la perception est déterminée d'abord par les nécessités de l'action. Son rôle central est l'activité vitale et toutes les autres idées et concepts sont des extraits de cette théorie générale. Merleau-Ponty insiste sur le fait que dans un acte perceptif, le sujet percevant n'est pas un spectateur passif de formes objectives. Les objets qui sont investis par nos préoccupations subjectives sont valorisés dans le champ perceptif [MERLEAU-PONTY-45]. Le monde perçu est le reflet de nos joies, de nos grimaces et de nos préjugés.

Dans notre travail expérimental, nous nous efforcerons de préciser quelle est l'action que le sujet est sensé effectuer lorsqu'on lui demande d'évaluer son environnement sonore.

2.2. Approche « évolutionniste »

Le rôle que l'on attribue à la perception auditive, souvent considérée comme passive en opposition avec la perception visuelle, est issu de discussions portant sur la nature fondamentale et évolutive du processus perceptif. Selon la théorie d'inspiration darwinienne [DARWIN-59], les systèmes perceptifs se sont développés et façonnés à travers le temps pour nous permettre d'acquérir et de traiter les informations de notre environnement. L'ensemble des informations sonores produites par l'environnement (biotope acoustique) peut être divisé en deux catégories principales : elles sont de nature inorganique ou de nature communicative [RADAU-67]. Partant du principe que les vertébrés partagent certaines fonctions auditives fondamentales (discrimination, localisation, analyse fréquentielle, analyse de scènes auditives) élaborées à partir de contraintes et pressions environnementales communes [POPPER-97], le système auditif des vertébrés a évolué (pour percevoir et différencier les sources sonores) selon les espèces, suivant deux types d'influences [POPPER-00]. L'une, engendrée par les sons inorganiques comme les sons du tonnerre, du vent, de l'eau et l'autre, générée par les sons dits communicatifs tels

l'expression orale, la musique, les vocalisations animales. Pour reprendre une discussion sur le forum d'auditory.org [AUDITORY-05], en termes de communication, la voie auditive a un certain nombre d'avantages sur les autres sens. Contrairement à la communication visuelle, elle passe à travers les obstacles comme les arbres, les roches, retenant la majeure partie de son information (à part peut-être sa localisation). Contrairement à la communication tactile, elle fonctionne même sans proximité. Enfin, contrairement à la communication chimique portée par les odeurs, elle est capable de changer rapidement au cours du temps. Ces caractéristiques uniques ont sans doute affecté la façon dont elle a évolué.

Ainsi, l'étude des spécificités des sons de l'environnement dans lequel notre organisme s'est développé depuis très longtemps peut constituer une base de recherche pour la connaissance de nos modalités perceptives [SHEPARD-81]. Des études statistiques de descripteurs temporels de bas niveaux¹ ont relevé des régularités sur les sons naturels (vocalisations humaines et animales, musiques, sons environnementaux) qu'il est élégant de mettre en relation avec les propriétés du système auditif et les mécanismes de codage neuronal des informations acoustiques [ATTIAS-97, NELKEN-99, SMOLDERS-79].

De même, Björk [BJÖRK-84] a montré que la perception de sons naturels (telles les vocalisations animales et humaines, ou les bruits de vent dans les arbres) est décrite à travers les deux émotions fondamentales caractérisant l'évaluation subjective d'un environnement, appuyant ainsi la thèse selon laquelle la phylogenèse a adapté le système perceptif auditif et émotionnel pour être sensible aux sons naturels. Une de ces émotions est le plaisir, représentée par un axe bipolaire plaisir/déplaisir, la seconde étant représentée par l'axe somnolence/éveil selon l'approche dimensionnelle des émotions de Russell [RUSSELL-80]). A ce sujet, des expériences ont montré que les sons, selon leur répartition spectrale et leur intensité, activent un organe primitif qui nous vient des poissons et qui est placé dans une petite cavité de l'oreille interne : le saccule. Cet organe envoie des signaux à l'hypothalamus qui contrôle entre autres les fonctions du plaisir, de l'équilibre et du désir sexuel à travers notamment la production de dopamine. En croisant des données subjectives relatives à la sensation de plaisir que produit sur des auditeurs un panel de sons aux réponses chimiques de leur saccule, il est apparu que certaines fréquences (aux alentours de 300 à 350 Hz) et certains niveaux (aux alentours de 90 décibels) procurent un gain hédonique et ont un effet sur l'équilibre [TODD-01]. Ainsi, la communication animale à niveau

¹ Descripteur de bas niveau : ici, propriété acoustique qui n'a pas de signification évidente du point de vue sonore pour les auditeurs. On peut citer par exemple la propriété statistique appelée « 1/f » [VOSS-77].

sonore souvent élevé, la musique amplifiée, percussive, entraîneraient une réponse sacculaire qui associerait au plaisir une envie de se mouvoir.

L'habilité de nos sens à acquérir des informations sur leur environnement, en particulier les sons inorganiques, a été utile à des fins de survie [RISSET-95]. L'acuité de l'ensemble des systèmes de capteurs sensitifs, a sûrement été cruciale pour la conservation des espèces. Ainsi, dans le cadre de l'audition, notre structure physiologique perceptive, déterminée par notre capacité à localiser des sources sonores, à traiter des communications avec détail, et à être averti du danger, est en constante évolution et s'adapte, se transforme, en résonnant avec les stimuli sonores produits par l'environnement naturel.

En écho à l'idée générale selon laquelle le rôle de l'audition, à travers les propriétés physiques des stimuli lui parvenant, a influencé de manière significative l'évolution du système auditif, trois définitions communément admises par la communauté scientifique sont formulées par des auteurs de la façon suivante :

« ...The function of our hearing is to inform us as reliably as possible about the world of sounds around us »¹.

[PLOMB-99 cité par NEUHOFF-04]

« The main purpose of the auditory system is [...] to provide relevant information about objects and events in the environment in relation to the perceiving organism and his/her intended actions »².

[GUSKY-97]

« The generic task of hearing is to acquire and process information from and about the environment, e.g., to identify sound sources and to assess their state of operation in space and time »³.

[BLAUERT-97]

¹ « La fonction de notre système auditif est de nous informer le plus sérieusement possible sur le monde des sons qui nous entoure ».

² « Le but principal du système auditif est [...] de fournir des informations pertinentes sur les objets et événements de notre environnement par rapport à l'organisme percevant et à ses actions prévues ».

³ « La tâche générique de notre audition est d'acquérir et de traiter l'information sur notre environnement, donc d'identifier les sources sonores et d'évaluer leur état dans l'espace et le temps ».

La pertinence de ces définitions est flagrante lorsque l'on s'imagine en déplacement pédestre urbain, à l'écoute de notre environnement sonore. Ainsi, notre perception auditive, qui a été initialement façonnée pour survivre doit pouvoir, pour Plomb, nous fournir une représentation sensorielle précise des informations acoustiques nous parvenant. Guský met l'accent sur l'importance de l'action comme moteur de la perception. Celle-ci va conditionner la décomposition des informations pour nous aider à organiser notre activité par rapport à elle. En milieu urbain, nous pouvons suggérer que chaque source sonore a une importance perceptive propre et notre représentation mentale appropriée à notre action dépendra de l'activité que nous sommes en train d'effectuer. Il est donc important de fixer le cadre de l'action pour évaluer l'importance perceptive de chacune d'entre elles. Enfin, nous retiendrons des propos de Blauert qu'au-delà de l'identification même des sources, nous sommes sensibles à des propriétés particulières telles que leur disposition et leur comportement.

Avant de pouvoir interpréter notre perception de l'environnement sonore, il est important de comprendre comment notre système auditif parvient à une représentation mentale de cet environnement.

2.3. Théorie cognitive

Sur la base de ces définitions sur le rôle de l'audition, plusieurs approches coexistent sur le mode de fonctionnement de l'audition et se différencient sur l'interprétation qu'elles ont de la relation audition - stimuli sonores. D'un côté, des auteurs ([JAMES-90], [HELMOLTZ-85]) soulignent l'importance de l'image des objets extérieurs émettrice d'ondes sonores qui est centrale dans le rôle perceptif et qui provoque en nous des sensations à travers un processus psychique. Helmholtz situe la formation de données sensorielles et leur interprétation comme étant le point central de la perception. La théorie, ayant déjà été pressentie par ces auteurs, selon laquelle la décomposition de l'information sensorielle auditive permet la naissance d'une perception cohérente a évolué pour être représentée actuellement par l'approche cognitive de l'audition.

Plus généralement, l'approche cognitive conçoit la perception comme le résultat de l'ensemble des opérations mentales qui permettent de donner une signification aux entrées sensorielles [BAGOT-96]. Son objectif est alors d'explicitier ces différentes opérations. En d'autres termes, l'étude des aspects cognitifs de l'audition a pour objectif de comprendre comment l'information auditive contribue fondamentalement au développement de la connaissance [MCADAMS-94]. Pour cela, le traitement de l'information est divisé en différents processus auditifs (organisation

perceptive, processus attentionnels, mémoire, reconnaissance...) dont la compréhension et l'étude des interactions permettraient, s'ils étaient mis bout à bout et implémentés dans un robot, d'utiliser l'information sonore reçue pour évoluer correctement dans son environnement [BREGMAN-94].

La première étape de traitement est l'analyse des scènes auditives ou organisation perceptive décrite par Bregman [BREGMAN-90]. C'est une des parties les plus abouties du point de vue théorique et qui laisse la place à tout un champ d'investigations informatiques (C.A.S.A.). Elle décrit les processus d'habileté que le système auditif emploie pour faire la distinction d'une simple source dans un environnement bruyant. Cette habileté est aussi importante pour une personne qui converse avec des amis dans une ambiance de « cocktail party » [BROADBENT-58] que pour un animal qui tente de repérer son prédateur parmi les bruits de la forêt. Cette théorie qui s'est construite à partir de nombreuses expériences psychoacoustiques postule que la stratégie du système auditif, pour pouvoir réunir en une unité perceptive l'ensemble des signaux, provenant dans une période de temps donnée, d'une seule source sonore de l'environnement, est d'utiliser des régularités acoustiques générales. Ces régularités sont des relations systématiques entre les propriétés acoustiques des sons issus d'une même source. Par exemple, le système auditif peut utiliser le fait que de nombreux sons naturels ont une nature harmonique. Si ce dernier détecte deux sous-ensembles de composants dont les membres sont tous des multiples d'une fréquence fondamentale différente, il peut conclure que le signal est le mélange de deux sons harmoniques. Ces régularités sont d'une telle élégance qu'il est intéressant de les noter :

- Il est extrêmement rare que des sons n'ayant aucun rapport entre eux démarrent et s'arrêtent exactement au même moment.
- Les propriétés d'un son isolé tendent à se modifier de façon continue et lentement (ou les propriétés d'une séquence de sons issues de la même source tendent à se modifier lentement).
- Lorsqu'un corps sonore vibre à une période répétée, ses vibrations donnent naissance à un pattern acoustique dont les fréquences des composants sont des multiples d'une même fréquence fondamentale.
- La plupart des modifications qui surviennent dans un signal acoustique affecteront tous les composants du son résultant de manière identique et simultanée.

Ainsi l'analyse des scènes auditives résulterait, pour une grande part, des capacités du système auditif à réunir les composantes qui présentent ces régularités et, au contraire, à séparer celles qui s'en écartent.

Les processus de traitement utilisant ces régularités lors de l'analyse des scènes auditives et qui nous permettent d'organiser mentalement en sources un environnement sonore acoustiquement très complexe, sont des processus qui dépendent fortement des caractéristiques du stimulus. Ils sont présumés automatiques et sont dit de « bas niveau ». Les traitements des niveaux plus « cognitifs » (de « haut niveau ») sont plus variables car ils dépendent des connaissances antérieures du sujet, de ses attentes, de ses motivations, de ses schémas cognitifs préexistants [BAGOT-02] mais également des ses émotions et de sa sensibilité [VÄSTFJÄLL-02].

Les rôles de ces deux types de traitements ou processus, qui contribuent à la formation d'une perception auditive selon un schéma cognitif, sont décrits par Reinier Plomb [PLOMB-02] en ces termes :

« Audition concerns the way in which the sound stimulus is processed both peripherally and more centrally so that its specific characteristics are preserved. Cognition concerns the way in which our previous experience with sounds is used to interpret the new signals. These two components of the process are frequently denoted the bottom-up and top-down aspects of perception, respectively. ¹»

Comme nous allons le voir plus loin, le désagrément sonore est souvent bien « expliqué » par la mesure du niveau sonore, c'est-à-dire que plus un bruit a un niveau énergétique élevé, plus il paraîtra désagréable. C'est une tendance. Néanmoins, la théorie cognitive nous informe que, après le stade de la décomposition de l'environnement en « objet auditifs » [MCADAMS-97], un jugement perceptif est influencé par des facteurs relatifs à la signification du bruit entendu. Notre travail fait l'hypothèse que dans un environnement sonore composé de multiples sources, l'évaluation du désagrément sonore par la mesure énergétique peut être complétée par l'introduction de « variables » liées aux sources sonores entendues. Ces variables rendent compte d'effets cognitifs non évaluables par une simple et unique mesure de niveau sonore. Le travail ici n'est pas d'expliquer ni même comprendre la structure et le fonctionnement des effets cognitifs

¹ « L'audition a trait à la manière dont les stimuli sonores sont traités à la fois de manière périphérique mais également de manière plus centrale de façon à préserver ses caractéristiques. La cognition a trait à la manière dont notre expérience passée avec les sons est utilisée pour interpréter de nouveaux signaux. Ces deux composantes sont couramment et respectivement appelées les processus ascendants et descendants de la perception.»

mais d'adopter une démarche simple qui consiste à mettre en relation les entrées du système auditif (les sons, les sources) aux sorties (leur évaluation par des jugements auditifs).

2.4. Théorie écologique

D'un autre côté, la théorie écologique introduite par Gibson dans le domaine visuel [GIBSON-66] et étendue en partie dans le domaine auditif, substitue la perception de l'« objet » initiateur du son décrite ci-dessus à celle de l'« événement » supposé avoir créé le son. La fonction première de l'organe sensoriel relatif à l'audition est donc d'acquérir des renseignements sur la direction d'un événement, nous permettant de nous orienter, et sur la nature de l'événement, nous permettant son identification [MACPHERSON-95]. (La source sonore se réfère aux objets à partir desquels les sons émanent, les événements sont considérés comme actions produites par les sources et provoquant des ondes acoustiques [GYGI-01]. Ici la source sonore a une signification identique à celle de l'« objet »). De ce fait, l'écoute « événementielle » nous permet d'identifier des caractéristiques physiques très spécifiques ; à la fois des objets, telles que la largeur de barres qui viennent d'être percutées [LAKATOS-97], la longueur de tiges tombées au sol [FREED-90], la dureté de maillets cognés [CARELLO-98], la taille, le type et la forme d'assiettes qui se cassent [KUNKLER-00], mais également des actions telles que la distinction entre la montée et la descente de marche d'escalier [GAVER-88], la configuration de claquements de mains [REPP-87] ainsi que l'anticipation de la trajectoire de sources sonores [NEUHOFF-04].

Le système auditif est donc sensible à certaines propriétés lui permettant de prélever directement les informations utiles de l'environnement sonore : les relations physiques entre les objets et les événements d'un côté et leur potentialité d'action d'un autre côté. Ces propriétés sont appelées *invariants* et nous permettent par exemple de reconnaître le son d'une voiture quelle que soit sa vitesse, son régime moteur, que ce soit un son synthétique ou une voiture passant en pleine rue ainsi que nous inciter à nous écarter si on a l'impression qu'elle peut nous atteindre. Ainsi, un individu dirige son comportement sur la base de propriétés invariantes provenant de son environnement. Ces invariants répondent alors à des besoins essentiels tels que se mouvoir, se protéger, se nourrir etc. Le corollaire de cette hypothèse est que le système auditif (à travers le temps et l'apprentissage) a évolué et a appris à prélever directement (et sans traitement comme dans la théorie cognitive) les informations utiles de l'environnement sonore.

La capacité de résonance avec les propriétés physiques de l'environnement que notre système auditif a développée à travers le temps au contact des sons naturels permet de fournir certaines explications à des phénomènes étudiés. Par exemple, des auteurs ont constaté que la différence d'intensité perçue est plus importante, à écart constant, pour des sons croissants que pour des sons décroissants. L'explication « écologique » fournie par Neuhoff [NEUHOFF-98] est que cette surestimation est le résultat d'une adaptation naturelle ou « ajustement » du système auditif à la structure du monde physique¹. Ce biais naturel a permis un avantage sélectif car une intensité croissante peut signaler un mouvement de source vers un organisme et ainsi provoquer l'éveil et avertir de l'arrivée d'un prédateur.

De nos jours, en environnement urbain et dans une certaine mesure, le son de pétard du cyclomoteur n'aurait-il pas quelque peu remplacé le rugissement du lion dans notre représentation du danger ?

Sans prétendre trouver les invariants qui rentrent en « résonances » avec notre structure neuronale [MACPHERSON-95], nous permettant une reconnaissance des sons de l'environnement ou des propriétés physiques des matériaux [ROUSSARIE-99], nous tenterons dans la troisième partie de la présente étude qui concerne la reconnaissance de sources, de nous inspirer de ce concept d'invariant, essentiellement dans le domaine spectral, pour identifier les sources à l'origine d'événements sonores.

2.5. L'écoute sonore

A partir d'une stimulation sonore reçue et à travers des expériences antérieures, notre système auditif nous renseigne sur une forme, un mouvement ou un matériau impliqué dans la production du son, plutôt que sur l'identité même de l'objet [CASEY-98]. S'inspirant de cette théorie pour décrire les modalités sensorielles auditives, William Gaver [GAVER-93] définit l'écoute de tous les jours (« everyday listening »), c'est-à-dire celle que nous sommes communément amenés à utiliser, comme étant l'écoute des événements plutôt que des sons.

L'écoute d'une ambiance sonore urbaine passe non seulement par l'identification des sources en présence mais également par l'extraction d'informations utiles associées aux événements sonores. De l'écoute d'une voiture qui passe dans une rue, nous en déduisons sa taille, sa vitesse, le type de rue dans laquelle elle roule. Ces informations ainsi extraites sont qualifiées « d'utiles »

¹ Des auteurs ([CANEVET-99] et [TEGHTSOONIAN-00]) peuvent avoir une analyse différente.

pour le promeneur dans le sens où ce sont elles qui vont influencer son comportement. L'écoute musicale [GAVER 1993] (ou réduite selon Schaeffer [SCHAEFFER-66]) se réfère quant à elle aux propriétés qualitatives des sons et s'exprime en termes d'intensité, d'acuité... Lorsque aucune source sonore ne peut être clairement identifiée, l'écoute événementielle se substitue généralement à l'écoute musicale comme lors de l'écoute d'un bruit de fond urbain [DUBOIS-04].

Pour les tests psychoacoustiques qui sont effectués dans la présente étude, c'est naturellement l'écoute de tous les jours qui est employée par les auditeurs. De ce fait, cette écoute leur est simplement suggérée dans les consignes qui leur sont données en leur demandant de s'imaginer lors d'un déplacement habituel sans contrainte particulière de temps. Ainsi, si l'on s'appuie sur l'hypothèse que la perception des informations sonores est directement guidée par un comportement humain rationnel et orienté vers un but [BERNARD-94], on peut supposer que leur conduite adaptative vis-à-vis des informations sonores perçues serait sensiblement identique lors d'un déplacement in situ. Nous plaçons ainsi les sujets dans des « conditions perceptives » bien précises, afin de leur donner un contexte d'écoute. Il est évident que les conditions expérimentales (écoute au casque et en laboratoire) ne peuvent pas recréer les conditions d'écoute réelles lors d'un déplacement en ville, mais la qualité d'enregistrement, du casque et la mise en situation essaient de replacer l'auditeur dans le contexte souhaité.

Enfin, nous pouvons évoquer Amphoux ([AMPHOUX-91] cité par [VOGEL-99]) qui définit trois types de formes sonores qui chacune suscitent une écoute différente.

Le *fond sonore*, que l'on n'entend pas consciemment mais que l'on oit. Dans un contexte urbain nous parlerons alors de *bruit de fond*.

L'*ambiance sonore*, caractéristique d'un lieu, que l'on peut entendre sans écouter, mais que l'on ne peut écouter sans entendre. Dans ce travail nous distinguons trois types d'ambiances sonores urbaines : une ambiance de type *rue circulée de quartier* (différent d'une ambiance type boulevard), une ambiance de type *parc* et une ambiance de type *marché*.

Les *signaux sonores*, qui sont toujours entendus et écoutés. En contexte urbain, nous en faisons l'analogie avec les sources sonores ou objets sonores. Ceux-ci représentent, sans distinction de notre part, les objets producteurs de sons tels un cyclomoteur ou un oiseau, ou une caractéristique associée à l'objet producteur de son tels une voix d'adulte, un chant d'oiseau ou un passage de cyclomoteur...

III. Approche psychophysique de l'audition – Théorie et méthodes

3.1. La psychoacoustique comme discipline de la psychophysique

L'objet général de la psychophysique est l'étude des relations entre les variations de l'univers physique qui nous entoure et les variations psychologiques concomitantes [BAGOT-96].

La psychophysique s'attache donc à relier les sensations à des mesures physiques des stimuli. Elle ne se pose pas la question de savoir comment est traitée l'information nous parvenant (le système sensoriel est considéré comme une boîte noire [CANEVET-00]) mais elle cherche des relations directes sous formes d'équations ou de lois pour relier le domaine physique au domaine sensoriel. Dans la psychoacoustique, le stimulus est sonore et la sensation, auditive.

La description physique des stimuli se fait par des grandeurs objectives possédant toutes des unités de mesures correspondantes. Par exemple, un stimulus sonore pur et constant sera défini par son niveau de pression (en décibels, dB), sa durée (en seconde, s) et sa fréquence (en hertz, Hz). Ce sont des continuums physiques décrivant des grandeurs physiques considérées comme orthogonales entre elles, c'est-à-dire que chacune d'entre elle peut varier indépendamment des autres. Les sensations produites par ces grandeurs physiques, en dehors de la dimension temporelle, sont la sonie ou force sonore subjective et la tonie ou hauteur tonale. L'objectif de la psychophysique est également d'établir une correspondance entre les continuums du domaine physique et les continuums correspondants du domaine sensoriel, eux-mêmes qualifiés d'orthogonaux. Cependant chaque grandeur physique n'influe pas de manière exclusive sur une seule grandeur subjective. Par exemple l'augmentation d'intensité d'un son pur provoque une certaine variation de la sensation de sa hauteur tonale, c'est-à-dire que la dimension subjective (hauteur tonale) est approximativement le correspondant perceptif de la fréquence mais elle dépend également de l'intensité acoustique.

Le stimulus est l'objet central de l'étude psychophysique. En effet, le but de cette science étant de caractériser les sensations créées par notre environnement, le choix des stimuli a une importance centrale quant à l'utilité des lois établies dans nos sensations quotidiennes. Pour comprendre de manière fine la manière dont le système périphérique auditif code l'information sonore, l'expérimentateur emploie souvent des stimuli sans réelle signification ou ayant peu de

sens pour les auditeurs tels des bruits blancs et des sons purs. Le contrôle de leurs paramètres physiques permet de les faire varier et d'en recueillir les différences de sensations engendrées.

L'intérêt pratique de la psychoacoustique est de prédire les sensations auditives par la mesure. Dans notre étude, la sensation est le désagrément sonore et les stimulations sont des événements complexes dont le contrôle et la variation des paramètres physiques sont quasiment impossibles. Ainsi la mesure employée est avant tout identitaire, c'est-à-dire qu'elle se réfère au type de source ayant entraîné la stimulation sonore, avant d'être une analyse fine du signal sonore.

Les résultats reliant le monde de la physique à celui des sensations sont issus des propriétés unidimensionnelles des continuums. Ils sont extraits à partir de méthodes psychophysiques dont les quatre principales sont : l'estimation de grandeur, l'estimation de rapport, la production de grandeur et la production de rapport. Nous choisirons l'estimation de grandeur qui permet au sujet de quantifier ses sensations.

3.2. Les critères psychoacoustiques comme descripteurs de sensations auditives

Les études psychoacoustiques de structures sonores complexes se rapprochent d'un autre but de cette science : la description qualitative des sons. Pour cela, différents critères considérés comme unidimensionnels ont été définis en rapport à la sensation qu'ils procurent : ce sont les critères psychoacoustiques. Ils ont bien évidemment une description subjective mais également des modèles permettant leur modélisation plus ou moins fine selon le type et la complexité des stimuli.

3.2.1. La sonie

La sonie (« loudness » en anglais) est l'intensité subjective d'un son. Pour des signaux stationnaires et de courte durée (quelques secondes), des modèles de calculs ont été introduits et le modèle de Zwicker, désormais normalisé [DIN 45 631 / ISO R532B], a une très bonne correspondance avec le résultat de tests subjectifs.

Le niveau de pression acoustique pondéré A, implémenté dans tous les sonomètres, a également été créé dans le but d'incorporer des données psychoacoustiques du système auditif humain dans une mesure et a été conservé pour sa pertinence de l'évaluation de la nocivité vis-à-vis des bruits industriels [BOTTE-88]. Mais si l'on regarde sur un même graphique [Figure 0] la

superposition entre les lignes d'iso-sonies et la courbe de pondération A (pour des sons purs et en champ libre), on observe tout de même des différences notables. Si à de petits niveaux (20 phons) il y a une bonne correspondance entre les deux courbes, on remarque qu'à des niveaux plus importants, il existe des différences importantes, notamment aux basses fréquences et aux fréquences autour de 4000 Hz.

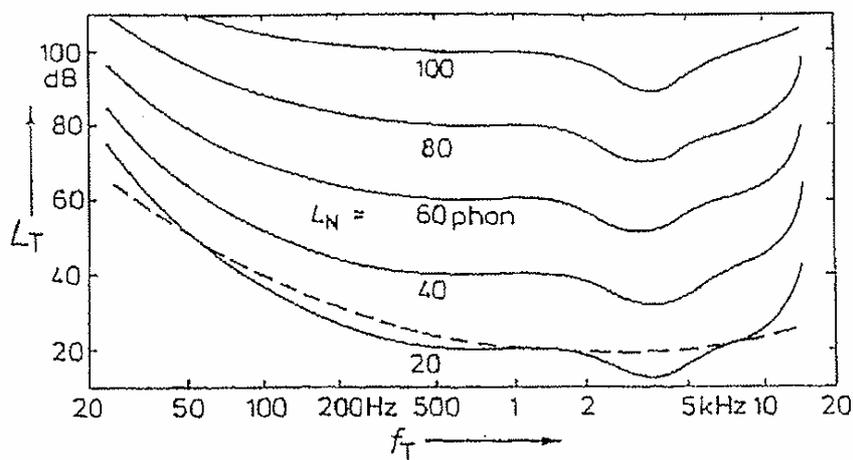


Figure 0 : Courbes d'isonies (écoute binaurale en champ libre) et courbe de pondération A (en pointillé) [DESS-01].

Cela s'explique par le fait que le calcul de la sonie prend en compte plusieurs effets auditifs non intégrés dans le décibel A : le facteur de transmission, l'influence des bandes critiques et le masquage spectral.

Le facteur de transmission rend compte de la transformation d'amplitude, variable selon la fréquence, du signal acoustique par suite de sa propagation dans l'oreille interne et l'oreille moyenne créant ainsi une légère résonance. Son effet se traduit par une amplification de pression pour les fréquences situées aux alentours de 4000 Hz. [Figure 1]

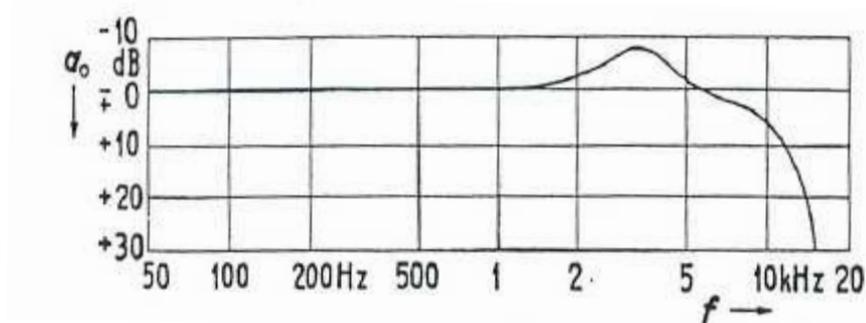


Figure 1 : Variation avec la fréquence du facteur de transmission entre le champ libre et l'oreille interne [ZWICKER-81].

Les bandes critiques sont en quelque sorte les filtres auditifs de l'oreille humaine car elles définissent la résolution fréquentielle de l'oreille. Ainsi, la sensation d'intensité des sons complexes croît avec l'étendue en fréquence de leur spectre dès que celle-ci dépasse une largeur appelée bande critique. Leur fréquence centrale, bien qu'arbitraire, a été fixée dans une classification qui consiste à segmenter la gamme audible en vingt-quatre intervalles juxtaposés gradués en Barks.

La répartition de l'énergie acoustique en bandes critiques est faite pour modéliser les capacités de filtrage mécanique de l'oreille interne. Les courbes d'effet de masque traduisent quant à elles les détails du filtrage ou masquage spectral. En effet, la sélectivité fréquentielle de l'oreille n'est pas rectangulaire et la figure 2 montre qu'un son de 1000 Hz n'excite pas uniquement la bande critique principale (bande 9) mais également les bandes critiques, proportionnellement au niveau d'excitation principal.

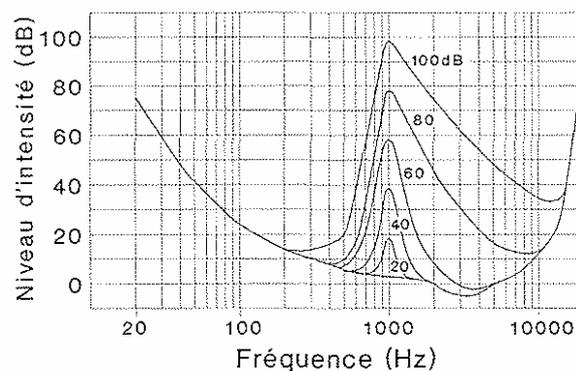


Figure 2 : Diagramme d'excitation d'un son de 1000 Hz à différentes intensités illustrant l'effet de masque. Ces courbes montrent l'élévation des seuils absolus en présence d'un son masquant de 1000 Hz et à différentes intensités indiquées sur les courbes. Par exemple pour une intensité d'un son masquant de 80 dB la zone de masquage est délimitée par une courbe passant par des seuils absolus tels que 50 dB à 800 Hz et 30 dB pour 3000 Hz [ZWICKER-99].

3.2.2. Calcul de la sonie

Le calcul de la sonie d'un son selon le modèle de Zwicker passe par le calcul de la sonie spécifique ou densité de sonie, elle-même nécessitant la détermination de l'excitation par le son dans chacune des vingt-quatre bandes de fréquences. A titre d'exemple, la figure 3 représente la densité de sonie pour la bande critique numéro 9. On dit que la sonie est spécifique à une bande critique. L'intégration de chacune des sonies spécifiques aboutit au calcul de la sonie totale.

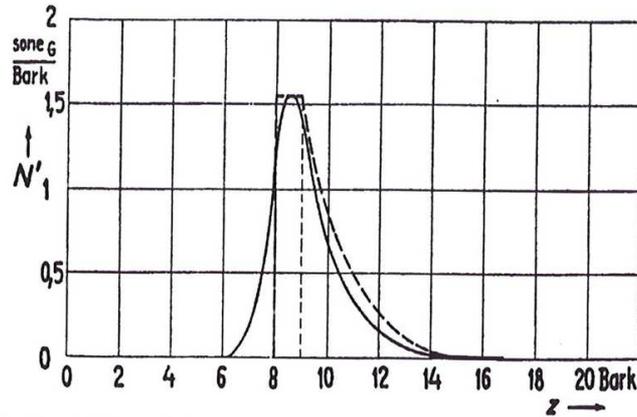


Figure 3 : Courbe de densité de sonie (ou sonie spécifique) en fonction de la fréquence en Barks produite par un son pur de 60 dB à 1 kHz. La courbe en pointillée représente l'approximation utilisée pour les calculs de sonie globale [ZWICKER-81].

Ainsi, pour calculer la sonie d'un son, on fait tout d'abord son analyse spectrale en bandes critiques. Pour chaque bande critique, on calcule son diagramme d'excitation. Cette excitation est ensuite transformée en densité de sonie. Puis la sonie totale est calculée en intégrant la densité de sonie sur les vingt-quatre bandes.

3.2.3. L'acuité

Les sons ne peuvent être évalués uniquement sur une échelle opposant les adjectifs fort et faible. Pour l'oreille humaine, les sons peuvent également paraître graves ou aigus, ou mats ou brillants, ou encore bruyants ou ayant une tonalité...

En conséquence une autre sensation a été définie comme étant l'acuité. Elle représente un facteur important dans la perception du timbre [BISMARCK-74] et est principalement déterminée par la forme de l'enveloppe spectrale et par la distribution en fréquences de l'énergie acoustique : à une forte concentration d'énergie dans les hautes fréquences correspond une forte sensation d'acuité [MEUNIER-04]. Son calcul passe également par la prise en compte du masquage fréquentiel et l'intégration des bandes critiques [ZWICKER-99].

3.2.4. La force de fluctuation et rugosité

Dans le cas de vibrations sonores dont l'enveloppe temporelle est sujette à des fluctuations, deux sensations auditives peuvent être relevées. Pour des modulations de fréquences faibles, la sonie du son change et l'on a une sensation de battements. C'est la force de fluctuation. Pour des modulations de fréquences plus élevées, la sonie reste à peu près constante mais une sensation de rugosité apparaît, comme pour un son qui serait produit en frottant une surface rugueuse [ZWICKER-99].

3.3. Les critères psychoacoustiques et le bruit

Le son de nos villes est souvent évalué comme étant bruyant dès lors qu'il est défini comme « non désiré ». Les principales sources de bruits en milieu urbain sont les artères fortement circulées (boulevards, boulevards périphériques, routes nationales...) où les sons mécaniques (véhicules, tramways) sont les sources dominantes. Nous avons également les bruits provenant d'équipements (centrale de traitement d'air en terrasse d'immeuble pour donner un exemple) et les bruits provenant de chantiers de construction. Nous appelons ces derniers respectivement bruits d'équipement et bruits de chantier.

Ces sources de bruits sont considérées comme gênantes et, au-delà de leur impact même sur l'audition, on leur reconnaît des effets sur l'homme qui peuvent être « extra auditifs » (effet sur le sommeil, sur la sphère végétative, sur le système endocrinien, sur le système immunitaire, sur la santé mentale) et subjectifs (gêne due aux bruits, effets du bruit sur les attitudes et les comportements, effets sur les performances et effets sur l'intelligibilité de la parole) [AFSSE-04]. Ces bruits sont caractérisés par un niveau souvent continu (en comparaison avec des bruits impulsionnels par exemple) et de forte puissance (L_{Aeq} supérieur à 60 dB(A) entre 22h et 6h et supérieur à 65 dB(A) le reste du temps, pour le bruit des infrastructures de transport terrestre [BRUIT-04]).

En dehors de toute considération non auditive (sensibilité, âge, culture, catégorie sociale...de la personne), ces bruits, et à travers eux leurs effets négatifs, sont généralement évalués par la mesure de leur dose énergétique. Sur des périodes relativement importantes (d'une journée à plusieurs années) la dose énergétique est évaluée par la mesure du L_{Aeq} . Sur des périodes plus courtes, la sonie de ces sons quasi stationnaires est également estimée de façon satisfaisante par la mesure du L_{Aeq} [NAMBA-04].

Pour des sons de durées plus courtes, les modèles de désagréments doivent tenir compte de la non-stationnarité des sons (pour les bruits de chantiers par exemple ou certains bruits routiers). La non-stationnarité des sons se manifeste au niveau perceptif par des variations de sonie, d'acuité, de rugosité...Ils font donc l'objet d'études spécifiques, car ce sont des critères psychoacoustiques pouvant rendre compte du caractère désagréable des sons ([BOULLET-05], [MEUNIER-02]).

De plus, les critères psychoacoustiques peuvent également rendre compte d'émergences spectrales qui produisent des effets de sifflement ou de bourdonnement et qui sont reconnues comme étant particulièrement désagréables (pour les bruits d'équipements par exemple).

3.4. Le timbre et les analyses multidimensionnelles

L'étude de sons complexes, périodiques ou non, amène également à définir la notion de timbre, sensation quant à elle à caractère multidimensionnel. Alors que le niveau est la correspondance physique principale évidente de la sonie, il n'y a pas de correspondance aussi évidente pour le timbre. Au départ, le timbre était défini comme étant ce qui fait que deux sons sont ressentis comme différents bien qu'ayant une sonie, une hauteur spectrale et une durée identiques [ASA-60]. Nous savons désormais que des facteurs tels que les transitoires (en particulier le transitoire d'attaque), l'enveloppe dynamique (forme temporelle) et l'enveloppe spectrale (position du centre de gravité du spectre), pour citer les plus importants, ont un rôle dans la différenciation du timbre d'instruments de musique. Les dimensions sont généralement obtenues par analyse multidimensionnelle [WINSBERG-1989] sur des tests d'estimations de similitude [MCADAMS-94]. Les notes sur les échelles de dissemblances sont censées refléter des comparaisons entre des composants spécifiques aux représentations mentales de chaque timbre.

Les algorithmes d'études multidimensionnelles projettent les jugements de dissemblances en une représentation spatiale du timbre où les distances euclidiennes entre les points correspondent aux dissemblances entre stimuli. La structure géométrique résultante reflète les dimensions perceptives sur lesquelles les auditeurs se sont appuyés pour évaluer le timbre ou, en d'autres termes, elle indique la structure de la représentation mentale qui leur a permis de faire des comparaisons ordonnées [QUEHL-01]. Une analyse comparée entre les dimensions de l'espace perceptif et les propriétés acoustiques des stimuli permet par la suite de donner une interprétation psychoacoustique à ces dimensions.

Cette technique est utile pour déterminer les dimensions perceptives qui permettent de différencier des sons isolés. Ce qui caractérise une ambiance sonore urbaine est la présence de sources imbriquées. L'utilité d'une analyse multidimensionnelle par jugement de dissemblances sur des séquences de ce type peut être limitée. En effet, les dimensions de l'espace perceptif extrait seraient directement relatives aux catégories ou types de sources sonores présentes dans les séquences [SUSINI-98].

L'autre technique d'analyse multidimensionnelle couramment utilisée est l'échelle sémantique différentielle. La base théorique de cette technique s'appuie sur le postulat que les perceptions de l'environnement extérieur ne sont pas le reflet unique de conditions ou d'états physiques et psychologiques, mais qu'elles sont également l'expression de connaissances inconscientes, comme par exemple le sens ou la signification des choses [QUEHL-01]. Autant la signification que l'analyse physique du stimulus déterminent l'évaluation subjective de ce dernier. Ainsi, la signification du son entendu peut être appréciée selon des qualificatifs choisis et évalués sur des échelles dont les extrémités sont représentées par des adjectifs opposés. A travers l'affirmation selon laquelle le langage représente un outil important pour décrire la signification d'un concept, cette technique permet de déterminer des dimensions reliant les sons et leur signification et d'établir des profils sensoriels. Les adjectifs employés peuvent être relatifs à des sensations (fort/faible, brillant/mat) à des émotions (stressant/rassurant...) ou au désagrément sonore (agréable/désagréable) et sont généralement obtenus soit après une description verbale libre et spontanée faite par les auditeurs sur les sons entendus, soit en s'appuyant sur la littérature existante.

De nombreuses études ([SOLOMON-58] pour des sons de sonars, [GABRIELSSON-79] pour des sons de haut parleurs, [BJÖRK-85] pour des sons naturels et [VON BISMARCK-74] pour des sons stationnaires) utilisant des échelles sémantiques différentielles, font ressortir de façon constante trois dimensions sensorielles principales expliquant à peu près à chaque fois les deux tiers de la variance des réponses. La première dimension est une évaluation hédonique (« evaluative » en anglais) car elle regroupe les échelles telles que bon/mauvais, heureux/triste, beau/laid ou plaisant/déplaisant. La seconde dimension a trait à la notion de puissance (« potency ») avec des échelles telles que grand/petit, fort/faible, lourd/léger. Enfin la troisième dimension indique une notion de mouvement (« activity ») et se réfère à des antonymes tels qu'actif/passif, rapide/lent ou chaud/froid.

La méthode d'échelle différentielle nous a semblé intéressante pour notre étude. Ainsi les notions d'agrément et de force sonore ont été étudiées sur l'ensemble des séquences. Néanmoins, de manière à rester cohérent avec les approches psychologiques, évolutionnistes, cognitives et écologiques, l'échelle sémantique différentielle a été aussi appliquée à chaque « objet source » des environnements urbains écoutés par les sujets. Cela pourrait traduire la troisième dimension liée à l'activité des sources.

IV. Le paysage sonore

C'est à partir du concept de qualité sonore que nous allons, dans la partie bibliographie, tenter de déterminer comment les gens évaluent et se représentent mentalement un environnement sonore.

4.1. La Qualité Sonore

Nous regroupons les objets, les bruits et les sons naturels dans la catégorie « sons isolés » dans le sens où ils sont étudiés de façon indépendante, sans interaction avec d'autres sons.

4.1.1. Les objets

Le concept de qualité sonore environnementale n'est pas encore tout à fait établi. Cependant, s'inspirer des nombreuses études sur la qualité sonore des produits nous permet de dégager les principales orientations pour définir et évaluer l'environnement sonore urbain.

Jusque dans les années quatre-vingts, l'évaluation de la qualité sonore était réduite à l'étude du niveau sonore global. Beaucoup d'objets étaient en effet considérés comme trop bruyants ce qui rendait leurs sons désagréables. Ils pouvaient également représenter un danger pour l'oreille dans le cas d'exposition prolongée. Pour améliorer le son produit par un objet, il suffisait de réduire son niveau d'émission. Une fois ce cap franchi, il est devenu évident que l'étude de la qualité sonore ne pouvait se réduire à l'étude purement quantitative du son, ne serait-ce que parce que deux sons ayant le même niveau en dB(A) ou en sonie peuvent « sonner » totalement différemment. L'évaluation de la qualité sonore a pris un caractère multidimensionnel à travers des recherches psychoacoustiques et psychologiques, opérant ainsi un glissement vers des études qualitatives. Celles-ci ont permis de mesurer les sensations induites par le son, d'étudier sa convenance par rapport à sa fonction et également de mesurer sa compatibilité vis-à-vis des attentes de l'auditeur. Blauert définit de la façon suivante l'étude de la qualité sonore : « Adequacy of a sound in the context of a specific technical goal and/or task¹ » [BLAUERT-94].

Dans l'article « Sound-Quality evaluation – a multi-layered problem » [BLAUERT-97], les auteurs soutiennent que l'appréciation qualitative d'un son est le fruit d'une comparaison entre un

¹ « L'adéquation d'un son dans le contexte d'un but ou d'une tâche précise. »

événement auditif avec un ensemble d'attentes concernant le produit. Le son doit être en adéquation avec la fonctionnalité du produit et le sujet a déjà une idée prédéterminée de ce qu'il doit être. Ainsi, la qualité sonore n'est pas une propriété intrinsèque car le jugement émis par les utilisateurs dépend de leur situation cognitive (mémoire, attente...) et émotionnelle. L'étude de la qualité sonore d'un produit peut être décomposée en trois niveaux non totalement indépendants : une étude physique permettant une description objective du son, une étude psychoacoustique prenant en compte les caractéristiques de la perception sonore humaine et une étude psychologique traitant de la manière dont les gens perçoivent et interprètent les sons.

Gusky [GUSKY-97] suggère trois critères psychologiques qui permettent d'évaluer une « bonne » ou « mauvaise » qualité sonore. Tout d'abord, une bonne compatibilité entre l'information sonore disponible et l'action requise. Elle permet par exemple à l'utilisateur de s'assurer et de se rassurer que sa coûteuse voiture de sport n'émet pas un son de cyclomoteur au démarrage. Il y a souvent, dans l'évaluation de ce critère, une interaction forte entre les informations visuelles et sonores disponibles. Ensuite, il y a un côté plaisant-déplaisant du son qui permet à l'auditeur d'évaluer qualitativement et instantanément la qualité d'un événement sonore. L'évaluation hédonique du son est influencée par l'image cognitive de celui-ci, c'est-à-dire par sa fonction et sa signification. Elle ne peut donc être décrite uniquement par l'évaluation de qualité sensorielle et donc par des descripteurs psychoacoustiques tels que la sonie ou la rugosité. Des analyses multidimensionnelles appliquées soit à des jugements de dissimilarités entre paires de sons soit à des échelles sémantiques différentielles sont des moyens cités pour faire ressortir des différences qualitatives qui ne peuvent ressortir avec l'utilisation de critères physiques ou psychoacoustiques. Enfin une bonne qualité sonore va de pair avec la possibilité d'identifier distinctement la source sonore à évaluer ainsi que la vitesse avec laquelle la tâche d'identification est réalisée.

Ainsi, une bonne qualité sonore d'un produit ne signifie pas forcément qu'il a un son plaisant. Par exemple, un avertisseur sonore a pour fonction de mettre en garde ainsi que d'être reconnu et identifié immédiatement. Ainsi, le son qu'il produit ne sera pas forcément agréable mais sa qualité sera déterminée par l'adéquation entre le son perçu et sa fonctionnalité.

4.1.2. Les bruits

Namba [NAMBA-94] définit le timbre d'un son comme étant à la fois un indice permettant d'identifier une source et une qualité pouvant se manifester oralement par une impression affective. Ainsi, dans le contexte de la recherche sur le bruit, l'auteur propose d'utiliser le terme de qualité du bruit (« noise quality ») à la place du mot timbre. De même, il argumente l'utilité de diviser le processus d'évaluation en deux étapes pour étudier la qualité sonore. Tout d'abord il y a une identification ou une reconnaissance de la source sonore sur la base de son timbre et son niveau à l'aide de fonctions cognitives telles que les processus perceptifs, la mémoire, l'attente. Puis une fois la source identifiée ou pas, le son est évalué par une impression affective exprimée par un jugement hédonique c'est-à-dire du côté plaisant-déplaisant. Quelques études ont permis de mieux cerner les facteurs entrant en compte dans l'identification d'une source. C'est ce que nous verrons dans la partie identification.

L'impression subjective du sujet, quant à elle, est multidimensionnelle et doit donc être analysée comme telle. Deux techniques principales existent [GUSKY-97] et permettent de faire ressortir des facteurs perceptifs : le différenciateur sémantique et l'étude de similarité. Cette dernière, ou comparaison par paire, présente le désavantage de ne pouvoir être effectuée qu'en laboratoire. De plus, elle nécessite un nombre limité de stimuli car le test peut devenir long et fastidieux.

Des études de laboratoires japonais sur des sons mécaniques isolés tels que des moteurs, des ventilateurs...font ressortir trois facteurs concernant la qualité du bruit. Le premier est relatif au côté agréable-désagréable (ou plaisant-déplaisant) du son, le second au côté puissant et le dernier au côté métallique. Le côté puissant est corrélé à des mesures de niveau (L_{Aeq} ou sonie), le côté métallique avec le critère l'acuité du son. Par contre aucune mesure, ni même des critères psychoacoustiques [ZWICKER-90] établis pour évaluer cet attribut perceptif, ne se rapporte au côté agréable-désagréable. Les auteurs concluent quant à l'importance de la signification subjective du son entendu dans l'appréciation hédonique d'un son.

4.1.3. Les sons courants et sons naturels

Une autre étude [BONEBRIGHT-01] a été menée sur des sons usuels ou courants (« everyday sounds ») dans le but d'étudier leur structure perceptive et d'y relier les critères acoustiques correspondants. Soixante quatorze sons allant du bruit de clefs à celui du velcro en passant par celui produit par l'ouverture d'une cannette ont été présentés à des auditeurs qui ont eu pour tâche de former des groupes de similarités en se basant uniquement sur les propriétés des sons et

non pas leur cause. Ces mêmes sons ont ensuite été évalués sur douze échelles sémantiques. Une analyse multidimensionnelle permet d'extraire trois dimensions perceptives. Certains attributs, et les mesures acoustiques correspondantes, caractérisent une dimension alors que d'autres sont distribués sur deux dimensions. La première dimension oppose les attributs lents, dispersés, brillants et durs (associés à des sons comme de l'argenterie qui tombe, un verre qui se casse, un smash de basket-ball) à leur antonyme (associés à des sons comme du sel qui coule, du velours frotté, un tableau essuyé...). La seconde reflète le côté temporel du son en opposant des sons saccadés (pompe à vélo, ciseau) à des sons stationnaires (ventilateur, eau qui bout), l'attribut perceptif bas/haut correspondant (« low/high ») étant assez peu explicite. La troisième dimension oppose des sons qualifiés de relaxants, doux et plaisants (ouverture et fermeture de thermos, bruit de pas) à des sons de perceuse et de sèche-cheveux.

Dans cette expérience, malgré des consignes claires, tout semble indiquer que c'est sur le type d'action en relation avec l'objet (donc l'événement sonore) que les auditeurs se sont appuyés pour différencier ou regrouper les sons présentés. Cela tend à montrer que la perception auditive est orientée sur la cause ou la source et non pas les caractéristiques purement auditives ou intrinsèques des sons.

Dans une étude de 1984, Björk [BJÖRK-84] part du principe que la qualité sonore d'un stimulus auditif peut être évaluée en termes d'émotions ainsi qu'en termes de sensations auditives et se propose de les étudier dans le but de mieux comprendre les facteurs influençant la qualité sonore. Les stimuli sont égalisés en niveau de pression (à 60 dB(A)) et sont joués à l'envers de telle façon que l'auditeur ne puisse pas reconnaître leur origine. L'auteur s'affranchit ainsi volontairement de tout effet de sens et oblige de ce fait les auditeurs à se concentrer sur l'information spectrale. Sur quinze sons naturels sélectionnés (voix, chants d'oiseaux, cris d'animaux, bruits du vent...) et estimés sur vingt-quatre échelles sémantiques différentielles représentant à la fois des émotions et des attributs auditifs, il trouve cinq facteurs principaux expliquant 91% de la variance des résultats. Quelques tendances sont alors observées. La première dimension qui représente le côté plaisant (« evaluation ») du son est corrélée à la mesure de la rugosité, la seconde dimension représentant le côté intéressant du son (« activity ») est plutôt reliée à la mesure de la hauteur ou de l'acuité. Enfin, la troisième dimension importante représente la puissance du son (« potency ») et est corrélée à la mesure de la sonie.

Les études de qualité par des analyses multidimensionnelles sur des sons isolés permettent donc d'extraire des dimensions perceptives discriminantes. Le croisement de ces dimensions avec

des attributs physiques permet généralement et de façon assez précise de passer d'une description subjective à une description objective. Comme ce sont des sons isolés et donc entendus comme tels, il est possible par la suite de modifier les caractéristiques physiques de sorte à en améliorer la qualité sonore. Par contre, pour des séquences réalistes comme celles sur lesquelles nous travaillons, c'est-à-dire représentatives d'un environnement sonore, comment ces relations objectives-subjectives peuvent-elles se retrouver ?

4.2. Paysage sonore in-situ

R. Murray Schafer, musicien et compositeur, décrit dès 1967 le paysage sonore comme une pièce musicale où les sons naturels sont donnés à écouter comme des instruments de musiques et où il nous accorde de le composer comme tel [SCHAFER-67]. Malheureusement les bruits homogènes et anonymes de trafic ont désormais submergé la diversité des paysages sonores existant dans nos villes.

Pour étudier le paysage sonore, ou plutôt la perception de l'image que l'on en a, il faut analyser cette image en termes d'espace, de densité, de volume et de silence. Dans ce but, les auteurs ont décrit les concepts analytiques de tonalité, de signal, d'empreinte, d'objet et de symboles sonores [KALLMANN-04].

La tonalité (« keynote ») fait référence au son dominant et omniprésent, perçu habituellement comme fond sonore. C'est en quelque sorte le bruit de fond de nos villes composé de bruits mécaniques. Les signaux sont les sons situés au premier plan, que l'on écoute consciemment et qui renferment des messages ou des informations. Parmi les signaux que nous pouvons distinguer, les empreintes sonores sont des objets sonores bien localisés. L'objet sonore, initialement décrit par Pierre Schaeffer [SCHAEFFER-58], est ici la plus petite particule autonome du paysage sonore. Enfin, les symboles sonores qui consistent en des signaux qui impliquent des réactions personnelles émanant de niveaux d'association collectifs et culturels tels le bruit de cloches ou le son d'activités traditionnelles (sirène des pompiers les premiers mercredis du mois...).

Une première taxonomie permet de différencier les paysages sonores qualifiés de « *hi-fi* » en opposition aux paysages sonores qualifiés de « *low-fi* ». Un paysage « *high-fidelity* » se caractérise par un rapport signal sur bruit plutôt favorable dans le sens où les sons sont entendus distinctement sans masquage ou encombrement. Il est variable en termes de niveau, de spectre et de rythme

alors que le « *low-fi* » a souvent un niveau constant créant un « mur du son » et isolant l'auditeur de son environnement. L'analyse de 2500 heures d'enregistrement sonore a permis de constater que les vocalisations d'animaux et d'insectes occupent des faibles bandes fréquentielles spécifiques permettant à tous d'occuper un espace « réservé » [KRAUSE-93]. Nous le retrouvons particulièrement dans la campagne où l'écoute est moins bruyante, plus attentionnée et donc plus informative que beaucoup de sites urbains, et notamment les boulevards ou rues extrêmement circulées où il est très difficile d'évaluer, grâce à l'écoute, notre distance par rapport aux sources.

L'étude d'un paysage sonore passe également par l'apprentissage de l'écoute [FERRINGTON-03]. Elle nous permet de mettre en place des stratégies pour différencier les sons, reconnaître les indices, donner un sens aux sons et de faire ainsi contrepoids à la domination de la culture visuelle. De ce fait, nous pouvons en découvrir les sons significatifs de part leur individualité, leur nombre et leur domination. Notification, classification, perception, morphologie et symbolisme sont ainsi des outils d'analyse qui permettent une connaissance détaillée d'un paysage sonore [SCHAFER-67]. La figure 4 illustre par exemple le rythme du paysage sonore de la côte ouest de la Colombie Britannique (Canada).

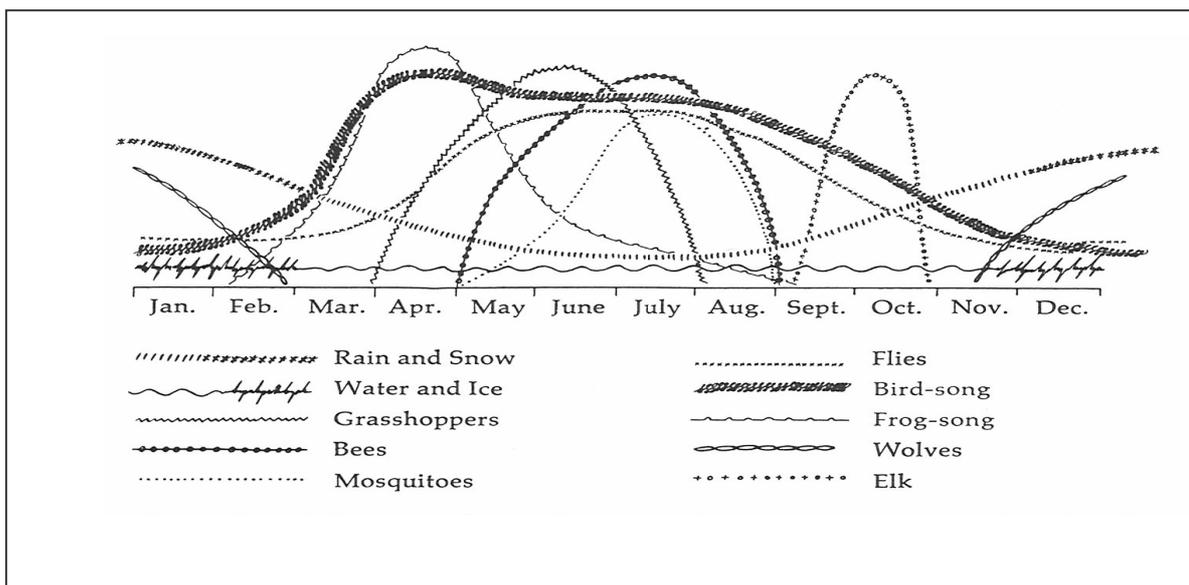


Figure 4 : Cycles du paysage sonore naturel de la côte ouest de la Colombie Britannique montrant les volumes des différentes sources sonores sur une année.

Enfin, le quotidien sonore de l'*homo urbanus* est caractérisé par un niveau sonore souvent important. Même lorsqu'il est en situation de pouvoir réduire la quantité d'informations lui parvenant, walkmans, chaînes hi-fi, télévisions...le replongent dans un monde sonore.

Cette boulimie de sons, ou la fuite de moments calmes et des ambiances hi-fi, représente selon Kendall Wrightson [WRIGHTSON-99], une peur consciente ou inconsciente, d'avoir à affronter des sensations. En effet, le calme ne tend-il pas à ramener des émotions à la surface ?

Actuellement, l'étude du paysage sonore consiste à identifier, afin de les préserver, les paysages sonores présentant des aspects positifs [BERGLUND-04], ou à caractériser les autres en vue de les rendre plus agréables. Ceci essentiellement en milieu urbain, comme l'indique la directive européenne du 25 juin 2002 relative à l'évaluation et la gestion du bruit dans l'environnement [DIRECTIVE 2002/49/CE-02]. Celle-ci vise essentiellement à informer la population de son environnement sonore sous forme de cartographies. Pour l'instant, seules des cartographies de bruits sont mises en place, laissant de côté tout un aspect perceptif essentiel dans l'étude et l'amélioration du paysage sonore.

Dans la suite du travail, nous parlerons indifféremment de paysage sonore ou d'environnement sonore.

Le paysage sonore s'applique aussi à la ville. Dans le paysage sonore urbain, on peut discerner plusieurs ambiances sonores qui sont décrites qualitativement en faisant appel à la mémoire et à l'expérience sonores d'usagers de la ville. Une méthodologie combinant deux modes de questionnement, graphique et verbal, mise en place par Amphoux [AMPHOUX-91], et reprise dans la thèse de Maffiolo [MAFFIOLO-99], permet d'accéder à des propriétés différentes de la représentation mentale des bruits. Le dessin sonore et la verbalisation libre appliqués à des sujets sur le « Paris sonore » ont permis de dégager des ambiances sonores pertinentes ainsi que les propriétés qui les caractérisent. Ainsi, dans l'esprit des gens, « le Paris sonore et les jardins publics se caractérisent par des dimensions acoustiques auxquelles sont étroitement associées des représentations de sources, du temps, des lieux et des situations dans lesquelles ces bruits ont été perçus ». Des ambiances ont été sélectionnées pour être enregistrées et reconstituées de façon réaliste à des auditeurs. Des procédures expérimentales telles que la catégorisation libre, la comparaison par paires ainsi que les échelles de dissemblance ont été appliquées sur un corpus de seize sons variant de 15 à 20 s. Elles ont montré que les sujets regroupent prioritairement les séquences sur la base du caractère (événementiel ou amorphe) et du contenu significatif (identification de sources). Les scènes événementielles sont décrites par des termes désignant les sources, les événements sonores et les activités qui s'y déroulent alors que les scènes amorphes sont décrites par des termes se rapportant aux propriétés physiques des sons car aucune source sonore n'est clairement identifiable. En plus d'avoir mis en évidence et testé la robustesse de cette

catégorisation, l'auteur conclut qu'il s'agit de deux traitements cognitifs distincts ; une scène sonore événementielle produit une écoute de reconnaissance (type « *top-down* » ou « *bas niveau* ») alors qu'une scène amorphe entraîne une écoute plus analytique de type « *bottom-up* » ou « *haut niveau* ». Autrement dit, l'appréciation hédonique d'une ambiance sonore est conditionnée par le type de scène entendue ainsi que par la signification de son contenu, c'est-à-dire par les sources sonores qui la constituent. La représentation agréable/désagréable serait donc davantage de nature catégorielle que de nature dimensionnelle.

Nous retiendrons de cette étude concernant les représentations mentales d'ambiances sonores qu'il y a lieu de différencier, dans le choix d'ambiances sonores à évaluer, les types de scènes *évènementielles* et *amorphes*. Les scènes amorphes, à mettre en connexion avec le terme « *low-fi* » employé par Schafer, ressortent généralement comme étant le propre des boulevards ou des rues largement fréquentées : sources difficilement identifiables, faible variation temporelle du niveau sonore [RAIMBAULT-03, LAVANDIER-02, MAFFIOLO-99]. Elles peuvent être qualifiées de désagréables du fait de leurs composantes essentiellement mécaniques [DUBOIS-04] et la mesure du niveau sonore représente un bon corrélat physique de leur représentation négative [RAIMBAULT-03]. L'étude des scènes évènementielles est plus complexe dans le sens où c'est la nature des sources la composant qui va avoir une influence sur le jugement qualitatif. Des ambiances sonores urbaines faisant partie de la catégorie évènementielle peuvent être un marché, un parc ou une rue circulée où les sources sonores sont identifiables. Ainsi, nous avons choisi, pour notre étude, de nous concentrer sur les scènes évènementielles qui ne sont pas aussi simples à évaluer que les séquences amorphes.

4.3. Les séquences sonores complexes : ambiances sonores urbaines

Les études sur les ambiances sonores requièrent des séquences extraites de l'environnement. Il est ainsi difficile de « jouer » sur tel ou tel paramètre physique comme la qualité des sources sonores occurrence (des véhicules en particulier) qui doit être étudiée en amont. Dans le cadre du projet S.V.E.N. (« Sound Quality of Vehicle Exterior Noise »), Notbohm et ses collègues [NOTBOHM-04] ont étudié l'influence de plusieurs paramètres (débit, largeur de voie, type de rue en U ou en tissu ouvert, vitesse des véhicules, niveau sonore) sur la qualité sonore de bruits de trafic urbain en faisant à chaque fois varier un de ces paramètres. Les auteurs ont ainsi extrait quatre séquences de 4 min chacune qui deux à deux ne diffèrent que d'un seul paramètre et se ressemblent dans la mesure du possible sur tous les autres. Ces séquences sont tout d'abord

évaluées par des auditeurs sur vingt-sept échelles sémantiques différentielles. Sur l'ensemble des séquences, une analyse en composantes principales révèle quatre axes factoriels expliquant 64.4% de la variance des réponses. Le premier axe est corrélé à des termes faisant référence à des émotions comme ennuyeux, (non) plaisant, (non) acceptable, (non) relaxant ou des caractéristiques négatives du son comme fort, intrusif, laborieux.

Cet axe révèle que le caractère désagréable est la première dimension à laquelle les auditeurs sont sensibles concernant des bruits de trafic urbain.

La deuxième dimension connote des émotions rattachées au danger car des descripteurs tels que surprenant, alarmant ou dangereux y sont corrélés.

Le troisième facteur est le rythme reflétant la structure temporelle de l'ambiance sonore avec des termes tels que rythmique, pulsatif, réverbérant.

Le dernier facteur oppose des séquences sur la base des sons naturels qu'elles contiennent. Quatre variables sont ensuite créées en moyennant la réponse des sujets sur les descripteurs reflétant un facteur. En comparant les résultats entre séquences, il en ressort que l'ambiance sonore d'un rond point est, en général, ressentie comme plus dangereuse, plus rythmée que celle d'une ambiance d'intersection.

Notons toutefois que le facteur décrivant les sons naturels (les auteurs n'indiquent pas s'il s'agit de voix ou d'oiseaux) a été jugé deux fois plus important pour le rond point que pour l'intersection. Cependant, lorsque les stimuli sont présentés aux auditeurs avec un niveau sonore représentatif de la réalité, l'ambiance sonore de la rue en U est jugée plus désagréable, dangereuse et rythmée que la rue en L, alors que lorsque les niveaux sonores sont augmentés (de 68 à 83 dB(A)) c'est la rue en L qui est jugée plus négativement.

Les auteurs émettent l'hypothèse explicative que, étant donné que ces deux stimuli contiennent un trafic peu dense caractérisé par des événements sonores, il se peut que l'augmentation du niveau fasse ressortir des informations relatives aux sources sonores et qui sont susceptibles de distraire ou déranger les auditeurs.

En croisant les données avec des populations allemandes et suédoises, les auteurs ont constaté un effet « culturel » sur l'appréciation hédonique divergente des séquences sonores.

Les résultats de cette étude laissent à penser que les auteurs ont ici étudié des paramètres (débit, largeur de voie, type de rue en U ou en tissu ouvert, vitesse des véhicules, niveau sonore) qui sont secondaires dans l'évaluation de la qualité sonore. Les principaux axes dégagés font ressortir sans surprise les appréciations hédoniques classiquement rencontrées (coté agréable/désagréable, coté éveillant, rythmé...) mais ne permettent pas de conclure de manière franche sur l'influence des paramètres étudiés. Le dernier axe dégagé ainsi que les analyses plus subtiles montrent que la composition sonore de l'environnement est un élément important qui n'a pas été pris en compte comme paramètre à faire varier.

Toujours concernant les ambiances sonores, Raimbault et al [RAIMBAULT-03] ont étudié la perception d'ambiances sonores urbaines de deux villes en se focalisant sur des ambiances de type événementielles : une ambiance de type boulevard, une ambiance de type rue piétonne, une ambiance de marché ainsi qu'une ambiance de type cour de récréation ; le but étant d'étudier si ces différents lieux représentaient des ambiances distinctes du point de vue perceptif et acoustique. Les sujets ont été sollicités au travers d'un questionnaire ouvert sur leur activité, leur connaissance antérieure des lieux et leur évaluation de l'ambiance sonore. Sur cette dernière, une liste de neuf adjectifs pertinents a été extraite afin d'appliquer la méthode du différenciateur sémantique. Il a également été demandé aux auditeurs de justifier verbalement leurs appréciations subjectives sur les notes attribuées. Une analyse factorielle permettant ainsi de visualiser les consensus dans les réponses des sujets a révélé trois facteurs expliquant 90% de la variance des réponses. Le premier facteur (67% de la variance) reflète le côté plaisant/déplaisant des séquences et oppose le boulevard aux autres lieux. Le second facteur (15% de la variance) reflète la dynamique sonore (organisé/désorganisé) et l'arrangement spatial, (régulier/irrégulier) et oppose les ambiances de type marché aux ambiances de type square. Le troisième facteur représentatif de la spatialité (proche/loin, présent/non présent) et de la clarté (distinct/désordonné) oppose des ambiances d'école à des ambiances de squares.

Une analyse croisée des données perceptives et acoustiques montre que l'on peut généralement associer des mesures physiques aux différents lieux étudiés. Des niveaux forts caractérisent les boulevards, des niveaux faibles caractérisent les jardins, un nombre d'émergences (rapport entre le nombre d'émergences et leur durée en seconde) élevé caractérise les écoles et les marchés. Aucune mesure physique ne caractérise de façon satisfaisante les places piétonnes.

A l'observation des résultats perceptifs, il ressort que les gens ont donné des appréciations contraires sur les mêmes stimuli, révélant ainsi un manque de consensus dû à deux écoutes distinctes d'un même événement sonore décrivant deux types d'attitude perceptive (ou conduite d'écoute). La première est l'écoute descriptive qui se manifeste par l'identification des événements sonores. Ainsi la nature des sources conditionne l'interprétation subjective de l'ambiance sonore. La seconde est l'écoute holistique où l'activité de l'utilisateur est sa principale préoccupation [RAIMBAULT-01]. L'évaluation hédonique d'une ambiance sonore est cette fois-ci conditionnée par des nécessités comme par exemple « pouvoir s'entendre », « être dans de bonnes conditions pour travailler » ou « se détendre pour faire ses courses ». Selon le type d'écoute, les jugements sur les échelles plaisant/déplaisant ou agréable/désagréable peuvent varier. On pointe ici du doigt le fait que la prise en compte de l'activité du sujet sort du cadre du désagrément pour rentrer dans celui de la gêne. C'est l'ambiguïté relevée par Väsffjäll [VÄSTFJÄLL-02] quand il dit que c'est le son et non pas l'homme qui doit être au centre d'une étude sur la qualité sonore. L'expérimentateur voulant étudier la qualité d'une ambiance sonore peut recueillir des jugements reflétant soit une évaluation affective (le son est au centre de l'étude), soit une réaction affective (le sujet est au centre de l'étude). Un son est jugé plaisant parce que sa sonorité, ou ce qu'il représente, est plaisant ou par ce qu'il nous « rend plaisant » ? L'évaluation de la qualité d'une ambiance sonore doit faire suite à une écoute descriptive et une réponse appropriée à une bonne qualité sonore est du type « ce son est plaisant » (ou procure du plaisir). Une écoute holistique met la personne et son activité au centre de l'étude et sa réponse est du type « ce son crée du plaisir » car le son est en adéquation ou n'interfère pas avec son activité. Dès lors que l'activité du sujet rentre ainsi en jeu dans l'appréciation du désagrément, nous nous rapprochons alors du concept de gêne sonore (voir paragraphe 5).

Dans les tests auditifs qui font l'objet de ce mémoire, nous souhaitons orienter les auditeurs vers une écoute descriptive de telle façon à avoir une évaluation de la qualité sonore. Dès lors, nous excluons toute possibilité que l'auditeur puisse se représenter une interférence entre les sons entendus et une activité trop orientée (par exemple le sujet ne doit pas s'imaginer être en conversation). Notre but est de recueillir des appréciations sur des situations qui peuvent être réelles. En demandant aux auditeurs de s'imaginer un déplacement tel qu'ils l'ont déjà vécu (par exemple un déplacement urbain pour un rendez vous sans être pressé, Annexes 4), nous plaçons les sujets dans des conditions sensiblement similaires d'activité tout en recréant une situation courante ce qui a pour effet de minimiser à la fois les divergences de point de vue et également les interférences entre l'activité du sujet et les sons.

De plus, nous avons vu que les études qui se sont attachées à décrire la qualité de l'environnement sonore urbain ont toutes en commun le fait qu'elles visent à extraire des attributs perceptifs ou sémantiques que les gens utilisent pour évaluer les ambiances sonores qui leur sont proposées en laboratoires. Si certains attributs sont liés à des propriétés acoustiques simples (par exemple le côté puissant peut être relié au niveau sonore, le côté brillant à la mesure de l'acuité), une partie non négligeable de la variance des jugements subjectifs n'est pas clairement expliquée par des mesures (par exemple le côté plaisant, fortement dépendant du type de source entendu, ne peut être que partiellement décrit par des critères psychoacoustiques comme la rugosité ou l'acuité).

Un environnement sonore urbain est complexe dans le sens où de nombreuses sources coexistent et sont variables en intensité comme dans le temps. Nous ne pouvons ainsi étudier des sons isolés sur des durées de quelques secondes. Le concept de qualité sonore décrit précédemment ne s'applique plus exactement. Nous avons vu que pour savoir comment les gens se représentent mentalement un environnement sonore, il est commode de leur demander de le décrire librement. Dans ce cas, ils décrivent typiquement les événements sonores entendus en les citant mais ne mentionnent pas les propriétés acoustiques de ces derniers [VANDERVEER-79], [BALLAS-82], [BALLAS-87]. Southwork et Anderson ont mis en évidence que l'évaluation qualitative et l'appréciation d'un environnement sonore dépendent de l'information sonore [SOUTHWORK-69] et donc de la signification des sources entendues [ANDERSON-83] en rapport avec l'attente que l'on a du lieu. Des sons mécaniques seront plus naturellement acceptés dans une rue que dans un parc. Ainsi, quand un auditeur évalue la qualité sonore d'une ambiance, il identifie les sources occurrentes et il fait l'adéquation entre ces sources et les lieux où elles sont perçues.

Nous proposons de définir la qualité sonore d'une ambiance en prenant comme base d'étude les sources qui la composent. En adéquation avec un lieu, ayant une fonctionnalité et entraînant une certaine attente auditive, c'est sur la reconnaissance et la répartition des sources sonores que s'établit l'évaluation de la qualité d'une ambiance sonore. Nous suggérons que la qualité sonore d'une ambiance résulte d'une évaluation subjective selon les termes agréable/désagréable et qu'en conséquence, pour pouvoir la décrire, il faut tenir compte de la nature et donc de l'identité des sources sonores la composant. Cette évaluation étant déterminée par des facteurs liés aux sources tels que le sentiment d'adéquation entre les sons et leur lieu d'occurrence, ainsi que la représentation mentale (positive, négative ou neutre) de ces dernières.

Lors de nos tests, le comportement intentionnel de l'individu dans cet espace est fixé de telle sorte à minimiser ses interactions avec les sons. Cependant ses états affectif (humeur) et émotionnel sont également des facteurs pouvant influencer la perception de la satisfaction de la qualité sonore. Lors de notre évaluation cet état évolue au cours de l'acte perceptif (et donc influence la perception). Nous les avons supposé neutres au départ de nos tests d'évaluation.

V. Gêne ou désagrément sonore ?

Nous avons vu que la première réaction effective qui pouvait être induite par les sons était reliée aux termes génériques de désagrément, agrément, gêne, plaisir, nuisance. Bien qu'il soit très complexe de délimiter une frontière entre ces termes et de donner une définition nette pour chacun d'eux, il semble néanmoins que nous pouvons extraire deux catégories qui ne sont, bien sûr, pas totalement indépendantes.

La gêne causée par un son ou un bruit est le terme représentant la catégorie la plus large qui soit (en anglais, le terme employé dans la littérature est « noise annoyance »). Elle peut être définie comme une réaction négative créée par un environnement acoustique et ayant un effet sur la santé ou le bien-être d'une personne [GUSKY-97]. De ce fait, la gêne est le résultat d'un son intrusif qui perturberait une activité comme par exemple une communication, un effort intellectuel, un moment de repos mais elle peut également résulter d'une non satisfaction vis-à-vis d'un environnement sonore [GUSKY-99]. Ainsi on peut distinguer deux principaux types de gêne. La gêne à long terme, qui est généralement modélisée par la mesure du niveau sonore sur de longues périodes, est de ce fait bien corrélée à un pourcentage de personnes gênées. La gêne à long terme, qui représente un jugement plutôt qu'une sensation ou une perception, est une somme de facteurs personnels et sociaux qui sont divisibles en deux catégories. Les modérateurs sont des facteurs indépendants du bruit tel l'âge, la sensibilité au bruit en général, la confiance envers les autorités en charge du bruit...alors que les médiateurs dépendent directement du son comme par exemple la peur de la source sonore, son évaluation subjective, une sensibilité vis-à-vis d'une source en particulier ...

La gêne sonore est un concept dont la difficulté d'établir sa définition reflète la complexité de sa modélisation. Il apparaît que, sur un ensemble très large d'études, seulement 30% de la variance des résultats sur la gêne soit expliquée par une relation prenant uniquement en compte la dose énergétique reçue. Il semble que la piste la plus prometteuse pour modéliser la gêne totale

est de la considérer comme un concept flou [BOOTLEDOREN-03]. Ce n'est bien entendu pas l'objet du présent travail.

La gêne à moyen terme est le résultat d'une perturbation ou d'une interférence de l'activité de la personne (perturbation du sommeil, perturbation dans une tâche intellectuelle) causée par l'exposition au bruit. Elle est suivie d'une réaction affective dans le sens où c'est le sujet qui est au centre de l'étude et dont l'effet du bruit peut lui induire des effets de stress. Le niveau sonore est encore un facteur important mais son évaluation en vue d'une modélisation physique est très difficile à effectuer car elle se fait avant tout en respect de la tâche du sujet. Les terminologies appropriées pour sonder la gêne sont du type : « êtes-vous gênés par cet environnement sonore pour effectuer cette tâche ? ».

Qualifier l'effet du bruit à court terme revient à faire l'évaluation affective de celui-ci en termes de qualité sonore, c'est-à-dire que c'est le bruit et sa signification cette fois-ci qui sont au centre de l'étude. En complément du niveau sonore perçu, les facteurs non acoustiques tels que les médiateurs ont une influence sur le jugement de ce que l'on appelle alors le désagrément sonore. Les termes qui permettent de traduire le désagrément sont le côté plaisant ou déplaisant ainsi que le côté agréable ou désagréable du son.

Pour notre travail, nous nous placerons dans l'étude de l'effet du bruit à court terme que nous appellerons désagrément sonore. Notre travail ne se situe donc pas au niveau de la gêne sonore.

Derrière les définitions proposées, il ressort un trait important qui est le caractère temporel de la qualité sonore. L'évaluation de la qualité, à travers le jugement du désagrément sonore, est délicate car elle reflète la perception du sujet et implique en conséquence sa mémoire des événements et leur intégration perceptive.

Ainsi, sur le long terme, une personne réduit la perception de son environnement sonore suivant deux paramètres. Le premier est le résultat de l'intégration de l'ensemble des événements sonores en une source unique et dominante, la seconde est l'impression de prégnance de cette source, directement liée à son niveau sonore. Le terme de gêne ainsi justement employé peut s'évaluer à travers la mesure d'un niveau continu sur une longue période (souvent une année). Ce niveau devant être pondéré suivant l'influence plus ou moins négative de la source dominante : voitures, avions, trains...

Sur le court terme, conséquemment à l'évolution de la composition de l'environnement, l'évaluation de la qualité sonore est en changement continu. Sur une durée très courte (de l'ordre d'une dizaine de secondes) un auditeur peut intégrer sans sélection l'ensemble des événements. Sur une durée légèrement plus longue, des stratégies de mémorisation se mettent en place et l'évaluation peut être différente car elle ne s'appuie plus sur les mêmes données perçues. Le travail qui suit se propose de mettre en relation des évaluations sur des très courtes séquences ainsi que sur des séquences pour lesquelles le sujet ne peut mémoriser l'ensemble des événements sonores.

VI. Conclusions

6.1. Synthèse

Dans une situation d'écoute habituelle, nous ne sommes pas confrontés à un seul son mais à un flux complexe d'ondes acoustiques venant de sources multiples, à partir desquelles nous devons extraire le son dont l'information nous est utile. L'acuité du système auditif a évolué pour le rendre réceptif aux informations qui étaient utiles à la survie de l'organisme. Mais maintenant que notre comportement n'est plus axé sur notre survivance, à quelles informations sonores sommes-nous sensibles ?

Notre but n'est pas de rechercher par quels processus auditifs ou mécanismes de perception notre système traite les informations lui parvenant, mais plutôt d'extraire des structures de notre perception dans un environnement sonore naturel et en l'occurrence urbain, d'interpréter la signification des dimensions perceptives évaluées et d'en déterminer les sources ou facteurs pour lesquels nous sommes les plus sensibles.

La qualité des ambiances sonores a souvent été évaluée via des questionnaires rendant compte des attributs perceptifs ou sémantiques d'ordre global et axés sur la qualité émotionnelle de l'environnement. Or, une ambiance sonore est avant tout caractérisée par un ensemble d'événements et représentée par les sources sonores la composant. La source sonore ou objet sonore représente le premier aspect de notre étude car c'est de leur signification que dépendent les jugements de qualité sonore. Notre travail se focalise sur l'étude de séquences événementielles [MAFIOLLO-99], caractérisées par la présence de sources sonores clairement identifiables. Nous nous concentrerons en particulier sur une ambiance de quartier de type rue circulée mais nous

verrons également, à travers un test in situ, comment les gens évaluent de manière séquentielle et globale des ambiances de type parc ou marché ouvert.

La plupart des études multidimensionnelles sur la qualité de l'environnement urbain font ressortir comme composante principale le caractère hédonique de l'ambiance sonore. Le second aspect qui est au centre de notre étude est donc le *désagrément sonore*. Ce terme est préféré à celui de gêne sonore car il représente une évaluation affective et non pas une réaction affective (en relation avec une activité ou un état d'esprit) par rapport au son perçu. Nous nous concentrons ainsi sur la qualité de l'ambiance sonore et de sa signification pour l'auditeur.

La force sonore est généralement le critère qui rend le mieux compte du désagrément, en particulier dans des environnements bruyants. Nous considérons ainsi l'évaluation de la force sonore (ou sonie) comme le troisième aspect central de notre étude.

L'évaluation de trois aspects précités (sources sonores, désagrément et sonie) va se faire à travers une série de tests psychoacoustiques. Les séquences sonores, qui sont de ce fait présentées à un panel d'auditeurs, sont représentatives d'une ambiance urbaine de quartier de type voie circulée (en opposition avec une voie de type boulevard).

Nous retenons également du cadre théorique que la perception est un processus actif, conditionné par des intentions et des buts [GUSKY-97]. Il est donc important de fixer la structure de l'action pour évaluer l'importance perceptive des trois aspects précités. Nous avons ainsi suggéré aux auditeurs de s'imaginer écoutant l'environnement sonore lors d'un déplacement urbain.

L'évaluation d'une ambiance sonore à travers le désagrément sonore, la force sonore et les objets sonores occurrents est de ce fait multidimensionnelle. Nous nous inspirons de la technique de l'échelle sémantique différentielle pour recueillir les données perceptives. Les qualificatifs employés sont évalués par la méthode de l'estimation de grandeur.

S'il est facile de comprendre que le désagrément et la force sonore peuvent être évalués chacun sur une échelle unidimensionnelle qui leur est propre, cela est moins évident en ce qui concerne les sources sonores. Il nous est donc utile de décomposer la perception des sources en un espace qui nous renseigne sur plusieurs aspects qui les caractérisent. C'est l'objet de la description de l'espace tridimensionnel présenté ci-après.

6.2. Espace descriptif des Objets Sonores

Nous venons de voir que l'identification et l'évaluation de la source sonore est un point central de notre recherche sur l'environnement sonore urbain. Par des méthodes psychoacoustiques que nous allons détailler par la suite, notre but est d'évaluer la perception relative que nous avons de chacune d'entre elles dans une ambiance globale. Il nous est donc utile de nous appuyer sur la définition d'un espace sonore subjectif qui soit représentatif d'un maximum d'informations que nous percevons en écoutant les sources indépendamment les unes des autres.

Le premier acte que notre système perceptif doit accomplir lors de l'écoute d'un environnement sonore sa décomposition en sources sonores afin d'y apporter une signification. Le terme d'*objet auditif* (« auditory object ») est généralement employé pour définir l'unité perceptive de l'audition qui précède l'attribution du sens (ou de la signification) qu'on lui donne [GRIFFITHS-04]. En d'autres termes, cet élément fondamental se réfère à « une représentation mentale d'un groupe d'éléments qui possèdent une cohérence interne dans leur comportement et qui sont ainsi interprétés comme provenant de [...] la même source sonore » [McADAMS-97]. Par exemple, le bruit d'une voiture roulant peut être un amalgame de chocs entre morceaux de métal, entre caoutchouc et bitume, accompagné d'une série d'explosions. Dans un environnement urbain, nous traduisons cette unité psychologique par le terme de source sonore, ou « objet sonore » par analogie visuelle.

Le principe de définir le schéma global de la perception de l'environnement sonore à partir de l'intégration d'unités perceptives distinctes rejoint le principe de l'analyse structurale. Celle-ci repose sur le fondement « qu'il est toujours possible de considérer la réalité comme un assemblage d'éléments reconnaissables, identifiables et codifiés, dont la combinatoire forme ce que l'on appelle la structure » [WOLOSZYN-98].

C'est avec ce même raisonnement que, dans l'ouvrage « Ambiances Architecturales et Urbaines », Philippe Woloszyn et Daniel Siret exposent un modèle perceptif des objets ambiants [WOLOSZYN-98]. Ceux-ci peuvent être des sons, des odeurs ou des objets visuels. A partir d'un environnement existant qui est une entité homogène et complexe, ils le décomposent, en un ensemble d'unités indépendantes appelées les « objets ambiants » ou simplexes. Chacun de ces objets est « déterminé par la perception que nous avons des interactions dans le temps et l'espace, suivant notre situation, entre les phénomènes physiques et l'environnement ». Cette perception

est représentée selon trois caractères transversaux représentant la dimension, la temporalité et la valeur du simplexe formant les trois axes de l'espace de perception.

Dans leur article, les trois axes sont représentés deux à deux pour former des plans de perception. La figure 5 est un exemple de représentation d'ambiance (sonore, visuelle et olfactive) dans les trois plans [WOLOSZYN-98].

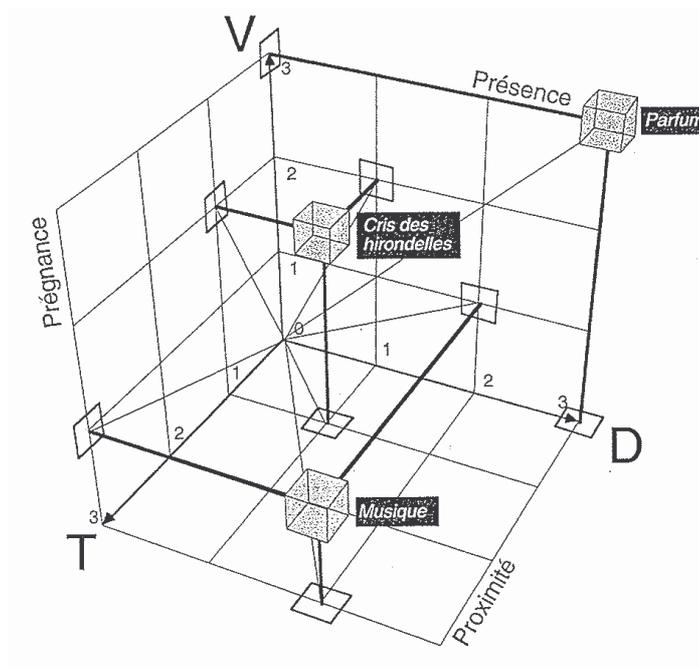


Figure 5 : Repère d'ambiance et mesure des objets ambiants à la terrasse d'un café un après-midi de printemps, selon les trois axes de l'espace perceptif (V pour valeur, D pour dimension et T pour temporalité) [WOLOSZYN-98].

Le plan de la prénance est formé par la réunion des axes de valeur (V) et de temps (T). Il exprime la « densité événementielle » de l'objet.

Le plan dimension(D)-valeur(V) mesure quant à lui la présence. C'est en quelque sorte le « remplissage » de l'espace subjectif par l'objet.

La proximité réunit les axes de dimension (D) et de temporalité (T). Nous pouvons la définir comme la « capacité de l'objet à atteindre le perceuteur dans le temps et dans l'espace ».

Or si cela peut se comprendre pour un parfum dont la « proximité » est forte et qui envahit l'espace, cela se conçoit mal pour une source sonore qui, si elle est proche est facilement localisée et l'auditeur a peu de chances d'être entouré par le son.

Par esprit de simplification, et pour mieux correspondre aux « objets sonores », lors des tests auditifs, les dimensions de prégnance, présence et proximité ont été présentées comme suit :

- **La prégnance** : c'est l'importance, dans le sens où une source sonore s'est démarquée des autres. Non pas à cause de sa durée, mais sur le fait qu'elle attire l'attention lorsqu'elle apparaît.
- **La présence** : c'est l'occurrence temporelle de la source. Cela revient à se demander si la source apparaît souvent durant la séquence.
- **La proximité** : c'est la distance par rapport à la source.

Ainsi, un objet sonore dans son environnement (par exemple un bus qui passe) sera évalué par sa prégnance, sa présence et sa proximité. On retrouve les dimensions des sons (énergétique, temporelle et spatiale) mais on ne retrouve pas la dimension spectrale qui est intrinsèque à la nature de la source et est ainsi reportée à travers l'identification de l'objet même. Les objets sonores que nous étudions et qui ont pour propriété de ressortir, a priori, du bruit de fond, sont ceux généralement rencontrés en milieu urbain : les voitures, les motos, les autobus ou camions, les cyclomoteurs, les voix d'enfants, les voix d'adultes et les chants d'oiseaux.

CHAPITRE 2. CARACTERISATION SUBJECTIVE DE L'ENVIRONNEMENT SONORE URBAIN

L'objectif de cette première partie est de caractériser la perception de l'environnement sonore urbain. L'originalité de ce travail est de prendre en compte l'identification des sources sonores qui composent l'environnement urbain. Des tests en laboratoire construits à partir de séquences enregistrées en ville sont donc mis en place. L'analyse des données collectées ainsi que le croisement avec des mesures physiques vont permettre de caractériser de façon subjective et objective la qualité sonore urbaine.

Au total, 6 tests différents sont réalisés. Ils sont numérotés dans le tableau ci-dessous (Tableau 1).

		Caractéristiques demandées		
		Prégnance, présence et proximité des sources sonores	Intensité subjective	Désagrément sonore
Séquence sonore présentée	Vingt stimuli de 15 s	1	2	3
	Recomposition de 5 min	4	5	6

Tableau 1 : Numérotation des tests réalisés en fonction de leurs caractéristiques.

Les six tests sont répartis sur trois expériences différentes.

La première expérience porte sur l'analyse de la prégnance, la présence et la proximité des sources sonores et tourne autour du test n° 1.

La seconde expérience porte sur l'extraction de modèles de désagrément sonore. Elle utilise les résultats des expériences n° 1, 2 et 3.

Enfin, la troisième et dernière expérience tente d'établir un pont entre les jugements portant sur les stimuli (tests n° 1, 2 et 3) et les jugements portant sur la séquence globale (test n° 4, 5, 6). En conséquence, elle s'appuie sur les résultats de l'ensemble des tests réalisés.

I. Méthode expérimentale

1.1. Les enregistrements

Les enregistrements ont été réalisés dans deux rues parisiennes. Trois séquences d'enregistrements de quinze minutes ont été effectuées. Elles nous ont permis d'extraire vingt séquences de 15 s chacune. Nous avons évité au possible de « couper » les passages de véhicules au début ou à la fin des stimuli afin d'obtenir une certaine fluidité. Les rues dans lesquelles ont été réalisés les enregistrements sont la rue de la Glacière et la rue Reille (treizième et quatorzième arrondissement de Paris). Elles sont représentatives de rues moyennement circulées comme il en existe beaucoup à Paris. Elles ont sensiblement la même configuration qui est la suivante : rues en U à simple (Rue de la Glacière) ou double sens (rue Reille), des pentes presque nulles, un même type de revêtement routier (asphalte) avec une granulométrie variable, des vitesses de véhicules tournant aux alentours de 50 km/h, des flots irréguliers de véhicules, c'est-à-dire avec des périodes de calme relatif entre les passages, faisant apparaître de manière plus distincte les chants d'oiseaux ou les voix.

Deux types d'enregistrements ont été réalisés simultanément : un pour les mesures, l'autre pour les tests auditifs.

Pour les tests auditifs, les prises des sons ont été effectuées avec une configuration ORTF – AB dite « couple AB » et munie d'une paire de microphones SHOEPS. L'écartement entre les microphones est de 17 cm et l'angle de leurs axes est de 110° comme indiqué sur la Figure 6. Ces valeurs sont reconnues pour obtenir la solution la plus acceptable entre la précision des sources partielles et leur répartition devant l'auditeur [CONDAMINES-78]. L'enregistrement s'est fait sur un DAT avec une fréquence d'échantillonnage de 44 100 Hz. Les données ont été transférées sur un ordinateur via une carte son de bonne qualité (Creative Extigy). Les vingt séquences de 15 s ont été découpées en utilisant le logiciel « Wave Lab » fourni avec la carte son.

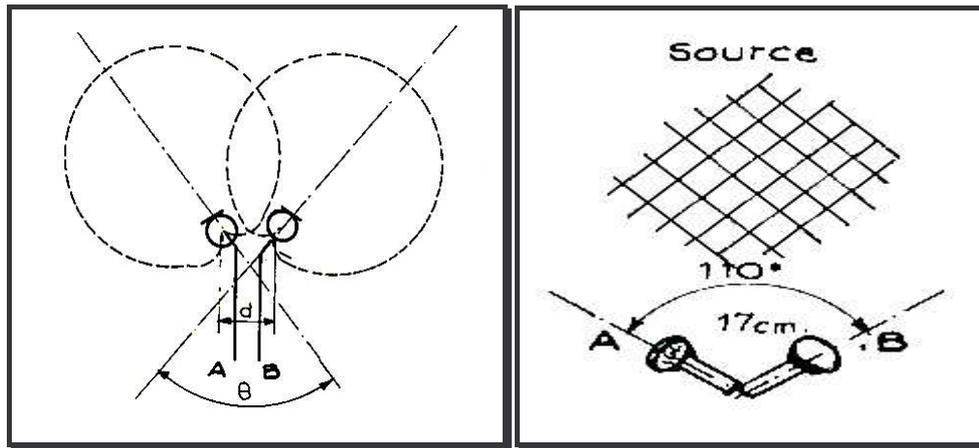


Figure 6 : Illustration de la configuration ORTF – AB. L'écartement d entre les microphones est de 17 cm, l'angle θ de leurs axes est de 110° . Illustration [CONDAMINES-78]

Pour les mesures, le système d'acquisition doit avoir les mêmes caractéristiques que ceux habituellement employés par les bureaux d'études acoustiques, les bureaux de contrôles, les collectivités...En effet, ce système doit enregistrer l'ensemble du signal sonore parvenant au micro. C'est donc une prise mono et omnidirectionnelle qui permet par la suite de faire des analyses physiques objectives de l'environnement sonore. Le système employé est la carte son externe SYMPHONIE de 01dB avec un micro Bruel & Kjaer de classe 1 reliée à un ordinateur portable. La fréquence d'échantillonnage est de 52 000 Hz, la bande passante s'étend de 20 à 20 000 Hz. Le protocole de mesure qui a été respecté lors de ces enregistrements est conforme à la norme 31-010 [NORME 31-010-96] intitulée « Caractérisation et mesurage des bruits de l'environnement ». Les principales spécifications sont les suivantes : le microphone doit être localisé à au moins deux mètres de la partie la plus avancée des bâtiments, entre 1.2 et 1.5 m au-dessus du sol, et à une distance comprise entre 2 et 3 m des véhicules.

1.2. Les stimuli sonores

1.2.1. Vingt séquences de 15 s

Les stimuli sonores présentés aux auditeurs doivent remplir plusieurs fonctions. Ils doivent être représentatifs d'ambiances sonores de type « événementielles », c'est-à-dire que toute source sonore doit être perçue distinctement. Les sources sonores les plus généralement citées en environnement urbain sont les voitures, les cyclomoteurs, les motos, les autobus, les voix d'adultes et les voix d'enfant [LEOBON-95]. Elles apparaissent toutes plusieurs fois dans les vingt stimuli. Leur répartition dans les stimuli se fait selon leur nombre pour les véhicules et selon leur apparition pour les voix et les oiseaux. Le tableau 2 résume les occurrences pour l'ensemble des stimuli. Les événements présents dans les stimuli doivent également être les plus exhaustifs

possibles. Ainsi nous avons des stimuli avec zéro, une et deux cyclomoteurs. Idem pour les bus. La plus grande variabilité en nombre est logiquement obtenue avec les voitures. En effet, nous allons de zéro à cinq voitures en passant par tous les nombres intermédiaires.

	Voiture	Cyclomoteur	Moto	Bus	Voix adulte	Voix enfant	Oiseaux
Stimulus 1	4				X	X	X
Stimulus 2	2	2			X	X	X
Stimulus 3	1		1	1			
Stimulus 4	5						
Stimulus 5	1	1					
Stimulus 6		1		1	X		
Stimulus 7		2					
Stimulus 8	3	1	1				
Stimulus 9	1	1					X
Stimulus 10				1			X
Stimulus 11	4	1			X	X	X
Stimulus 12	1						X
Stimulus 13	3						X
Stimulus 14		1	1		X		X
Stimulus 15	4					X	
Stimulus 16	1				X	X	X
Stimulus 17	2	2			X	X	X
Stimulus 18	2			1			
Stimulus 19				2	X		
Stimulus 20			1				X

Tableau 2 : Occurrences des sources sonores pour l'ensemble des stimuli.

1.2.2. Séquence de 5 minutes

Les vingt séquences décrites ci-dessus sont regroupées dans une seule séquence continue de 5 min. L'effet utilisé pour que les vingt séquences regroupées donnent l'impression d'une seule unité de temps et de lieux est le « fondu-enchaîné (ou « cross fading » en anglais), réalisé par le logiciel PRO TOOLS [Figure 7]. A partir des enregistrements des trois séquences de quinze minutes (paragraphe 1.1), des stimuli de dix-sept secondes ont été extraits. Ils correspondent à chacune des vingt séquences de quinze secondes mais avec une seconde supplémentaire de chaque côté. Puis, les deux signaux sont traités de la façon suivante : le niveau du signal de départ décroît sur deux secondes alors que le son enchaîné croît sur une même période et selon une même pente (cependant inversée). Les deux signaux sont ensuite mixés pour que le début et la fin des séquences de départ de quinze secondes coïncident. Nous obtenons à la fin une séquence de 5 min exactement, soit vingt fois 15 s.

L'impression de continuité de la prise de son de la séquence est ainsi totalement recréée.

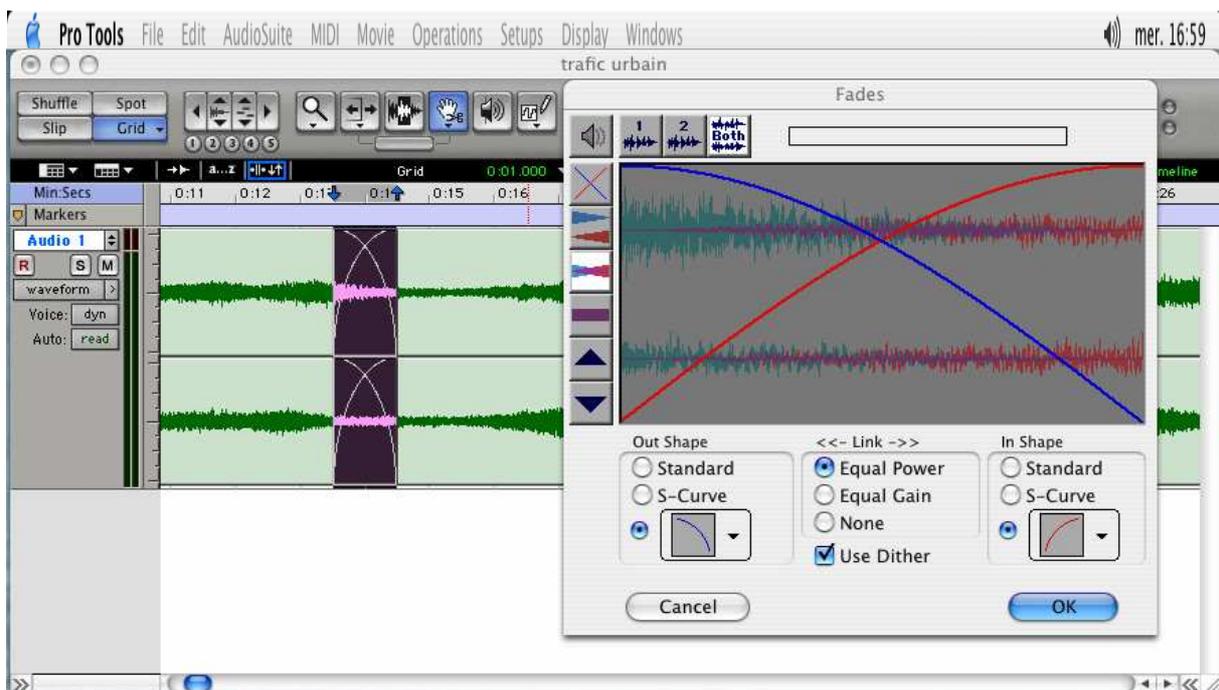


Figure 7 : Illustration du cross fading réalisé avec Pro Tools. Au premier plan nous voyons en détails les deux signaux (en stéréo) qui sont mixés sur une durée de 2 s, les courbes bleues et rouges représentent les atténuations d'intensité pour les deux signaux. En arrière plan, nous voyons les signaux une fois le cross fading réalisé.

1.3. Les sujets des tests

Chaque expérience psychoacoustique est conduite avec un panel de vingt à trente personnes. La plupart étant acousticiens ou ingénieurs du son, leur âge varie de vingt-deux à trente-neuf ans et la répartition entre hommes et femmes est sensiblement la même. Le point important est qu'ils sont tous résidents de Paris ou des environs, et ce depuis plusieurs années. On peut donc supposer que leur environnement sonore quotidien est à peu près identique et également qu'il diffère peu des séquences qui leur sont présentées. De plus, chacun des auditeurs confirme oralement qu'il n'a pas de sensibilité anormalement élevée vis-à-vis des sources sonores présentes dans les stimuli. Les acousticiens et ingénieurs passent régulièrement des audiogrammes pour leur métier. Aucun n'a de problème d'audition.

1.4. Calculs des critères physiques

A partir de l'enregistrement de chaque stimulus de 15 s, le niveau de pression équivalent L_{Aeq} est calculé toutes les 200 ms. Grâce à l'écoute des stimuli, un codage manuel des sources par catégorie est réalisé sur cette courbe de $L_{Aeq,200ms}$. La figure 8 montre le codage temporel effectué pour les sources de type véhicule. Les sources s'entendent distinctement et ainsi il n'y a pas de « recouvrement » dans le codage des sources. Pour les voix et les oiseaux [Figures 9 et 10], les codages sont effectués sur des fenêtres séparées car ces sons se superposent avec les bruits de véhicules.

Par la suite, il est possible de calculer, pour chaque catégorie ainsi codée dans le temps, des descripteurs particuliers. Par exemple, des descripteurs relatifs au niveau d'une source (le niveau de pression équivalent pondéré A particulier L_{Aeq} , les niveaux percentiles L_{A10} et L_{A5} , le niveau de pression maximum pondéré A L_{AMax}), mais également des descripteurs relatifs au temps comme le pourcentage de temps de présence (%T) d'une catégorie sur la durée totale du stimulus ainsi que le nombre d'événements (le nombre de voitures par exemple). Le codage des sources et les calculs ont été effectués grâce au logiciel dBTrait32.

D'autres indicateurs acoustiques sont calculés, mais cette fois-ci sur la durée totale des stimuli et non plus spécifiquement sur telle ou telle source. Ces indicateurs globaux sont le L_{Aeq} , le L_{A10} , le L_{A5} , le L_{AMax} , la sonie moyenne calculée sur le modèle de Zwicker, la sonie dépassée pendant 10% du temps, l'acuité, la rugosité, et la force de fluctuation.

Ces critères dits objectifs servent de relais pour tenter d'expliquer à l'aide de valeurs physiques, les valeurs subjectives obtenues pendant les tests sur les trois dimensions de prégnance, de présence et de proximité des sources, ainsi que les dimensions de force sonore et de désagrément.

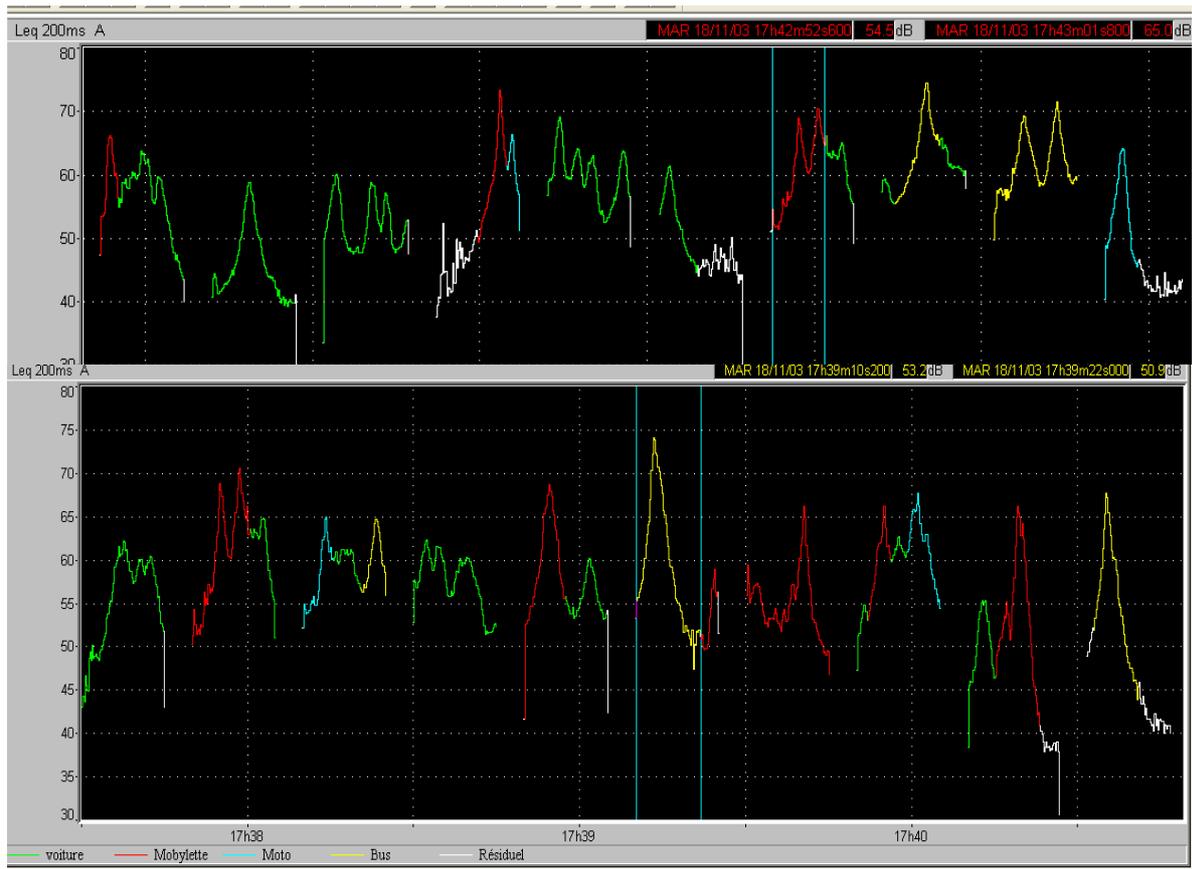


Figure 8 : Codage des sources sonores de type véhicules sur les courbes d'évolution du $L_{Aeq,200ms}$ des vingt stimuli du test n° 1. Chaque couleur correspond à une catégorie de source. Le codage est réalisé grâce à l'écoute des stimuli et à l'aide du logiciel dBTrait32. En haut, les stimuli allant des numéros « 11 » à « 20 », en bas, les stimuli allant des numéros « 1 » à « 10 ».

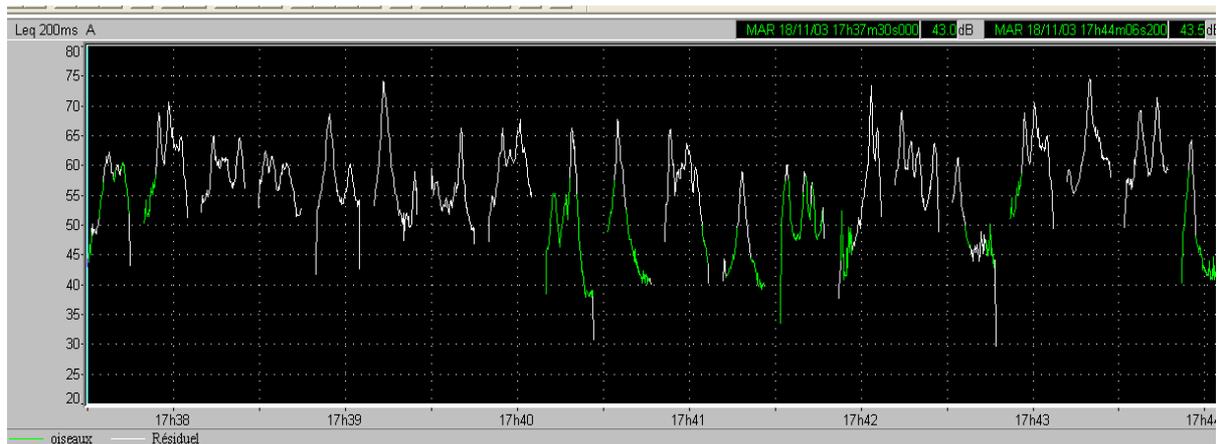


Figure 9 : Codage des oiseaux sur les courbes d'évolution du $L_{Aeq,200ms}$ des vingt stimuli du test n° 1. Le codage est réalisé grâce à l'écoute des stimuli et à l'aide du logiciel dBTrait32.

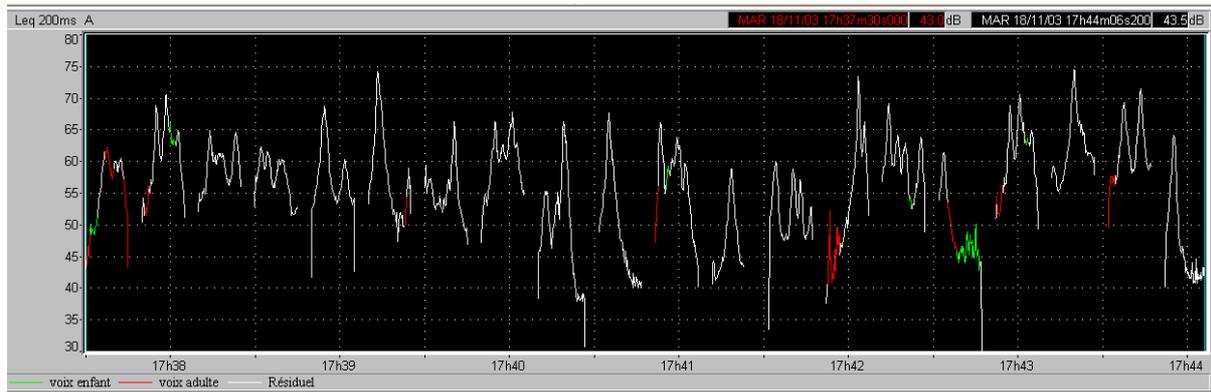


Figure 10 : Codage des voix d’adultes et voix d’enfants sur les courbes d’évolution du $L_{\text{Aeq},200\text{ms}}$ des vingt stimuli du test n° 1. Le codage est réalisé grâce à l’écoute des stimuli et à l’aide du logiciel dBTrait32.

II. Caractérisation des séquences de 15 secondes

2.1. Test n° 1 : Prégnance, présence et proximité des sources sonores

2.1.1. Questionnaire et procédure

Cette expérience s’appuie sur le test suivant : l’ensemble des stimuli est présenté dans un ordre aléatoire et différent pour chaque auditeur à travers un casque ouvert *SEINHEISER HD 600*. Les sujets doivent répondre aux questions qui leur sont présentées par l’intermédiaire d’une interface dessinée sur un écran d’ordinateur [Figure 11]. Cette interface est réalisée à l’aide du logiciel MATLAB. Ce logiciel gère la diffusion des stimuli sonores quand l’auditeur fait appel à la touche « Lecture » en cliquant dessus. Chaque stimulus peut être écouté plusieurs fois. L’auditeur peut ainsi rentrer des estimations chiffrées sur la prégnance, la présence et la proximité, en face des catégories de sources sonores qui lui sont indiquées. Une fois qu’il a effectué cette tâche, il lui suffit de cliquer sur la touche « OK » pour que le stimulus suivant se fasse entendre et que ses notes soient validées. A la fin du test, après l’écoute des vingt stimuli, l’ensemble des données d’acquisition (nom et prénom de la personne, date du test et notes subjectives) sont compilées dans un fichier.

L’interface permet également, à travers un onglet appelé « paramètre », de produire un son pur de référence d’une durée de 10 s qui permettra de calibrer le système afin que les niveaux sonores des stimuli soit les mêmes pour chaque sujet. Enfin, il est possible de choisir, toujours dans l’onglet « paramètre », la diffusion de stimuli qui servent d’exemples pour illustrer les bornes

extrêmes des adjectifs. La figure 11 représente l'interface de ce test telle que présentée aux auditeurs.

Parametre

TEST PERCEPTIF - BD 2003

Votre Nom:

Type de source sonore :

Note de prégnance :

Note de présence :

Note de proximité :

Voiture

Autobus-Camion

Moto

Mobylette

Voix adulte

Voix enfant

Oiseaux

Lecture

OK

Reset

La source prend-elle plus ou moins de place?

La source est-elle plus ou moins présente ?

La source est-elle plus ou moins proche?
10 = très proche
1 = très lointaine

5 4/20

Figure 11 : Interface du test de prégnance, de présence et de proximité des sources sonores pour les vingt stimuli. Cette interface est réalisée avec le logiciel MATLAB.

Une feuille d'explication est fournie aux auditeurs [Annexe 1]. En plus de lui fournir les modalités de déroulement du test, cette feuille précise que pour chaque stimulus entendu, les sujets doivent mettre une note de présence, de prégnance et de proximité en rapport avec chaque source présente. Les trois adjectifs ont été décrits de la manière suivante (comme présenté dans le chapitre 1) :

- **La prégnance** : c'est l'importance, dans le sens où une source sonore s'est démarquée des autres. Non pas à cause de sa durée, mais sur le fait qu'elle ait attiré votre attention lorsqu'elle est apparue.
- **La présence** : c'est l'occurrence temporelle de la source. Cela revient à se demander si la source apparaît souvent durant la séquence
- **La proximité** : c'est votre distance par rapport à la source (attention une petite note indique une source éloignée).

Les notes proposées sont des entiers allant de zéro (pour une absence de source sonore) à dix. Pour une meilleure appréciation de l'échelle et de la signification des adjectifs, des stimuli « extrêmes » sont présentés aux sujets avant le début du test. Chaque stimulus est entendu avec le même niveau que le niveau réel in-situ. De plus, il est filtré par la fonction de transfert inverse du casque. Ceci afin de s'affranchir des caractéristiques de la réponse fréquentielle du casque [Figure 12].

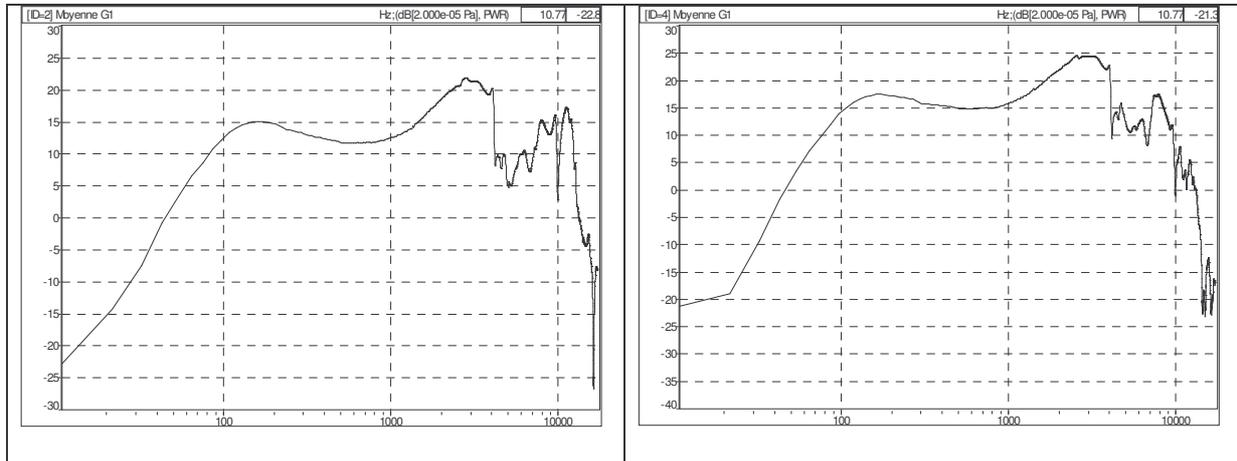


Figure 12 : Fonctions de transferts du casque HD 600 pour l'oreille gauche et l'oreille droite. Les réponses impulsionnelles ont été enregistrées à travers une tête artificielle (Kemar). Les fonctions de transferts sont obtenues par la transformée de Fourier des réponses impulsionnelles (fenêtre rectangulaire, de 0 à 20000 Hz sur mille six cent lignes, ce qui donne une résolution de 12.5 Hz) à partir du logiciel dBFa32.

Comme déjà précisé, les sujets ont la possibilité de réécouter les stimuli mais la consigne est de répondre de façon la plus spontanée possible. Ainsi, les sujets n'ont jamais réécouté plus de trois fois le même stimulus. La durée totale du test a varié de 30 à 45 min. Le contexte a été recréé en demandant au sujet de s'imaginer un déplacement urbain pour donner un même cadre d'action à tous les sujets. Par exemple ils ont un rendez-vous mais ils ne sont pas pressés. En conséquence, comme les sujets sont probablement dans le même état d'esprit, nous pouvons supposer que les différences de jugements dus au contexte « d'action » sont minimisées.

2.1.2. Normalisation des données, moyennes et corrélations

a). Normalisation

Bien que les réponses des sujets soient notées sur les mêmes échelles, elles sont tout de même normalisées de manière à rendre constante la variance des réponses de chaque personne. Cette normalisation a pour effet de modifier légèrement les notes des sujets (soit en étirant, soit en réduisant leur dynamique) en tenant compte de l'ensemble des réponses. Ainsi, si un sujet a mis des notes plus basses que la tendance générale, elles seront légèrement augmentées et inversement.

La normalisation est la suivante :

$$X_{norm,i,j,p} = X_{i,j,p} \sqrt{\frac{\sum_{i,j,p} X_{i,j,p}^2}{\sum_{i,j} X_{i,j,p}^2}}$$

Où

$p = 1$ à P personnes

$i =$ stimuli 1 à 20,

$j =$ questions 1 à 21 (trois questions pour sept sources différentes),

$X_{i,j,p}$ = réponse initiale de la personne p pour le stimulus i et la question j ,

$X_{norm,i,j,p}$ = réponse normalisée de la personne p pour le stimulus i et la question j ,

$\sum_{i,j} X_{i,j,p}^2$ = somme des carrés de toutes les réponses de la personne p ,

$\overline{\sum_{i,j,p} X_{i,j,p}^2} = (\sum_{p=1}^P \sum_{i,j} X_{i,j,p}^2) / P$ = moyenne de la somme des carrés de tous les sujets.

La normalisation effectuée sur les réponses des sujets n'a eu que très peu d'effet sur les notes initiales. En effet, les sujets ont globalement bien utilisé la totalité de la dynamique de l'échelle proposée.

b) Moyennes des réponses

Avant de poursuivre par des analyses factorielles construites sur des moyennes, il faut vérifier que celles-ci sont interprétables.

Pour cela, on étudie l'histogramme des réponses des sujets sur un critère. Si cet histogramme a une forme gaussienne, cela signifie que les notes attribuées sont dispersées autour de la valeur moyenne indiquant ainsi un « consensus » sur la signification et l'évaluation de l'attribut par l'ensemble des sujets.

Quand une source est détectée, la majorité des histogrammes des valeurs subjectives a sensiblement une allure de gaussienne mais différents gabarits peuvent également apparaître. La répartition des effectifs peut dans certains cas être concentrée en grande majorité sur deux ou trois notes uniquement qui sont voisines, sans toutefois prendre une allure de gaussienne (Figure 13a).

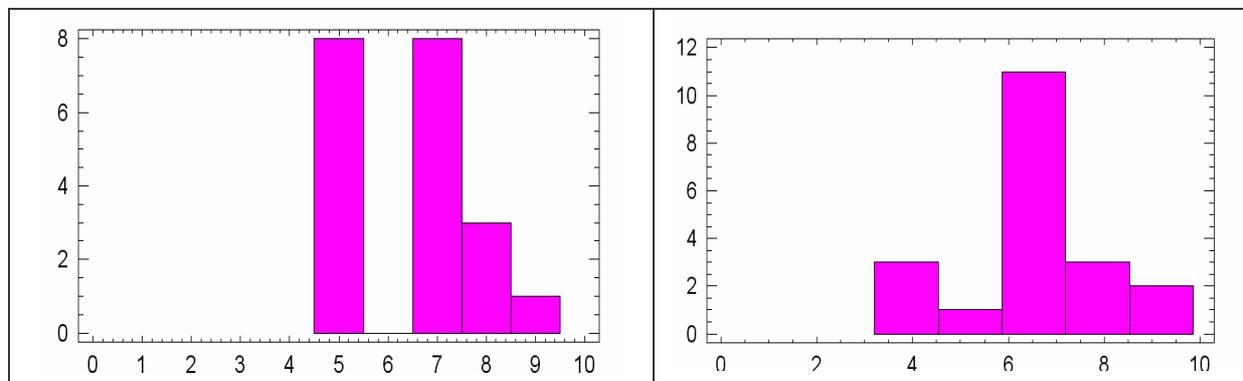


Figure 13a : Histogramme des réponses illustrant une répartition sans allure franche de gaussienne mais ayant une cohérence dans l'étalement des notes. Présence des cyclomoteurs du stimulus « 14 » et proximité des oiseaux du stimulus « 20 ».

La répartition des effectifs peut également s'étaler sur un nombre parfois important de notes différentes mais en prenant cette fois-ci une allure de gaussienne (Figure 13b).

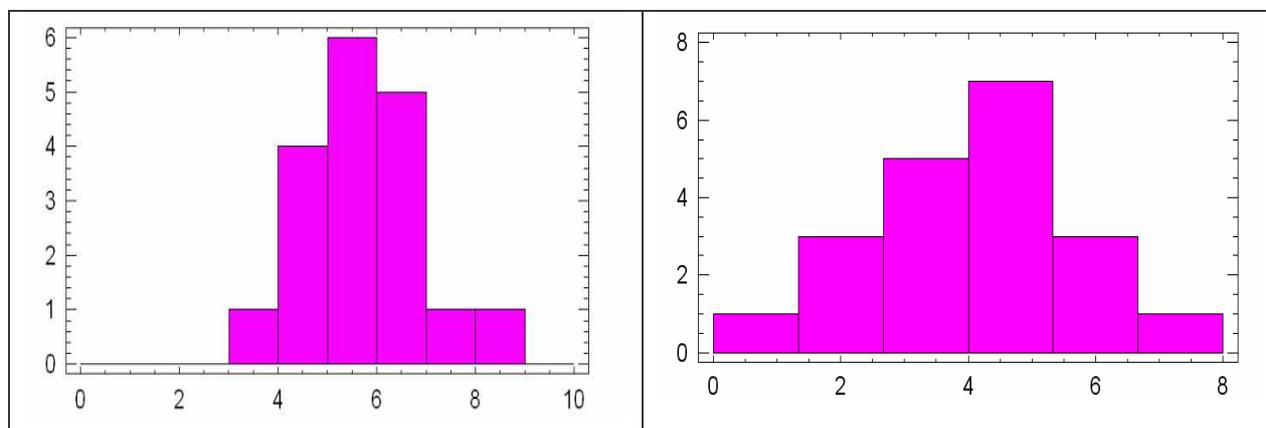


Figure 13b : Histogramme des réponses illustrant une répartition avec une allure de gaussienne mais ayant un nombre important de notes différentes. Prénance des voitures pour le stimulus « 3 » et présence des voix d'enfants pour le stimulus « 16 ».

Enfin, lorsqu'une source en particulier n'est pas présente dans le stimulus, la majorité des auditeurs attribue la note « 0 ». Néanmoins, il peut y avoir quelques erreurs sans conséquences importantes sur le calcul de la moyenne, comme nous pouvons le voir sur la figure 13c.

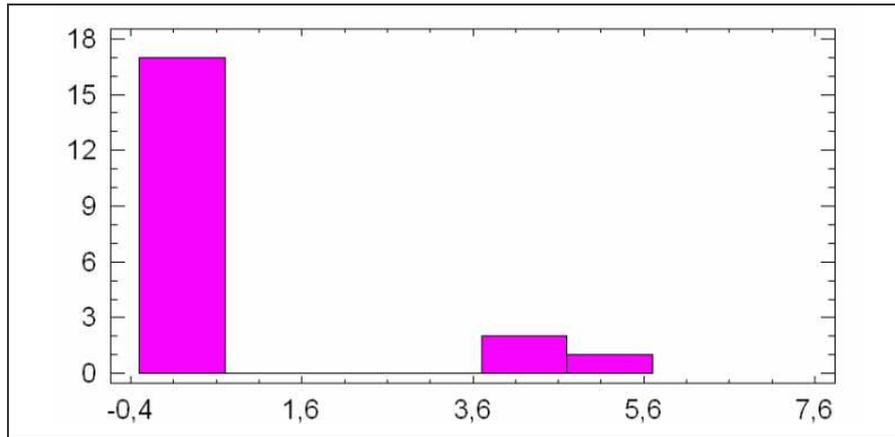


Figure 13c : Histogramme des réponses pour la prégnance des voitures du stimulus « 19 ». Il n’y a pas de voiture dans ce stimulus, cependant les trois erreurs n’ont que très peu d’influence sur la valeur de la moyenne

Ces différentes allures d’histogrammes ont toutefois très peu d’influence sur le résultat de la note moyenne qui est de ce fait interprétable.

Il y a cependant des histogrammes qui révèlent un manque de consensus dans les réponses des sujets. Par exemple, la figure 13d (prégnance des motos sur le stimulus numéro 3) illustre le fait que certains auditeurs ont entendu des motos et d’autres pas.

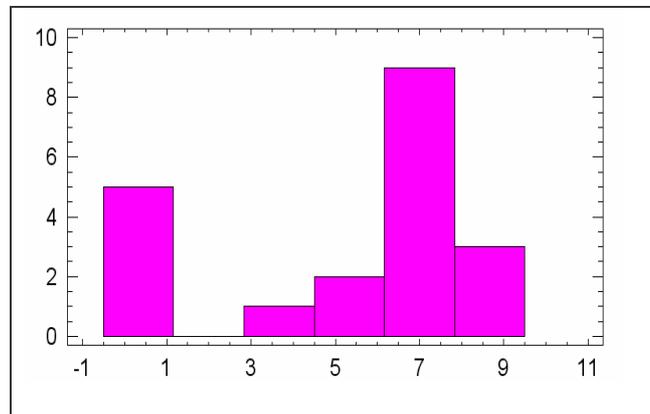


Figure 13d : Histogramme des effectifs illustrant le fait que deux groupes de sujets ont répondu de manière inverse sur la présence de moto du stimulus « 3 ».

Lors de confusions de ce type (il y a deux cas qui concernent l’identification des motos), nous avons conservé les réponses des sujets. Nous verrons par la suite que l’utilisation de l’analyse factorielle des correspondances, qui permet d’analyser les consensus des réponses, nous révèle si des notions ont été interprétées différemment ou si l’identité de certaines sources sonores a été la source de confusions.

Les résultats de l'ensemble des données du test n° 1 avec écart type des moyennes sont présentés sous forme de graphique en Annexe 2.

c) Corrélations

En considérant l'ensemble des sources sonores, les trois adjectifs (prégnance, présence et proximité) sont en général assez corrélés. Le tableau 3 indique la valeur de ces corrélations.

	<i>Prég</i>	<i>Prés</i>	<i>Prox</i>	<i>Prég</i>	<i>Prés</i>	<i>Prox</i>
	Voiture			Bus		
Prégnance		0.50	0.71**		0.69	0.88*
Présence	0.50		0.78**	0.69		0.34
Proximité	0.71**	0.78**		0.88*	0.34	
	Moto			Cyclomoteur		
Prégnance		0.90	0.92		0.87**	0.96**
Présence	0.90		0.99*	0.87**		0.81**
Proximité	0.92	0.99*		0.96**	0.81**	
	Oiseaux			Voix d'adulte		
Prégnance		0.90**	0.96**		0.90**	0.96**
Présence	0.90**		0.91**	0.90**		0.90**
Proximité	0.96**	0.91**		0.96**	0.90**	
	Voix d'enfant					
Prégnance		0.97**	0.99**			
Présence	0.97**		0.96**			
Proximité	0.99**	0.96**				

Tableau 3 : Coefficients de corrélation entre la prégnance, la présence et la proximité de toutes les sources sonores. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

Les corrélations ont été calculées uniquement sur les stimuli comprenant la source en question, évitant ainsi que les notes « 0 » (signifiant une absence de source) n'entrent dans les calculs. Cette précaution permet de ne prendre en compte que les corrélations interprétables. Si les « 0 » étaient pris en compte, les corrélations auraient été artificiellement très bonnes. La figure 14 montre une corrélation avec et sans prise en compte des « 0 ».

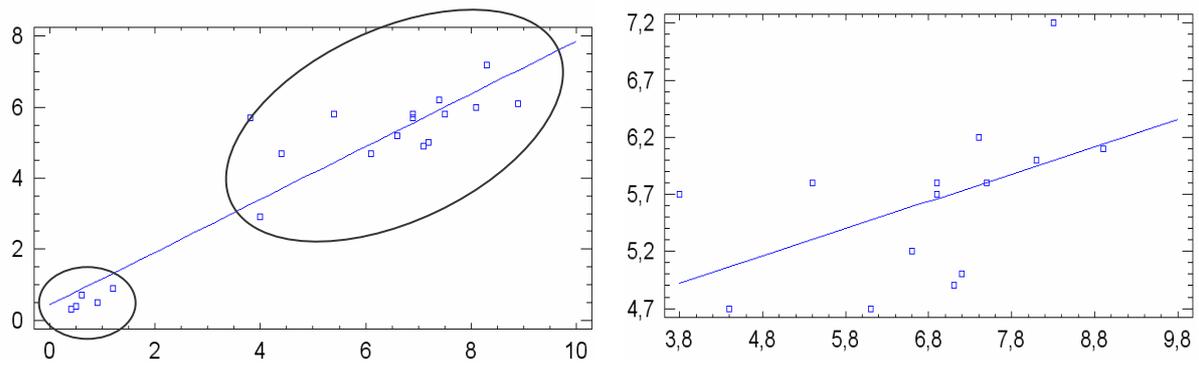


Figure 14 : Graphiques des corrélations entre la présence et la prégnance des voitures. A gauche sont représentés l'ensemble des vingt stimuli formant deux « paquets » ($r = 0.93$), à droite les quatorze stimuli comprenant des voitures ($r = 0.50$).

Les bonnes corrélations du tableau 3 s'expliquent essentiellement par la construction des stimuli plutôt que par une mauvaise distinction entre les trois adjectifs. Bien que les trois adjectifs fassent référence à des notions distinctes (l'importance pour la prégnance, la durée pour la présence et la distance pour la proximité), celles-ci ne sont pas indépendantes si l'on considère le cas de sources sonores en environnement urbain. En effet, les sources, tant les véhicules que les voix, ont la particularité de se mouvoir par des mouvements continus et réguliers (la source apparaît au loin, s'approche et puis disparaît comme elle est apparue) liant en conséquence la prégnance, la présence et la proximité de leur signature sonore. Les oiseaux par contre sont considérés comme étant plus statiques mais la diversité de positions et d'intensités de ces derniers dans chacun des stimuli n'est pas assez élevée pour que les notions demandées soient faiblement corrélées. De plus, les stimuli ont été enregistrés en deux points fixes, ainsi, les véhicules ont tendance à passer à peu près à la même vitesse et à la même distance du micro ce qui ne permet pas une grande variabilité dans les trois dimensions.

C'est naturellement lorsque les variabilités sont les plus importantes (le nombre de voitures dans l'ensemble des stimuli varie de « 0 » à « 5 », le nombre de bus et de cyclomoteurs varie de « 0 » à « 2 ») que les corrélations sont les plus faibles. Les notions de présence, prégnance et proximité ont donc bien été distinguées par les sujets.

En conclusion, les importantes corrélations entre les adjectifs montrent qu'il n'existe pas, dans une même catégorie de sources, une grande variabilité de cas ou de trajectoires sonores. C'est donc entre catégories de sources que des disparités de jugements peuvent apparaître.

2.1.3. Analyses factorielles

Les réponses du test n° 1 et l'analyse des données physiques des vingt stimuli nous permet d'obtenir un tableau comprenant les données suivantes par stimuli :

a) Description des valeurs

Valeurs subjectives : Prénance Voiture (PgVoit), Présence Voiture (PsVoit), Proximité Voiture (PxVoit), Prénance Cyclomoteur (PgMob), Présence Cyclomoteur (PsMob), Proximité Cyclomoteur (PxMob), Prénance Bus (PgBus), Présence Bus (PsBus), Proximité Bus (PxBus), Prénance Moto (PgMot), Présence Moto (PsMot), Proximité Moto (PxMot), Prénance Voix Adulte (PgVA), Présence Adulte (PsVA), Proximité Adulte (PxVA), Prénance Voix Enfant (PgVE), Présence Enfant (PsVE), Proximité Enfant (PxVE), Prénance Oiseaux (PgOis), Présence Oiseaux (PsOis), Proximité Oiseaux (PxOis).

Valeurs objectives : Les valeurs relatives à l'intensité sont calculées sur les courbes de L_{Aeq} codées par catégorie de source : L_{Aeq} , L_{A10} , L_{A5} , L_{AMax} . Par exemple pour un stimulus nous pouvons calculer le L_{A10} relatif à l'ensemble des voitures présentes dans ce stimulus. Il sera noté dans les graphiques $L_{A10}Voit$.

Nous calculons également le pourcentage de temps d'apparition (%T) par type de source sur la durée totale d'un stimulus. Par exemple, le pourcentage de temps d'apparition des oiseaux se notera dans les graphiques %TOis.

Enfin, nous calculons le nombre (Nb) de véhicules de chaque catégorie pour chaque stimulus. Ainsi, le nombre de cyclomoteurs sera par exemple noté NbMob.

Lorsque cela est utile à l'interprétation des espaces perceptifs, des valeurs globales du L_{Aeq} sont calculées et sont notées GL_{Aeq} pour la valeur du L_{Aeq} sur les 15 secondes, $GL10$ pour le L_{Aeq} dépassé pendant 10% du temps, et $GL90$ pour le L_{Aeq} dépassé pendant 90% du temps.

b) Descriptions des analyses

Les représentations géométriques des données des analyses factorielles transforment en distances euclidiennes les propriétés statistiques entre éléments du tableau de données subjectives obtenues lors du test n° 1. En réduisant l'information des tableaux en un petit nombre de variables classées par rapport à de nouveaux axes (axes principaux d'inertie), les analyses factorielles représentent des outils permettant de visualiser des regroupements, des oppositions et des tendances, impossibles à distinguer dans un large tableau de nombres.

Nous avons en effet un nombre n de variables très important (présence, prénance et proximité pour chaque source sonore et pour chaque stimulus) donc un espace à n dimensions.

Les analyses factorielles vont réduire cet espace en un nombre de dimensions (appelées également axes principaux d'inertie ou axes factoriel ou encore composantes principales) indépendantes entre elles expliquant un maximum de variabilité des réponses.

De même que dans la thèse de Stéphanie VIOLLON, le critère principal retenu pour trouver le nombre de dimensions à retenir pour l'interprétation des résultats est « le décrochement observable sur les diagrammes d'inertie, tout en vérifiant que le pourcentage d'inertie expliqué des derniers axes ne tombent pas dans des valeurs trop faibles » [VIOLLON-00]. Ce décrochement ou cassure est illustré sur la figure 15. Il se peut néanmoins que nous ayons à étudier la signification d'axes qui n'expliquent qu'une faible partie de la variabilité des résultats, notamment lors des analyses sur l'ensemble des stimuli [Figure 16].

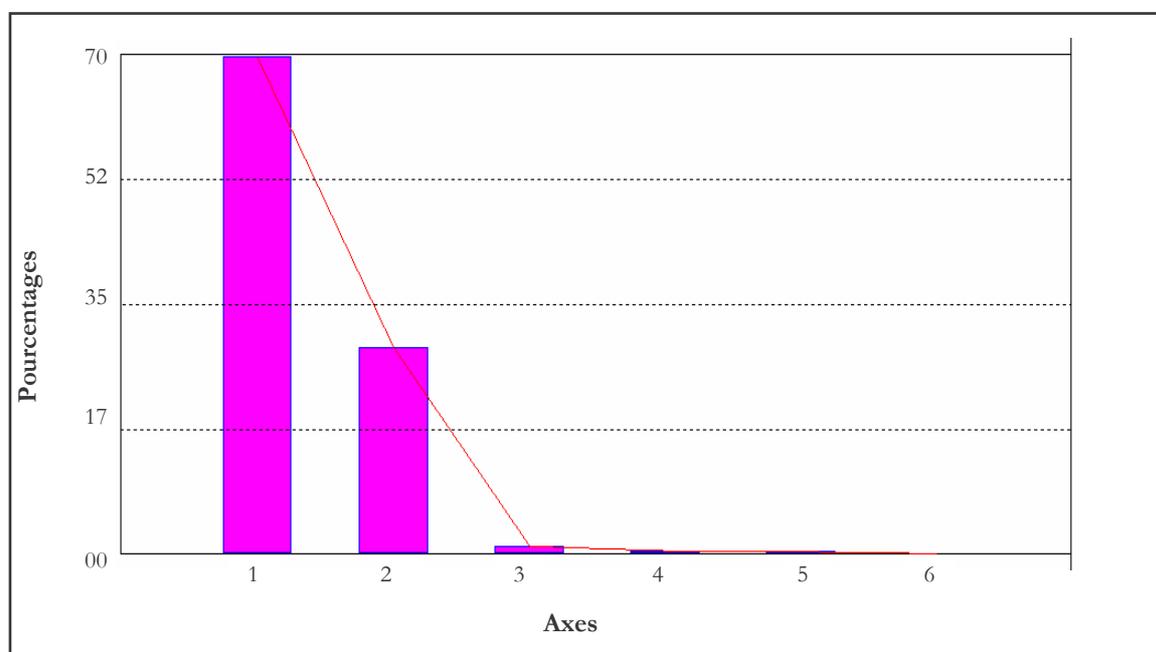


Figure 15 : Représentation du nombre d'axes factoriels illustrant une « cassure » entre le deuxième et troisième axe. Dans ce cas, nous retenons uniquement les deux premières dimensions.

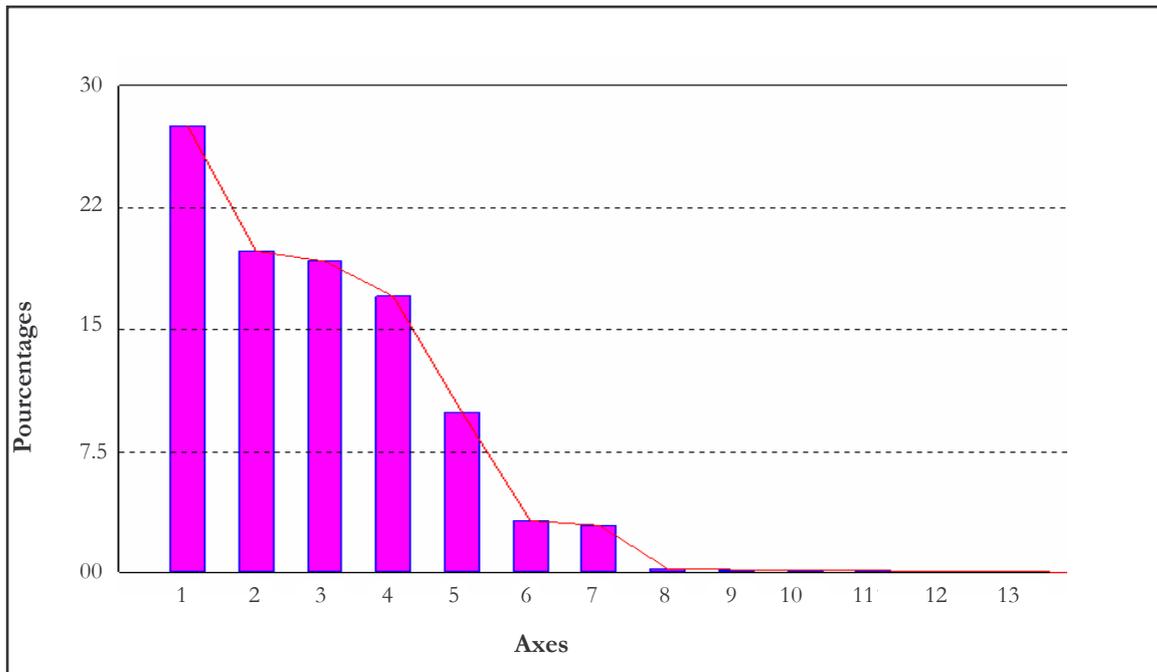


Figure 16 : Représentation du nombre d’axes factoriels illustrant une « cassure » (entre autre) entre le premier et deuxième axe. Dans ce cas, nous retenons les sept premiers axes.

La projection des données physiques en tant que variables supplémentaires sur les axes factoriels nous permet également de visualiser les corrélations entre les dimensions perceptives (prégnance, présence et proximité des sources) et les valeurs objectives facilitant ainsi une association entre les deux espaces.

Deux types d’analyses factorielles sont effectués sur les jugements du test n° 1. Une Analyse en Composantes Principales (ACP) calculées sur les moyennes et une Analyse Factorielle des Correspondances (AFC) calculées sur les tableaux de contingences.

L’ACP permet de croiser des individus (ici les vingt stimuli) à des variables numériques dites variables quantitatives (ici les réponses des sujets sur la prégnance, la présence et la proximité des sources dites variables actives puis dans un second temps les données physiques calculées dites variables passives ou supplémentaires car elles n’entrent pas en compte dans la détermination des axes). Cette analyse multidimensionnelle va nous permettre d’extraire les plans de l’espace subjectif qui sont significatifs en termes de variabilité des réponses des sujets. Sur ces plans significatifs en rapport avec le terme (prégnance, présence ou proximité) et la source, nous pouvons projeter simultanément les variables physiques afin de mieux interpréter la « signification » perceptive de ces dimensions liées à telle ou telle source.

Le défaut que peut présenter une ACP seule réside dans le fait que les données quantitatives qui y sont employées sont des moyennes des réponses des sujets. Ainsi, même si nous avons

vérifié la distribution normale des réponses permettant de faire une moyenne, sans induire trop d'erreurs, il est intéressant de faire ressortir les consensus des réponses des sujets afin d'analyser la pertinence des dimensions trouvées par l'ACP.

L'AFC traite, comme types de données, des tableaux de fréquence. Dans notre cas, si nous construisons des catégories discrètes sur les réponses des sujets, les fréquences représentent le nombre de personnes qui ont répondu dans telle ou telle catégorie. Cette analyse permet alors de croiser ces nouvelles variables qualitatives sur les n individus ou stimuli. Le but étant ainsi d'analyser les liaisons entre ces variables.

Les descriptions théoriques des analyses factorielles faites par la suite dans ce mémoire sont largement inspirées du livre d'Escofier et Pagès intitulé « Analyses factorielles simples et multiples » [ESCOFIER-98].

2.1.3.1. Analyse en Composantes Principales

Afin de mieux appréhender les explications qui vont jaloner ce travail, nous invitons le lecteur à écouter les stimuli sonores qui accompagnent la version informatique (CD-R) du présent mémoire. A défaut d'écoutes, il pourra se référer aux figures 8, 9 et 10 de ce document montrant les courbes de L_{Aeq} des stimuli ainsi que la répartition des sources sonores.

a) Individus et variables

Les individus sont les stimuli contenant les sources sonores mentionnées. Les variables actives sont les réponses des sujets, donc les moyennes des prégnances, des présences et des proximités de chaque source et pour chaque stimulus. Les variables passives sont les mesures physiques, toujours calculées à l'égard des stimuli et des sources sonores.

b) Procédure

Les ACP sont réalisées graduellement. Tout d'abord en se focalisant uniquement sur les catégories de sources prises séparément. Les corrélations entre la prégnance, la présence et la proximité avec les valeurs physiques sont réalisées à cette étape en faisant abstraction des « 0 ». L'utilité de cette analyse est double. Tout d'abord elle nous permet de faire ressortir les sensibilités des auditeurs aux dimensions perceptives à l'égard des sources sonores. De plus, elle indique, à travers des tableaux de corrélations, quelles mesures physiques sont reliées à ces dimensions.

Puis les ACP sont faites en regroupant les catégories au fur et à mesure : par véhicules, par voix et oiseaux et enfin en regroupant toutes les catégories ensembles. Cette deuxième série d'analyses permet de faire ressortir une hiérarchie perceptive des sources sonores.

c) Interprétations

L'ACP, comme toute analyse factorielle, est utilisée dans le but de réduire la dimension des données en conservant le plus d'informations possible. L'information est représentative de la variabilité des réponses des sujets (variables actives) et est représentée à travers des axes dits « principaux ».

Deux types de représentations graphiques nous servent à interpréter les résultats. Pour chaque analyse ils sont exposés de façon détaillée en Annexe 3.

D'un côté, nous pouvons représenter les axes principaux d'inertie (qui deux à deux forment les plans factoriels) sur lesquels les individus sont projetés. Dans ce cas la distance entre deux points s'interprète comme une ressemblance. La répartition des individus le long des axes principaux est interprétée comme étant significative du taux de variabilité des réponses expliquée par les axes. On parle ainsi de principaux facteurs de variabilité dans la mesure où ils rendent compte le plus possible (sous contrainte d'orthogonalité des axes) de la diversité des individus (cette variabilité est bien sûr construite à partir des variables actives).

De l'autre côté, nous représentons les mêmes axes principaux sur lesquels cette fois-ci sont projetées les variables. La représentation est cette fois circulaire (cercle de corrélation) et c'est le cosinus d'un angle entre deux variables qui est significatif. Il s'interprète comme le coefficient de corrélation entre les deux variables. On peut y projeter aussi les variables passives, ici les mesures physiques.

A- ACP pour les sources prises séparément

Cette analyse nous permet de faire ressortir les sensibilités des auditeurs aux dimensions perceptives à l'égard de chaque source sonore. Elles se déduisent de la visualisation des variables sur les cercles de corrélation (les valeurs numériques des corrélations sont également indiquées dans les tableaux de corrélation) ainsi que de la projection des stimuli sur les plans factoriels. De plus, l'analyse indique, à travers les tableaux de corrélations, quelles mesures physiques sont reliées à ces dimensions.

Les représentations graphiques de l'ACP sont exposées en Annexe 3.

a) ACP sur les voitures

Alors que l'adjectif « présence » a été présenté aux auditeurs comme ayant rapport avec le temps d'apparition de la source et que l'adjectif « prégnance » a été présenté comme ayant attrait à l'importance de la source, nous constatons qu'il n'y a pas de correspondances prévisibles entre ces dimensions perceptives et les mesures physiques [Tableau 4].

	<i>Prégnance</i>	<i>Présence</i>	<i>Proximité</i>
Nombre	0.63*	0.84**	0.79**
%T	0.70**	0.54*	0.40
L_{Aeq}	0.27	0.57*	0.66*
L_{Amax}	0.57*	0.61*	0.80**
L_{A5}	0.45	0.56*	0.76**
L_{A10}	0.30	0.48	0.55*
Axe 1	0.82**	0.86**	0.95**
Axe 2	0.54*	-0.45	-0.05

Tableau 4 : Coefficients de corrélation entre la présence, la prégnance, la proximité des voitures et les critères physiques calculés sur les courbes de L_{Aeq} codées voiture, ainsi qu'avec les deux principaux axes de l'analyse en composantes principales. Les résultats avec ** indiquent une probabilité avec un niveau de confiance p<0.01 et les résultats avec * indiquent une probabilité avec un niveau de confiance p<0.05.

En effet, la présence des voitures est plutôt corrélée aux critères relatifs au nombre alors que la prégnance est plutôt corrélée aux critères relatifs au temps [Tableau 4].

De plus, l'analyse en composantes principales révèle deux axes principaux expliquant 95% de la variance des données.

Le premier axe (78.1% de la variance) est très bien corrélé aux jugements de proximité. Cela signifie que c'est la proximité qui nous permet de mieux discriminer les différents stimuli contenant des voitures. La variation du nombre de voitures dans les stimuli se traduit par des variations de niveau maximum et cela est d'abord perçu par les auditeurs comme des variations de distance perceptive. Cette perception est d'ordre « spatial ». On peut citer ici une étude réalisée pour la SNCF qui montre que les voitures ne sont généralement pas perçues comme des « objets bruit » mais faisant partie intégrante d'un objet appelé « route » [RAIMBAULT-03]. Il semble que les sujets perçoivent les voitures comme des événements uniquement lorsqu'elles émergent clairement du bruit de fond alors que si leur activité n'est pas suffisante, les sujets n'y font pas attention. Cette particularité a déjà été relevée par V. Maffiolo [MAFFIOLO-99]

lorsqu'elle sépare les ambiances sonores urbaines de types « amorphes » ou « événementielles ». Les sources sonores des stimuli présentés aux auditeurs sont assez diverses et contrairement aux autres sources, les voitures ont tendance à être perçues comme faisant parti d'un « bruit de fond ».

Les voitures sont donc associées à la route qui génère le bruit de fond. Les mesures liées au temps (ou au nombre) et à l'intensité sont mal corrélées [Tableau 5] ce qui montre que les sujets ont entendu ces différences sur deux dimensions orthogonales :

	Nbre	%T	L _{Amax}
Nbre		0.71**	-0.07
%T	0.71**		-0.48
L _{Amax}	-0.07	-0.48	

Tableau 5 : Corrélations entre les mesures liées au temps (%T), au nombre et à l'intensité. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

b) ACP sur les Cyclomoteurs

Concernant les cyclomoteurs, la prégnance ainsi que la proximité sont logiquement corrélées au niveau (en particulier le niveau maximum L_{Amax}). La présence des cyclomoteurs est corrélée aux deux critères temporels : le nombre et le pourcentage de temps d'apparition [Tableau 6a].

	<i>Prégnance</i>	<i>Présence</i>	<i>Proximité</i>
Nombre	0.55	0.73*	0.45
%T	0.48	0.75*	0.41
L _{Aeq}	0.76*	0.47	0.85**
L _{Amax}	0.82**	0.68*	0.84**
L _{A5}	0.74*	0.51	0.82**
L _{A10}	0.67*	0.41	0.78*
Axe 1	0.98**	0.93**	0.96**
Axe 2	0.11	-0.36	0.23

Tableau 6a : Coefficients de corrélation entre la présence, la prégnance, la proximité des cyclomoteurs et les critères physiques calculés sur les courbes de L_{Aeq} codées cyclomoteurs, ainsi qu'avec les deux principaux axes de l'analyse en composantes principales. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

L'analyse multidimensionnelle révèle un axe principal qui explique 92.4% de la variance des réponses. Cet axe a un coefficient de corrélation supérieur à 0.93 pour les trois adjectifs que sont la prégnance, la présence et la proximité. Comme cela a été souligné plus haut, ces importantes

corrélations sont dues essentiellement à la construction des stimuli qui ne permettent pas d'effectuer des différenciations [Tableau 6b].

	Nb	%T	L _{Amax}
Nb		0.87**	-0.04
%T	0.87**		-0.24
L _{Amax}	-0.04	-0.24	

Tableau 6b : Corrélations entre les mesures liées au temps (%T), au nombre et à l'intensité

Tous les cyclomoteurs passent à la même distance du microphone, ils ont tous à peu près la même vitesse. Plus leur nombre augmente, plus le pourcentage de temps augmente mais leur intensité est sensiblement la même à chaque fois.

c) ACP sur les bus

Comme pour le cas des cyclomoteurs, les auditeurs ont clairement identifié les bus comme étant des « objets sonores » : leur présence est corrélée à leur temps d'apparition ce qui montre qu'ils sont clairement identifiés avec un début et une fin. De plus, la prégnance est corrélée aux critères énergétiques et en particulier au niveau maximum L_{Amax} (r = 0.98) [Tableau 7].

	<i>Prégnance</i>	<i>Présence</i>	<i>Proximité</i>
Nombre	0.15	0.70	0.01
%T	0.65	0.85	0.46
L _{Aeq}	0.78	0.43	0.83
L _{Amax}	0.96**	0.60	0.92*
L _{A5}	0.92*	0.53	0.84
L _{A10}	0.85	0.46	0.87*
Axe 1	0.98**	0.75	0.86
Axe 2	0.07	-0.65	0.48

Tableau 7 : Coefficients de corrélation entre la présence, la prégnance, la proximité des bus et les critères physiques calculés sur les courbes de L_{Aeq} codées bus, ainsi qu'avec les deux principaux axes de l'analyse en composantes principales. Les résultats avec ** indiquent une probabilité avec un niveau de confiance p<0.01 et les résultats avec * indiquent une probabilité avec un niveau de confiance p<0.05.

L'ACP montre que la variabilité des résultats est expliquée à 76.6% par le premier axe. Celui-ci est très corrélé au jugement de la prégnance (r = 0.98). Comme précisé plus haut, la prégnance est très bien corrélée au niveau maximum L_{Amax}. Ainsi cette mesure est un bon indicateur pour aider à différencier les bus entre eux dans une rue.

On retrouve le même schéma qu'avec les cyclomoteurs. Ici le second axe (22% de la variance) tend à être expliqué par le nombre (Nb) ou le temps de présence (%T).

d) ACP sur les motos

Les stimuli qui ont des motos ne présentent aucune variabilité car il n'y a qu'une seule moto par stimulus.

Aucune analyse ne peut alors être interprétable.

e) Conclusions sur les véhicules

L'analyse en composantes principales sur les stimuli comprenant des véhicules nous a permis de relever les éléments suivants :

- Dans une ambiance urbaine de type rue, nous sommes sensibles au caractère spatial des voitures. En effet, parmi les trois dimensions sur lesquelles les auditeurs ont eu la possibilité d'émettre des jugements perceptifs, c'est à travers les évaluations concernant la proximité, donc leur distance par rapport aux voitures, qu'ils jugent de manière pertinente les dissemblances entre les différents passages qui leur sont présentés. L'analyse que nous pouvons en faire est que les voitures, dans une situation purement urbaine, sont le bruit de fond et font partie intégrante de la route, quand bien même leur intensité les fait ressortir de ce bruit de fond. Elles ne sont donc pas véritablement perçues comme des « objets sonores ». L'habitude d'entendre des voitures, plutôt que d'autres sources (type cyclomoteur, moto ou bus), pourrait être un facteur qui fait que notre attention n'est pas dirigée sur le caractère énergétique de ce véhicule. Nous pouvons avancer l'explication que lorsqu'une source ne présente pas un danger ou n'y fait pas référence, notre attention, qui se focalise sur les informations « utiles » pour notre déplacement, se réfère au caractère spatial de la source.
- Les cyclomoteurs et les bus ont été évalués selon leur niveau pour la prégnance et selon leur nombre (ou durée d'apparition) pour la présence. Comme cela avait été suggéré lors des consignes de tests.

Les différences de notions entre la présence, la prégnance et la proximité sont bien perçues lorsque la construction des stimuli le permet.

L'approche adoptée dans le présent travail est de relier la signification des dimensions (prégnance, présence et proximité) de la source comme « objet bruit » indépendamment de l'étude de ses caractéristiques acoustiques propres. Dans ce sens nous avons uniquement utilisé des critères relatifs au niveau ou au temps car nous considérons que les propriétés spectrales des sources sont intrinsèques à l'identité de celles-ci. Nous n'excluons pas qu'il peut être également intéressant de mettre en relation les émergences spectrales existantes pour des véhicules tels que les cyclomoteurs et les bus (en comparaison du bruit de fond urbain qui a une répartition spectrale de type bruit rose) avec le fait qu'ils attirent l'attention. De même, le son de la majorité des voitures semble être caractérisé par un bruit de contact entre les pneumatiques et la chaussée, bruit qui s'avère être relativement peu « agressif ».

f) ACP sur les oiseaux

Dans un environnement urbain, les chants d'oiseaux sont la plupart du temps « mixés » avec d'autres sons. L'énergie ainsi mesurée et restituée à travers la courbe de L_{Aeq} codée « oiseaux », résulte de la superposition de plusieurs sources sonores. En conséquence, il n'est pas surprenant d'avoir une corrélation relativement faible entre les critères énergétiques comme le L_{Aeq} ou le L_{Amax} calculée sur l'évolution du L_{Aeq} codé et la prégnance ou la proximité, censées être relativement liées au niveau sonore.

D'un autre côté, la présence, présentée comme étant le temps d'apparition de la source est plutôt bien corrélée au pourcentage de temps d'apparition %T [Tableau 8].

	Prégnance	Présence	Proximité
%T	0.56	0.77**	0.60*
L_{Aeq}	0.22	- 0.06	0.19
L_{Amax}	0.56	0.37	0.50
Axe 1	0.98**	0.96**	0.98**
Axe 2	0.14	- 0.27	0.13

Tableau 8 : Coefficients de corrélation entre la présence, la prégnance, la proximité des oiseaux et les critères physiques calculés sur les courbes de L_{Aeq} codées oiseaux, ainsi qu'avec les deux principaux axes de l'analyse en composantes principales. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

Le premier axe de l'analyse en composantes principales explique quasiment la totalité de la variance des réponses (95%). Cela signifie que les trois notions n'ont pas été différenciées ce qui est probablement dû à la construction des stimuli. En effet, il y a peu de variabilité au niveau des chants d'oiseaux parmi les onze stimuli.

Cet axe est logiquement corrélé aux trois adjectifs de manière significative [Tableau 8].

g) ACP sur les voix d'adultes

Pour les voix d'adulte, nous avons le même problème qu'avec les oiseaux pour trouver des mesures physiques qui soient corrélées aux dimensions de prégnance et de proximité (les corrélations sont même négatives dans le Tableau 9). La courbe de L_{Aeq} correspondant aux moments d'occurrence des voix d'adultes prend en compte l'énergie provenant non seulement des voix mais également des autres sources et en particulier des véhicules. Cela illustre une limite dans l'utilisation de la courbe de L_{Aeq} pour traduire la sensation d'importance des voix d'adultes dans une ambiance sonore urbaine où plusieurs sources sont mélangées.

	<i>Prégnance</i>	<i>Présence</i>	<i>Proximité</i>
%T	0.62	0.32	0.57
L_{Aeq}	- 0.82**	- 0.70*	- 0.65
L_{Amax}	- 0.73*	- 0.69	- 0.60
Axe 1	0.98**	0.96**	0.98**
Axe 2	0.12	- 0.28	0.14

Tableau 9 : Coefficients de corrélation entre la présence, la prégnance, la proximité des voix d'adultes et les critères physiques calculés sur les courbes de L_{Aeq} codées voix adulte, ainsi qu'avec les deux principaux axes de l'analyse en composantes principales. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

Le cercle des corrélations issu de l'analyse en composantes principales illustre le fait que le premier axe explique la quasi-totalité de la variabilité des réponses. Il est de ce fait largement corrélé aux trois adjectifs subjectifs ($r > 0.96$ dans les trois cas).

h) ACP sur les voix d'enfants

Les corrélations du tableau 10 montrent que les critères de niveaux extraits de la courbe de L_{Aeq} codée spécifiquement s'avèrent être insuffisants pour trouver une mesure pouvant « expliquer » la sensation de proximité et de prégnance des voix d'enfants.

	<i>Prégnance</i>	<i>Présence</i>	<i>Proximité</i>
%T	0.69	0.83*	0.68
L _{Aeq}	- 0.79	- 0.84*	- 0.76
L _{Amax}	- 0.72	- 0.74	- 0.69
Axe 1	0.99**	0.98**	0.99**
Axe 2	0.06	- 0.16	0.10

Tableau 10 : Coefficients de corrélation entre la présence, la prégnance, la proximité des voix d'enfants et les deux principaux axes de l'analyse en composantes principales. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

La sensation de présence est quant à elle significativement corrélée à la durée d'apparition de la source ($r = 0.83$).

Le cercle des corrélations issu de l'analyse en composantes principales illustre le fait que le premier axe explique la quasi-totalité de la variabilité des réponses. Il est de ce fait largement corrélé aux trois adjectifs subjectifs ($r > 0.96$ dans les trois cas).

B- ACP pour les sources groupées

Cette deuxième série d'analyses a pour but de faire ressortir une hiérarchie perceptive des sources sonores. Les représentations graphiques de l'ACP sont exposées en Annexe 3.

a) ACP sur les voix

L'analyse fait apparaître deux composantes principales.

- La première explique 69.5% de la variance des réponses. Cet axe est significativement corrélé à la prégnance, à la présence et à la proximité des stimuli contenant des voix d'adultes ou des voix d'enfants. En examinant la composition des stimuli, on remarque que les deux sources voix enfants et voix adultes sont souvent associées dans les mêmes stimuli. De ce fait, le premier axe reflète la séparation faite par les auditeurs entre les stimuli contenant « beaucoup » de voix et ceux contenant « peu » de voix.
- Le second axe sépare les stimuli contenant uniquement un seul type de voix.

Nous pouvons également remarquer que le regroupement des stimuli contenant des voix d'enfants et des voix d'adultes a pour effet de rapprocher la sensation de présence

avec le pourcentage de temps d'apparition des sources. Ce sont les quatre stimuli ne contenant qu'une seule des sources (stimuli 6, 14, 15 et 19) qui amplifient les corrélations.

b) ACP sur les voix et sur les oiseaux.

L'ACP réalisée sur l'ensemble des réponses faites sur les stimuli contenant au moins des oiseaux ou des voix fait apparaître trois axes principaux qui, ensemble, expliquent 98.2% de la variance des réponses.

- Le premier axe (58.6% de la variance des réponses) est lié à l'activité des voix et oppose les stimuli contenant beaucoup de voix à ceux contenant peu de voix. Cet axe est inversement corrélé à la présence, à la proximité et à la proximité des oiseaux [Tableau 11]. Cela s'explique par le fait que dans notre expérience, les stimuli qui font entendre des voix ont tendance à ne pas contenir d'oiseaux.
- Le second axe (26.4% de la variance des réponses) est lié à l'activité des oiseaux. Il est de ce fait corrélé de manière significative à la présence, à la présence et à la proximité des oiseaux [Tableau 11].
- Le troisième axe (13.4% de la variance des réponses) oppose les stimuli contenant des oiseaux et des voix d'adultes aux stimuli contenant uniquement des voix d'enfants.

	Voix d'adultes			Voix d'enfants			Oiseaux		
	Preg	Pres	Prox	Preg	Pres	Prox	Preg	Pres	Prox
Axe 1	0.81**	0.91**	0.85**	0.78**	0.76**	0.77**	- 0.61*	-0.69**	- 0.62*
Axe 2	- 0.01	- 0.16	- 0.1	- 0.53	- 0.50	- 0.51	- 0.74**	- 0.66**	- 0.73**
Axe 3	0.57*	0.22	0.06	- 0.32	- 0.37	- 0.34	0.23	0.17	0.23

Tableau 11 : Coefficients de corrélation entre la présence, la présence, la proximité des voix et des oiseaux et les deux principaux axes de l'analyse en composantes principales. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

c) ACP sur tous les véhicules

L'analyse en composantes principales réduit l'ensemble des données (prégnance, présence et proximité pour chaque catégorie de véhicules) en quatre axes principaux qui expliquent 99.1% de la variabilité des réponses.

- Le premier axe (38.3% de la variabilité des réponses) est relié à la perception des bus. L'apparition des voitures sur cet axe est expliquée par le fait que, dans les stimuli, le nombre de voitures est inversement proportionnel au nombre de bus. Elles sont néanmoins présentes dans chacun des stimuli contenant des bus.
- Le second axe (32.2% de la variabilité des réponses) est expliqué par la dispersion des réponses concernant les motos.
- Le troisième axe (19.2% de la variabilité des réponses) est un facteur relié à la perception des cyclomoteurs [Tableau 13].

Ces trois facteurs qui expliquent la très grande majorité de la variabilité des réponses concernent la perception de trois types de véhicules : les bus, les motos et les cyclomoteurs. Les voitures n'interviennent pas distinctement dans la variabilité des réponses. Cela tend à appuyer l'argument en faveur d'une perception différente des voitures vis-à-vis des autres types de véhicules. Les voitures ont tendance à être perçues de façon moins « variable » que les autres véhicules.

De plus, l'importance perceptive des sources, c'est-à-dire de celles qui ont tendance à attirer l'attention et qui ont permis aux auditeurs de différencier les stimuli, commence à se dessiner. Nous avons d'un côté les bus, les cyclomoteurs et les motos qui sont responsables de la formation des trois premiers axes (et dont la variabilité va de « 0 » à « 2 ») et de l'autre côté les voitures qui n'apparaissent pas clairement malgré leur grande variabilité (elles vont de « 0 » voiture pour les stimuli six, sept, dix, quatorze, dix-neuf et vingt, à « 5 » voitures pour le stimulus quatre). Nous allons voir, à travers l'analyse factorielle des correspondances, que nous retrouvons la même hiérarchie à travers l'analyse des consensus des réponses.

- Le quatrième axe (9.4% de la variabilité des réponses) oppose les stimuli contenant des véhicules aux stimuli n'ayant pas de véhicules en tant qu'« objet bruit » (comme les bus, les motos, les cyclomoteurs et dans certains cas les voitures). Cet axe est logiquement corrélé à la mesure du L_{A90} global qui est en fait une mesure du bruit de

fond. Nous nous référons de nouveau à la thèse de Maffiolo [MAFFIOLO-99] lorsqu'elle fait la distinction entre les situations « événementielles » où l'on distingue clairement les sources et les situations « amorphes » dans lesquelles l'identification est très difficile.

	Voiture			Bus		
	Preg	Pres	Prox	Preg	Pres	Prox
Axe 1	- 0.79**	- 0.74**	- 0.79**	0.84**	0.85**	0.82**
Axe 2	0.46*	0.39	0.40	0.24	0.29	0.24
Axe 3	- 0.12	- 0.18	- 0.15	0.26	0.28	0.26
Axe 4	- 0.30	- 0.47*	- 0.41	- 0.39	- 0.32	- 0.43
	Moto			Cyclomoteur		
	Preg	Pres	Prox	Preg	Pres	Prox
Axe 1	0.20	0.19	0.17	- 0.43	- 0.43	- 0.43
Axe 2	- 0.80**	- 0.83**	- 0.81**	- 0.62**	- 0.61**	- 0.61**
Axe 3	- 0.51*	- 0.48*	- 0.49*	0.64**	0.65**	0.64**
Axe 4	- 0.23	- 0.18	- 0.25	- 0.14	- 0.09	- 0.14

Tableau 13 : Coefficients de corrélation entre la présence, la prégnance, la proximité des véhicules et les trois principaux axes de l'analyse en composantes principales. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

d) ACP sur toutes les sources

L'ACP sur l'ensemble des données (présence, prégnance et proximité pour chaque source sonore) permet de réduire l'information à un nombre de variables significatives qui sont classées par rapport à sept axes principaux.

- Les trois premiers axes (66.3% de la variabilité des réponses) sont représentatifs de la perception des sources telles que les bus, les cyclomoteurs, les motos et les voix d'enfants. Etant donné qu'il existe une grande variabilité des réponses sur ces données, cela signifie que les auditeurs ont été sensibles à ces types de sources.
- Le quatrième facteur (17% de la variance des réponses) est lié à l'activité des oiseaux.
- Les trois derniers facteurs expliquent des notions plus subtiles. Par exemple, ils illustrent la distinction entre les stimuli avec ou sans événements sonores nets ou bien encore la séparation entre les voix d'adultes et les voix d'enfants.

En définitif, seules les voitures n'apparaissent pas clairement sur l'un de ces axes. Bien que la variabilité de ces sources (de zéro à cinq voitures) ait a priori été la plus importante, ce n'est pas cela qui soit apparu comme le plus marquant aux auditeurs. Encore une fois on retrouve que les voitures ne sont pas interprétées comme des sources qui créent des événements : elles font partie du bruit de fond. Les bus et les cyclomoteurs créent des événements sonores caractérisés par leur niveau maximum, tandis que les voix ou les oiseaux sont plus facilement caractérisés par leur temps de présence.

Enfin, les voix d'enfants, en apparaissant dans les trois premiers axes, montrent qu'elles ont été perçues de manière aussi importante que les véhicules prégnants.

2.1.3.2. Analyse factorielle des correspondances

a) Transformation des données

L'AFC a pour but de compléter l'ACP déjà réalisée. Les axes perceptifs formés par l'AFC sont interprétables en termes de consensus des réponses des sujets. Ainsi les consensus permettent de trouver quelles sont les sources sonores qui permettent aux sujets de mieux ressentir des différences entre les séquences de 15 s qui leur ont été présentées, et, si possible, de savoir quelles en sont les dimensions associées (prégnance, présence ou proximité). Finalement, cette analyse va nous permettre d'infirmer ou de confirmer les conclusions issues de la première analyse.

A cet effet, il faut laisser apparaître de manière plus significative les « distorsions » qui peuvent exister dans les réponses des sujets. Ainsi, au lieu d'être directement moyennées pour les besoins de l'ACP, les réponses de prégnance, de présence et de proximité des sources sonores sont arbitrairement transformées en catégories sémantiques de la façon suivante :

Les réponses ayant la note « 0 » sont classées dans la catégorie « pas ». Les réponses ayant des notes allant de « 1 » à « 3 » sont classées dans la catégorie « peu » (prégnant, présent ou proche). Les réponses ayant des notes allant de « 4 » à « 6 » sont classées dans la catégorie « assez ». Enfin, les réponses ayant des notes allant de « 7 » à « 10 » sont classées dans la catégorie « très ». Une fois ces catégories réalisées par personnes, nous effectuons un regroupement en histogramme des effectifs.

A titre d'illustration, pour l'individu « stimulus 1 » et la variable « prégnance voiture », nous avons « 0 » personne qui ait répondu la note « 0 », une personne qui ait répondu entre « 1 » et « 3 », 9 personnes qui aient répondu entre « 4 » et « 6 », et 10 personnes qui aient répondu entre « 7 » et « 10 ». Ainsi, nous allouons la note « 0 » à la catégorie « pas prégnant », la note « 1 » à la catégorie « peu prégnant », la note « 9 » à la catégorie « assez prégnant » et enfin la note « 10 » pour la catégorie « très prégnant » ; cela pour les voitures du stimulus un. La figure 17 illustre ce propos.

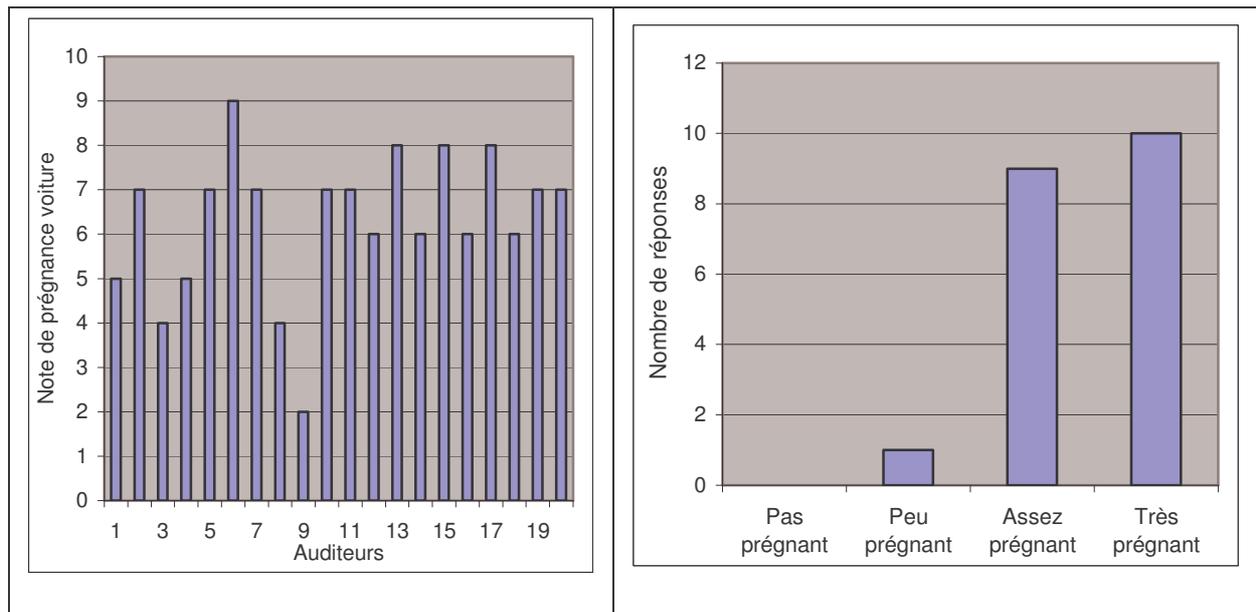


Figure 17 : Illustration de la transformation des données pour l'analyse factorielle des correspondances. Ces données correspondent aux notes de prégnance voiture pour le stimulus numéro un. A gauche les notes mises par les vingt auditeurs, à droite l'histogramme des réponses selon les quatre catégories sémantiques.

b) Individus et variables

Les individus sont les vingt stimuli de quinze secondes contenant les sources sonores mentionnées. Les variables qualitatives sont les catégories sémantiques décrites ci-dessus : « pas », « peu », « assez » et « beaucoup », suivies de l'adjectif prégnant, présent ou proche et d'un type de source sonore. Elles sont également appelées par la suite variables nominales. Elles sont dites « actives » car elles participent à la formation des axes. Les mesures physiques sont des variables dites « passives » car elles sont simplement projetées sur les axes déjà formés.

c) Procédure

Les AFC sont réalisées de façon graduelle. Tout d'abord en formant un groupe « véhicules », puis avec un groupe « oiseaux et voix et » enfin en regroupant toutes les données.

d) Interprétations

L'AFC traite un tableau de contingence qui résulte du comptage des réponses des sujets aux variables nominales décrites ci-dessus. Ainsi, l'information de la répartition par effectifs des réponses est conservée. L'AFC va réduire les dimensions de départ sur un nombre réduit d'axes expliquant la plus grande variabilité possible des résultats. Nous exprimerons ces résultats sur des graphiques représentant les plans factoriels construits par la réunion de deux facteurs. Sur ces plans, la distance entre deux points s'interprète comme une ressemblance entre les profils de ces points. Du fait que le tableau de contingence soit construit sur la répartition par effectifs des réponses, les axes extraits de l'AFC expriment les consensus dans les notations des sujets. Les graphiques représentant les dimensions nous aident à voir aisément à quelles variables nominales elles se rapportent. Toutefois nous complétons cette interprétation visuelle par des tableaux de corrélations. Ceux-ci affichent les corrélations ou coefficients de Pearson entre les variables qualitatives (actives et passives) et les principaux axes factoriels.

Dans la projection sur les plans factoriels, les notations des données nominales sont les suivantes : « adjectifs pg, ps ou px » pour prégnance, présence ou proximité, suivies de l'abréviation en deux ou quatre lettres de la source sonore (« Oi » pour oiseaux, « Voit » pour voiture, « Mob » pour cyclomoteurs...) puis le chiffre correspondant (pas de chiffre pour « pas », « 1 » pour « peu », « 2 » pour assez et « 3 » pour « beaucoup »).

Par exemple « pg_bus_2 » correspond à « assez de prégnance de bus », « px_Oi_0 » correspond à « pas de proximité d'oiseaux » etc.

A- AFC sur les véhicules.

Tous les stimuli contiennent au moins un passage de véhicule.

Les objectifs de cette analyse factorielle, à travers une analyse des consensus, sont de trouver quelles sont les sources sonores qui permettent aux sujets de mieux ressentir des différences entre les séquences de 15 s qui leur ont été présentées, et, si possible, de savoir quelles en sont les dimensions associées (prégnance, présence ou proximité).

La structure réduite par l'AFC propose quatre facteurs expliquant 84% de la variabilité des données nominales.

- Le premier facteur (35% de la variance des données) est lié à l'identification des autobus. Les extrémités de cette dimension opposent les séquences ayant de fortes composantes relatives aux bus sans contenir d'autres sources audibles (présence de deux bus dans le stimulus « 19 » et présence de un bus dans le stimulus « 10 ») aux séquences sans bus, en passant par des séquences contenant un bus pouvant avoir des niveaux sonores parmi les plus élevés de l'ensemble des stimuli (stimuli « 6 », « 3 » et « 18 ») mais incluant également d'autres sources audibles [Figure 18]. Le consensus des réponses est probablement dû au niveau maximum des bus qui est plus important que pour les autres véhicules.

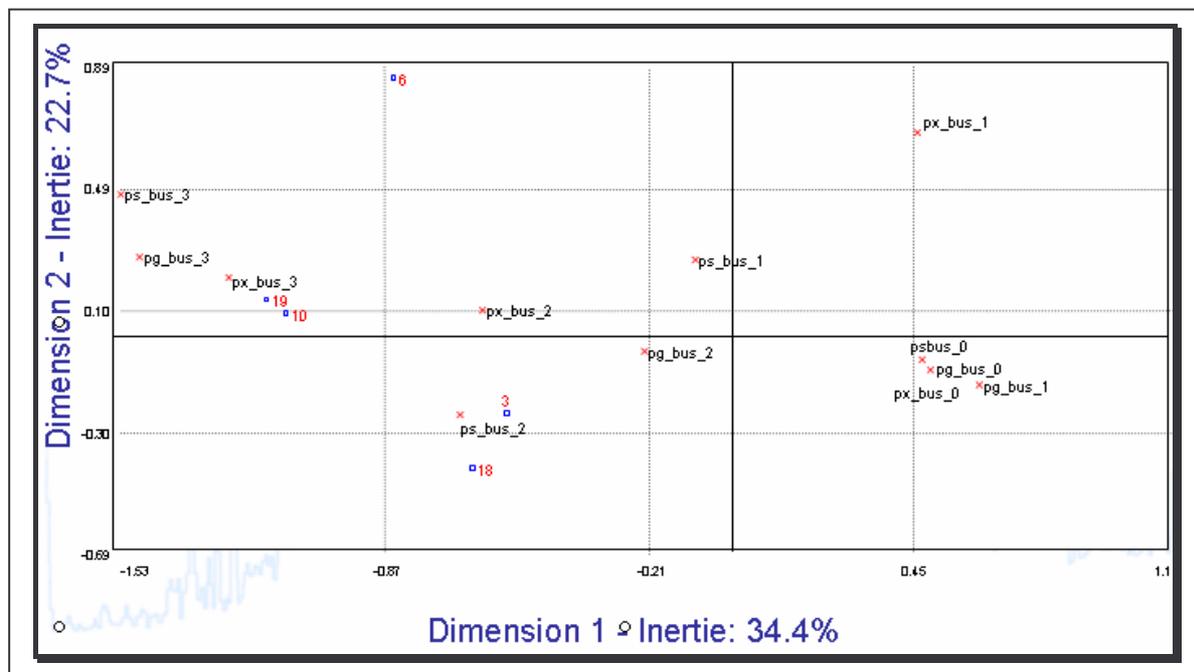


Figure 18 : Représentation des stimuli contenant des bus et leur trois niveaux nominaux sur le plan factoriel 1-2.

Ce premier axe nous indique que c'est la perception de critères relatifs aux bus qui obtient le maximum de consensus. Cela montre à la fois que les bus ont donc été correctement identifiés et que les jugements des auditeurs sont allés dans le même sens. De ce fait, les bus contenus dans certains stimuli ont permis à ces derniers d'être évalués de manière différente des autres.

- Le second facteur (23% de la variance des données) est lié à l'identification des cyclomoteurs [Figure 19]. Il oppose les stimuli contenant plus ou moins de cyclomoteurs.

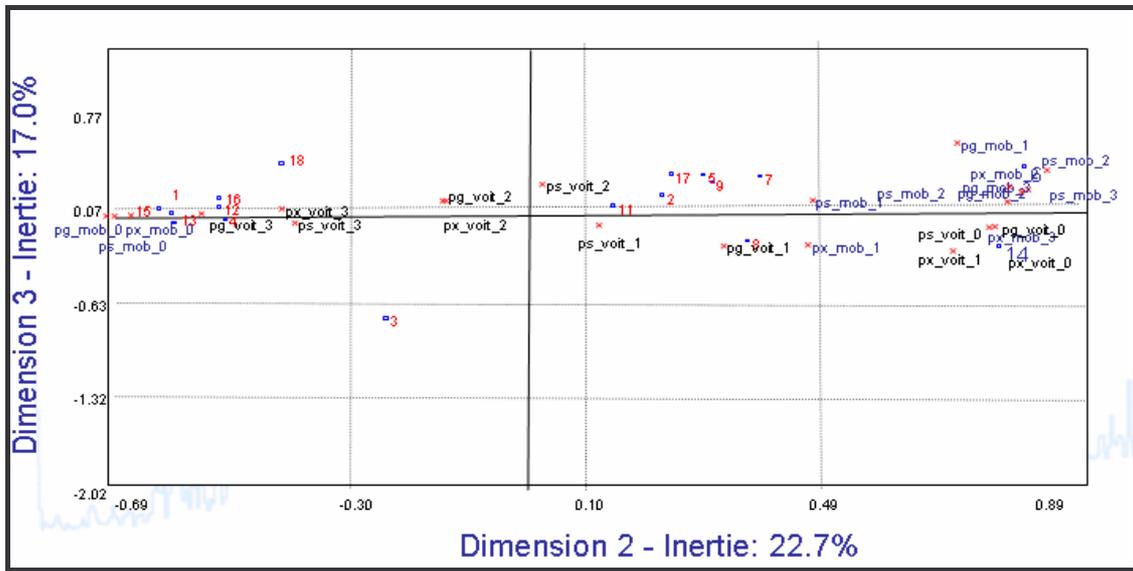


Figure 19 : Représentation des stimuli contenant des cyclomoteurs et/ou des voitures et leurs trois niveaux nominaux associés sur le plan factoriel 2 - 3.

Bien que l'identification des voitures soit distribuée sur les deux dimensions comme l'indiquent les corrélations du tableau 14, ces dernières ne sont pas directement responsables de la formation de ces axes dans le sens où elles n'ont pas la même « importance perceptive » que les bus ou les cyclomoteurs. L'explication est à trouver dans la construction des stimuli. En effet, on remarque que plus nous avons de voitures dans les stimuli, moins nous avons de bus ou de cyclomoteurs, et inversement. De ce fait, les deux premiers axes sont formés avant tout par les consensus qu'il y a eu sur la perception des cyclomoteurs et des bus.

- Le troisième facteur (17% de la variance des données) est expliqué par l'identification des motos [Figure 20], spécialement du stimulus « 20 » où la moto est le seul véhicule.

La prégnance, la présence et la proximité sont toutes autant corrélées à cette dimension [Tableau 14].

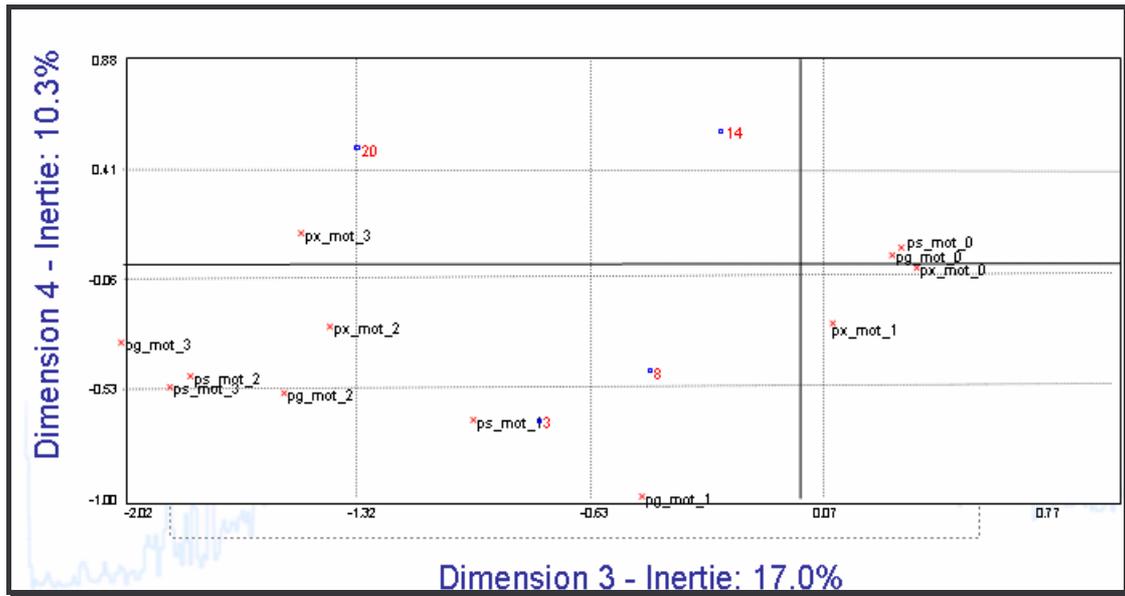


Figure 20 : Représentation des stimuli contenant des motos et leurs trois niveaux nominaux associés sur le plan factoriel 3 - 4.

	<i>Dimension 1</i>		<i>Dimension 2</i>		<i>Dimension 3</i>
	Bus	Voit	Mob	Voit	Moto
Prégnance	- 0.86**	0.64**	0.77**	- 0.65**	- 0.88**
Presence	- 0.87**	0.59*	0.76**	- 0.55*	- 0.84**
Proximité	- 0.84**	0.66**	0.77**	- 0.56**	- 0.85**
L _{Amax}	- 0,87**	0.55*	0.79**	0.58**	- 0.80**
%T	- 0.88**	0.40	0.60*	- 0.76**	- 0.86**
Nombre	- 0.87**	0.41	0.67**	- 0.56	- 0.82**

Tableau 14 : Coefficients de corrélation entre la présence, la prégnance, la proximité des véhicules (variables actives), les mesures physiques (variables passives) et les trois principaux axes de l'analyse factorielle des correspondances. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

Les trois premiers axes sont donc formés par les consensus obtenus, dans l'ordre croissant d'importance, par les bus, les cyclomoteurs et les motos. Dans l'ensemble des stimuli, ces véhicules ont une variabilité en termes de nombre, allant de « 0 » à « 2 » pour les bus et les cyclomoteurs et de « 0 » à « 1 » pour les motos. Les voitures, bien qu'ayant une variabilité en nombre et en intensité supérieure aux autres véhicules, ne forment pas directement d'axe. Ainsi, nous pouvons conclure que, lors des écoutes, les auditeurs ont été plus sensibles à une variation de « 0 » à « 2 » bus (ou de « 0 » à « 2 » cyclomoteurs) produite dans les stimuli qu'à une variation entre « 0 » et « 5 » voitures.

- Le quatrième facteur (10.3% de la variance des données) explique les erreurs dans l'identification des voitures et des motos pour les stimuli « 3 » et « 8 » qui sont opposés, sur cette quatrième dimension, aux stimuli « 20 » et « 14 » [Figure 21]. Les stimuli « 20 » et « 14 » contiennent effectivement des motos, les stimuli « 3 » et « 8 » contiennent également des motos mais aussi des voitures.

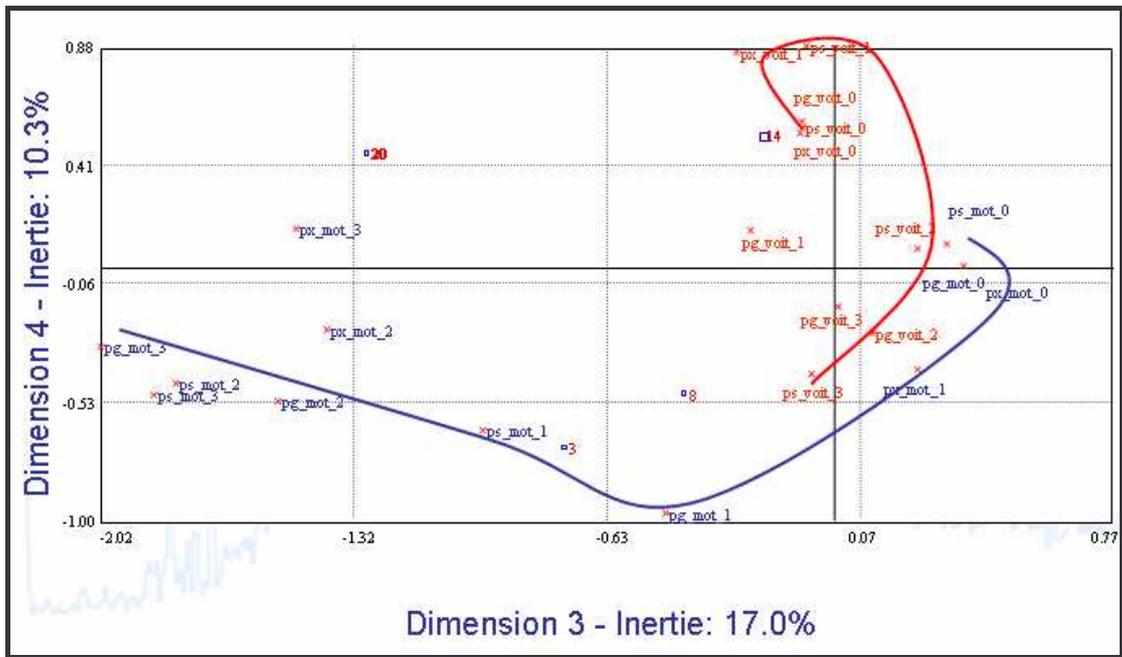


Figure 21 : Représentation des stimuli contenant des motos ou des voitures et leurs trois niveaux nominaux associés sur le plan factoriel 3 - 4

La figure 22 illustre par un histogramme des réponses (« présence motos ») les erreurs de jugements (ou d'identification) sur les stimuli « 3 » et « 8 » reflétant cette fois-ci un manque de consensus dans les réponses des sujets. Ces erreurs sont dues à une confusion entre les motos et les voitures. Logiquement, aucune mesure physique n'est corrélée de façon nette à cet axe.

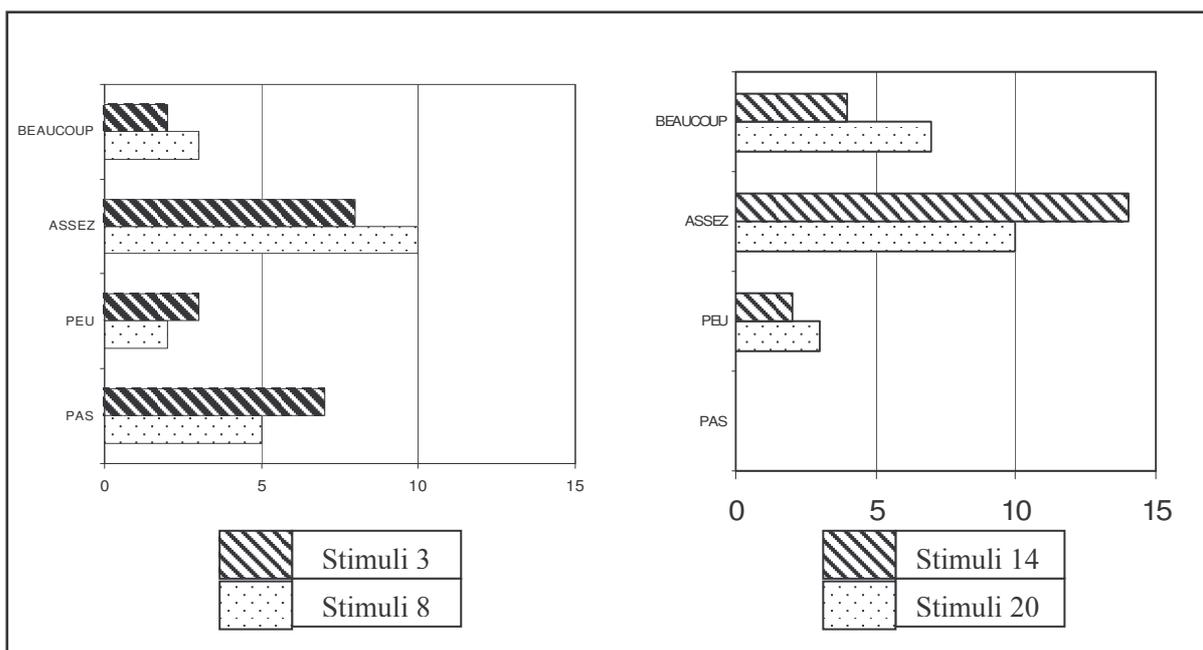


Figure 22 : Histogramme des effectifs concernant les réponses sur le “présence” des motos pour 4 stimuli différents. S’y opposent les erreurs dans les stimuli « 3 » et « 8 » aux bonnes réponses dans les stimuli « 14 » et « 20 ».

Ce quatrième axe permet de retrouver les erreurs de jugements en partie détectées lors de l’analyse des histogrammes de paragraphe 2.2.1.b (Figure 13d) qui sont dues à de mauvaises identifications.

B- AFC sur les voix et les oiseaux

La structure réduite propose trois facteurs expliquant 81.7% de la variance des données nominales. Etant donné que la prégnance, la présence et la proximité sont fortement corrélées, pour des raisons de lisibilité, nous n’affichons que la proximité des sources sonores dans les plans factoriels.

- Le premier facteur (43.4% de la variance des données) est expliqué par l’identification des voix d’enfants. Il oppose les séquences avec et sans voix d’enfants [Figure 23].

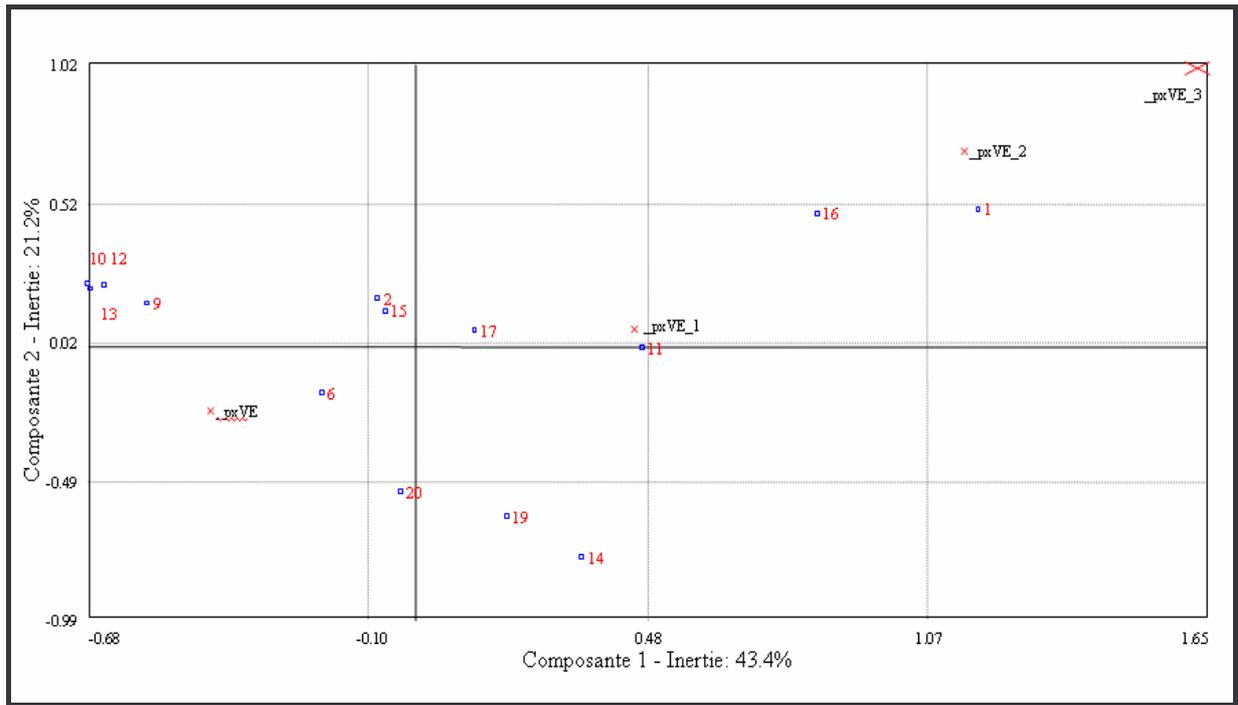


Figure 23 : Représentation des stimuli contenant des voix d'enfants et leur niveau nominal de proximité associé sur le plan factoriel 1-2.

- Le second facteur (21.2% de la variance des données) est expliqué par l'identification des voix d'adultes [Figure 24].

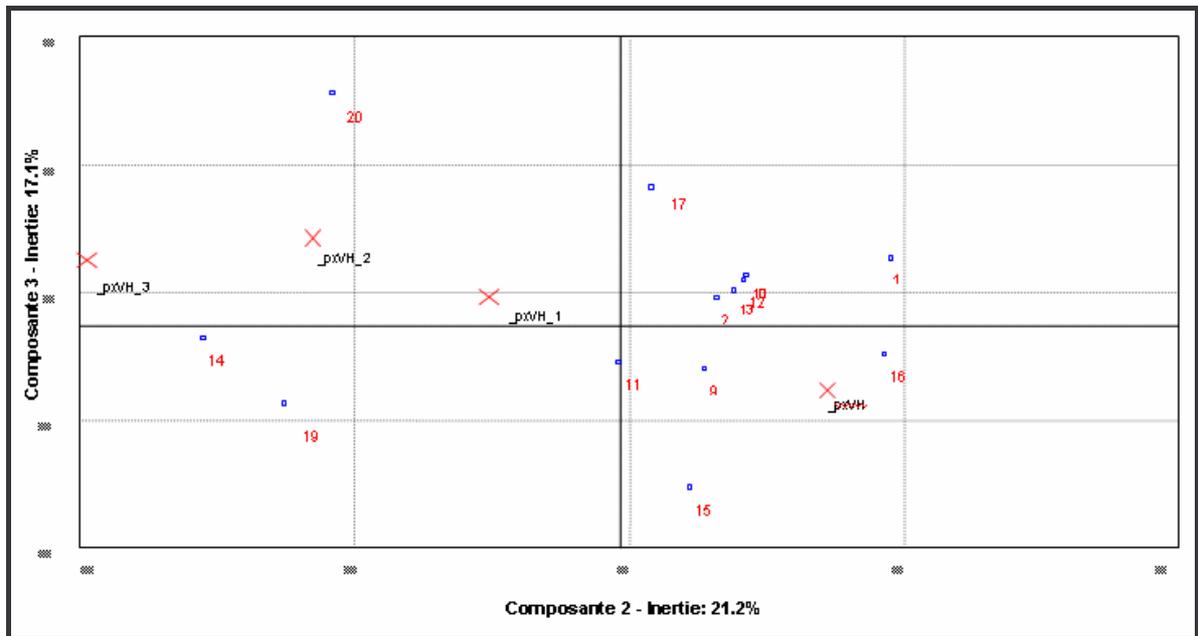


Figure 24 : Représentation des stimuli contenant des voix d'adultes et leur niveau nominal de proximité associé sur le plan factoriel 2-3.

- Le troisième facteur (17.1% de la variance des données) est expliqué par l'identification des oiseaux [Figure 25].

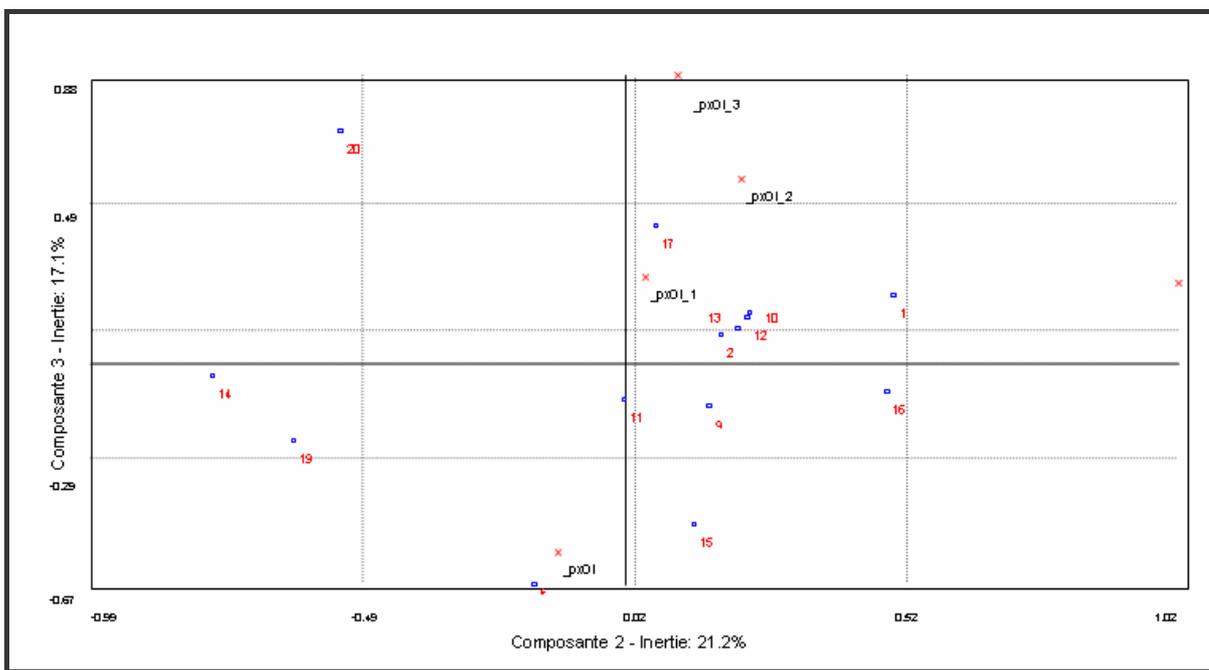


Figure 25 : Représentation des stimuli contenant des oiseaux et leur niveau nominal de proximité associé sur le plan factoriel 2-3.

	<i>Dimension 1</i>		<i>Dimension 2</i>		<i>Dimension 3</i>
	Voix enfants	Voix adultes	Voix adultes	Voix enfants	Oiseaux
Prégnance	0.83**	0.57*	0.74**	0.54*	0.85**
Presence	0.80**	0.56*	0.86**	0.51*	0.74**
Proximité	0.83**	0.57*	0.80**	0.59*	0.85**
L_{Amax}	0.74**	0.57*	0.71**	-0.43	0.78**
L_{Acq}	0.72**	0.55*	0.71**	-0.43	0.75**
%T	0.84**	0.61*	0.83**	0.49	0.61*

Tableau 15 : Coefficients de corrélation entre la présence, la prégnance, la proximité des voix et des oiseaux et les trois principaux axes de l'analyse en composantes principales. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

Nous retrouvons la même « hiérarchie » perceptive que lors de l'ACP, à savoir les voix d'enfants qui ont la plus grande importance perceptive (et apportent le plus de consensus), suivies des voix d'adultes et enfin des oiseaux [Tableau 15].

C- AFC sur toutes les sources

L'AFC sur l'ensemble des sources permet de retrouver les sept axes de l'ACP. Chacun de ces axes correspond à l'une des sept catégories de sources étudiées.

2.1.4. Discussion

L'AFC a été un moyen, à travers l'expression des consensus des réponses des sujets, de confirmer les résultats de la première analyse. En effet, les axes principaux trouvés lors des deux analyses sont sensiblement les mêmes.

Pour interpréter les analyses, il faut tout d'abord noter que les résultats de ces analyses sont avant tout le reflet de la construction des stimuli, même si ces derniers ont été extraits de séquences réelles. Ainsi, les corrélations entre les trois adjectifs présentés (prégnance, présence et proximité) sont importantes car ces derniers sont liés par la structure de la configuration urbaine. Nonobstant, les axes principaux montrent que certaines sources sonores sont perçues comme étant plus importantes que d'autres. En particulier, les auditeurs sont plus sensibles à une différence entre « 0 » et « 2 » bus qu'à une différence entre « 0 » et « 5 » voitures.

Dans les séquences présentées, les voitures ne sont pas clairement perçues comme des « objets bruits », avec un début et une fin, à l'instar des cyclomoteurs, des motos et des bus, mais comme faisant partie intégrante de la route.

Un autre fait qui ressort de nos analyses est l'importance que prennent les voix dans une ambiance sonore. Elles ressortent plus en termes de consensus pour les auditeurs. Les voix, comme éléments significatifs de présence humaine et porteuses d'informations, ont en effet déjà été reconnues comme étant des éléments qui attirent l'attention [VIOLLON-00].

2.1.5 Synthèse des expériences sur la prégnance, la présence et la proximité des sources sonores

Le but de ces expériences (test n° 1 et test n° 4) est de caractériser l'environnement sonore urbain du point de vue des éléments le composant. Pour cela, un espace subjectif est défini. Il est spécifique aux sources sonores et prend en compte les dimensions énergétique (prégnance), temporelle (présence) et spatiale (proximité) ; la dimension spectrale étant intrinsèque à la source.

A partir d'expériences de jugements des ces trois dimensions sur vingt séquences sonores d'ambiance urbaine, voici les principales observations et hypothèses que nous tirons de ces tests :

- La sensation de prégnance des voitures est corrélée aux mesures relatives au nombre et au temps de présence de celles-ci alors que la sensation de prégnance des autres véhicules (bus, motos et cyclomoteurs) est corrélée à des mesures relatives aux niveaux sonores. De plus, la meilleure des trois dimensions permettant de différencier les stimuli contenant des voitures est la proximité, elle-même liée au niveau sonore. Les auditeurs seraient donc sensibles à une représentation spatiale des voitures.
- De plus, nous avons mis en évidence un décalage entre des valeurs objectives et leurs corrélats subjectifs. En effet, on constate que si dans l'ensemble des stimuli, les voitures sont les sources qui obtiennent la plus grande variabilité en termes de nombre, d'intensité et de temps de présence, il n'en est pas de même dans les jugements recueillis. Les axes principaux formés par les analyses factorielles reflètent une variabilité importante des jugements (ainsi que des consensus) concernant la prégnance, la présence et la proximité des bus, cyclomoteurs, motos et voix d'enfants. Les auditeurs se sont donc basés en priorité sur ces sources pour différencier les vingt stimuli qui leur ont été présentés. Une conséquence pratique de ces observations est que les auditeurs sont plus sensibles à une variation entre « 0 » et « 2 » bus qu'à une variation entre « 0 » et « 5 » voitures.
- Ces observations tendent à montrer que, contrairement aux autres véhicules précités, considérés comme des « objets bruits », car émergents clairement du bruit de fond, les voitures sont perçues comme faisant partie intégrante de la route.
- Pour l'ensemble des catégories de véhicules établies (voitures, motos, bus et cyclomoteurs), la sensation de présence est proportionnellement reliée au nombre de

véhicules. Nous obtenons ainsi une bonne correspondance entre un paramètre subjectif et son corollaire objectif.

- La sensation de présence des oiseaux et des voix est correctement reliée au critère temporel qu'est la durée d'apparition. Cependant aucune mesure objective ne décrit correctement leur prégnance. En effet, si les auditeurs ont pu correctement évaluer leur « importance » dans l'ensemble des stimuli, il n'est pas possible d'extraire une mesure énergétique propre à ces sources, à partir d'un codage manuel des courbes de L_{Aeq} . Cette étude met donc en évidence l'insuffisance de l'utilisation du L_{Aeq} pour mesurer l'énergie de ce type de sources car souvent superposées à des véhicules.

Il est important de signaler que ces observations ont été effectuées sur un nombre limité de stimuli. Bien que nous ayons effectué des recoupements avec d'autres conclusions de travaux de recherche, il est clair que c'est au travers d'investigations plus poussées (et notamment in-situ) que l'on pourra évaluer le degré de pertinence de ces résultats, non seulement sur des ambiances similaires mais également sur d'autres configurations sonores.

2.2. Tests n° 2 et n° 3 : Force et désagrément sonore sur les séquences de 15 s

2.2.1. Questionnaire et procédure

Les deux tests, l'un pour le désagrément, l'autre pour la force sonore, ont été réalisés de manière séparée, pour que les notions ne se confondent pas. Pour une question de commodité, la description des deux tests a été regroupée dans le même paragraphe.

La procédure et le déroulement de ces deux tests (deux et trois) ont été identiques au test n° 1. Les figures [26-27] représentent les interfaces des tests telles que proposées aux auditeurs. La durée totale d'un test a été cette fois-ci de 10 min environ.

Une feuille d'explication est fournie aux auditeurs [Annexe 4]. Pour que les réponses soient les plus spontanées possibles, nous avons demandé aux sujets de rentrer les notes sous forme de nombres entiers allant de « 0 » à « 10 ». Des correspondances verbales sont annotées au dessus des graduations. Celles-ci vont, pour les adjectifs intense ou désagréable, de « très peu » (notes « 1 » et « 2 »), à « extrêmement » (notes « 9 » et « 10 ») en passant par « légèrement » (notes « 3 » et « 4 »), « moyennement » (notes « 5 » et « 6 ») et « fortement » (notes « 7 » et « 8 »). Aucun sujet n'a rentré la note « 0 » ni pour l'intensité, ni pour le désagrément.



Figure 26 : Interface du test de jugement de la force sonore pour les vingt stimuli.

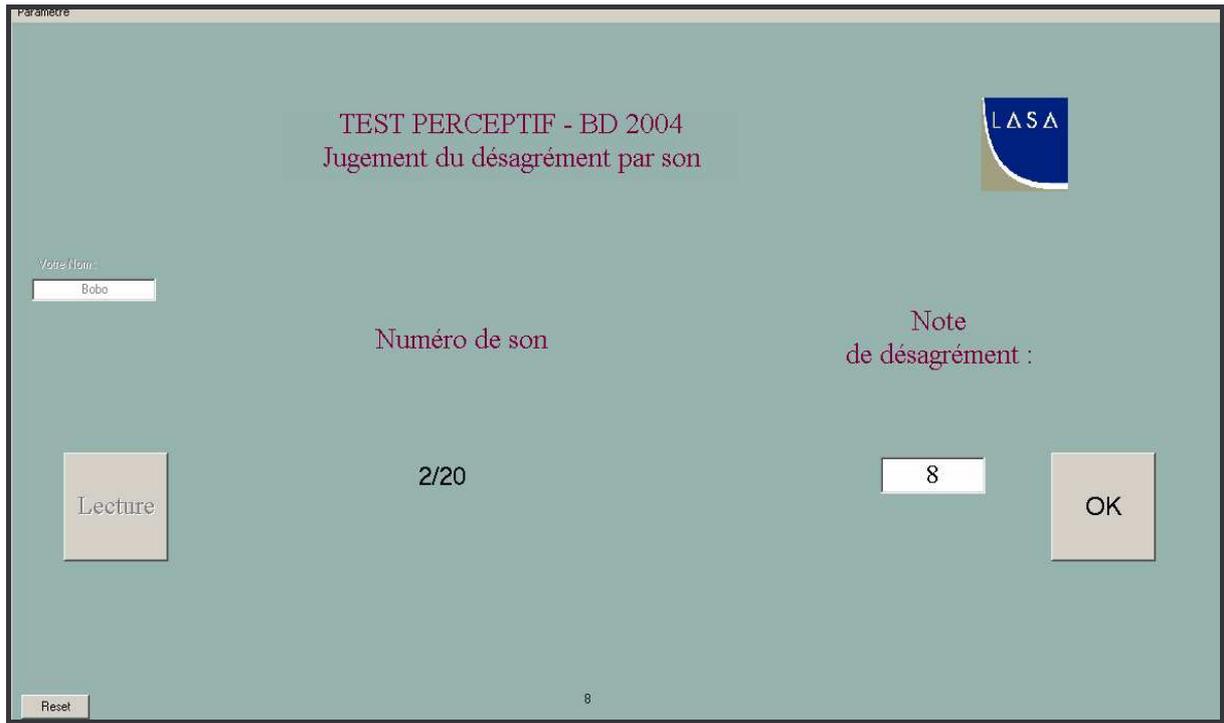


Figure 27 : Interface du test de jugement du désagrément sonore pour les vingt stimuli.

2.2.2. Normalisation des données

Bien que les réponses des sujets soient notées sur les mêmes échelles, elles sont tout de même normalisées. La normalisation est la suivante :

$$X_{norm,i,p} = X_{i,p} \sqrt{\frac{\sum_{i,p} X_{i,p}^2}{\sum_i X_{i,p}^2}}$$

Où

$p = 1$ à P personnes

$i =$ stimuli 1 à 20,

$X_{i,p}$ = réponse initiale de la personne p pour le stimulus i ,

$X_{norm,i,p}$ = réponse normalisée de la personne p pour le stimulus i ,

$\sum_{i,j} X_{i,p}^2$ = somme des carrés de toutes les réponses de la personne p ,

$\overline{\sum_{i,p} X_{i,p}^2} = (\sum_{p=1}^P \sum_i X_{i,p}^2) / P$ = moyenne de la somme des carrés de tous les sujets.

2.2.3. Relations entre force sonore estimée et désagrément sonore

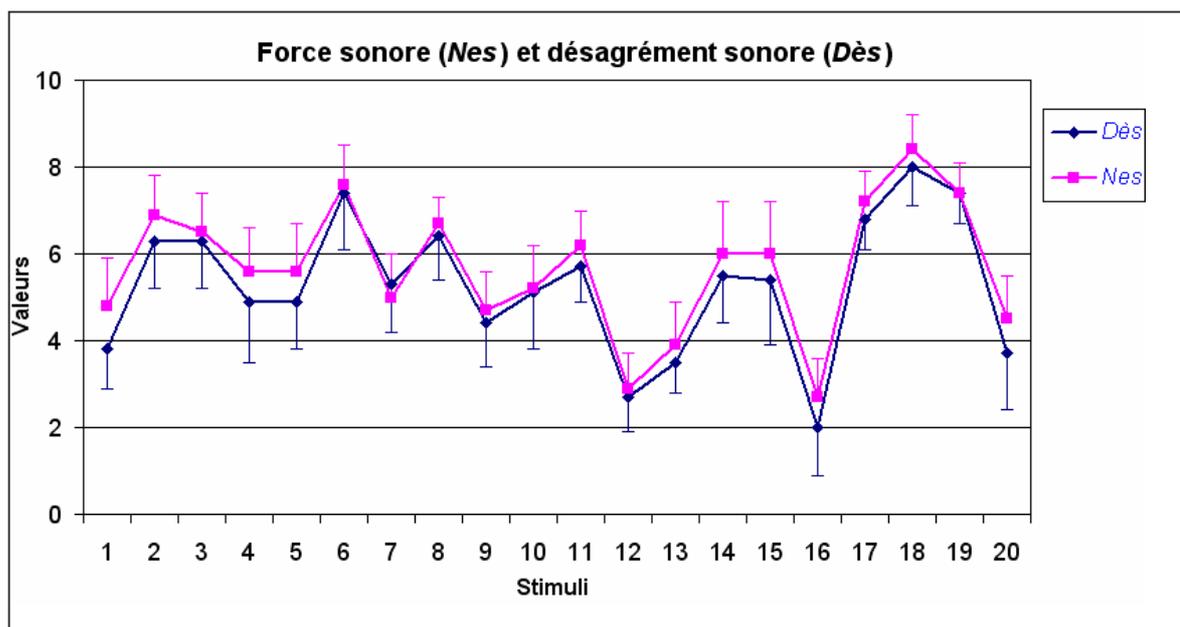


Figure 28 : Graphique du désagrément sonore estimé et de la force sonore estimée avec écart type sur les vingt stimuli de 15 s représentant une ambiance sonore urbaine. Les moyennes ont été effectuées sur vingt-six sujets pour la force sonore et vingt-deux sujets pour le désagrément.

On voit tout d'abord sur ce graphique [Figure 28] que la force sonore estimée et le désagrément sont très proches. Ils sont liés par un coefficient de détermination ajusté $R^2_{aj} = 96.3\%$.

Une observation point par point montre que pour l'ensemble des stimuli, à l'exception du stimulus numéro sept, les valeurs de la force sonore sont supérieures aux valeurs du désagrément. Le stimulus numéro sept contient deux passages de cyclomoteurs. C'est le seul stimulus parmi les vingt à ne contenir qu'un seul type de source identifiable.

Les stimuli un, quatre, cinq et quinze sont les stimuli dont les écarts perçus entre la force sonore et le désagrément sont les plus importants. Si l'on se reporte au tableau 2, on constate que ces stimuli contiennent les sources sonores suivantes : voitures, voix et oiseaux.

Ces observations tendent à montrer que la différence de jugement entre le désagrément et la force sonore est dépendante du type de source perçue. Si on fait l'hypothèse que les voix et les oiseaux ont une influence positive et sont donc perçus de manière agréable (tout en supposant la neutralité perceptive des voitures), ces variations d'écarts peuvent donc s'expliquer si l'on considère la représentation mentale de la source. Cette représentation mentale est a priori qui consiste à avoir une idée déjà établie du caractère agréable ou désagréable de ce que représente la source sonore. C'est sa signification.

Il se peut également que cette représentation mentale soit liée au sentiment d'adéquation entre les sources et les lieux dans lesquels elles apparaissent, ainsi que nous le suggère la perception hédonique des sons des voitures qui est évaluée différemment selon le lieu où l'on se trouve (rue circulée ou parc).

2.2.4. Relation entre la force sonore et les critères physiques

Nous réalisons ici une étude de corrélation pour évaluer quelle mesure physique est la plus proche de la sensation de force sonore. Les valeurs objectives testées, qui sont bien sûr des mesures relatives à l'intensité sonore, sont les suivantes : Sonie moyenne (N) calculée sur le modèle de Zwicker, sonie maximum dépassée pendant 5% du temps (N5), L_{Aeq} moyen, L_{Amax} , L_{A10} , L_{A5} .

Les résultats des coefficients de détermination ajustés R^2_{aj} (carré du coefficient de corrélation ajusté au nombre de variables explicatives, voir paragraphe 3.3.3) sont visibles dans le tableau 16 a suivant :

	N	N10	L_{Aeq}	L_{Amax}	L_{A10}	L_{A5}
Force sonore	94.5%	83.3%	93.4%	70.9%	90.1%	79.3%

Tableau 16 a : Tableau des coefficients de déterminations ajustées R^2_{aj} entre la force sonore et des grandeurs physiques pour les vingt stimuli de 15 s. Toutes les corrélations ont une probabilité avec un niveau de confiance $p < 0.001$.

La meilleure mesure de la force sonore est donc la sonie moyenne (N) dont la pertinence ($R^2_{aj} = 94.5\%$) est presque égale à celle du niveau de pression équivalent moyen L_{Aeq} ($R^2_{aj} = 93.4\%$).

Bien que les différentes procédures expérimentales employées pour estimer la sonie puissent créer des biais, il est intéressant de rapprocher les résultats du tableau avec d'autres résultats de la littérature.

Fastl et Zwicker [ZWICKER-90], [ZWICKER-99] proposent d'utiliser le N10 (sonie dépassée de 10% du temps) puis le N4 (sonie dépassée pendant 4% du temps) pour estimer la sonie globale de sons dont l'intensité varie dans le temps. Ces résultats sont issus d'études portant sur des bruits d'avion, de trafic urbain ainsi que de trafic ferroviaire. Cependant, les auteurs n'indiquent pas la durée des stimuli sur lesquels les tests d'estimations de sonie ont porté.

Meunier et al [MEUNIER-02] ont étudié la sonie sur des stimuli d'une durée de 6 s. Les sons étudiés sont synthétiques et représentent des « extrapolations temporelles » de trois types de sons : une voyelle humaine « a », un son pur à 1000 Hz et un bruit blanc filtré par un filtre passe bas à 4000 Hz. Des modèles de régression linéaire tendent à faire ressortir la sonie moyenne cette fois-ci, d'une manière identique aux travaux de ce mémoire.

Lors d'une expérience réalisée in situ et que nous détaillons au paragraphe 3.2 (sur trois ambiances différentes : rue circulée, parc et marché) et sur des séquences d'une minute et 30 secondes, la sonie estimée est toutefois plus proche de la sonie maximale dépassée pendant 10% du temps [CAN-05].

Nous pouvons ainsi émettre l'hypothèse que, sur des sons de durée assez courte, c'est-à-dire de moins de 20 s, les sujets ne font pas appel à des « stratégies » de mémorisation (en retenant les événements marquants, en intensité par exemple) mais qu'ils sont capables d'intégrer de manière objective l'intensité sur l'ensemble de la durée du stimulus faisant ainsi ressortir le calcul de la sonie moyenne.

A l'observation de la relation linéaire établie entre la force sonore et la sonie moyenne [Figure 29], il apparaît que le stimulus le plus excentré du modèle est le stimulus « 16 », dont l'évaluation subjective est inférieure à la prédiction de la mesure. Il a donc été jugé moins intense par rapport au modèle. Ce stimulus est (avec d'autres) celui qui contient le moins de véhicules c'est-à-dire une voiture au début de la séquence puis ce sont des voix et des oiseaux qui remplissent l'espace sonore. Il est sensiblement du même profil que les stimuli « 10 » (bus au début et oiseaux à la fin) et « 20 » (moto au début, oiseaux à la fin) qui sont mieux décrits par le modèle [Figure 29] et donc par la sonie moyenne. Dans ce cas, peut-être que le jugement de l'intensité a été biaisé par la représentation mentale de la source [MAFFIOLO-99]. La voiture étant, en conséquence, jugée moins « agressive » qu'une moto ou qu'un bus.

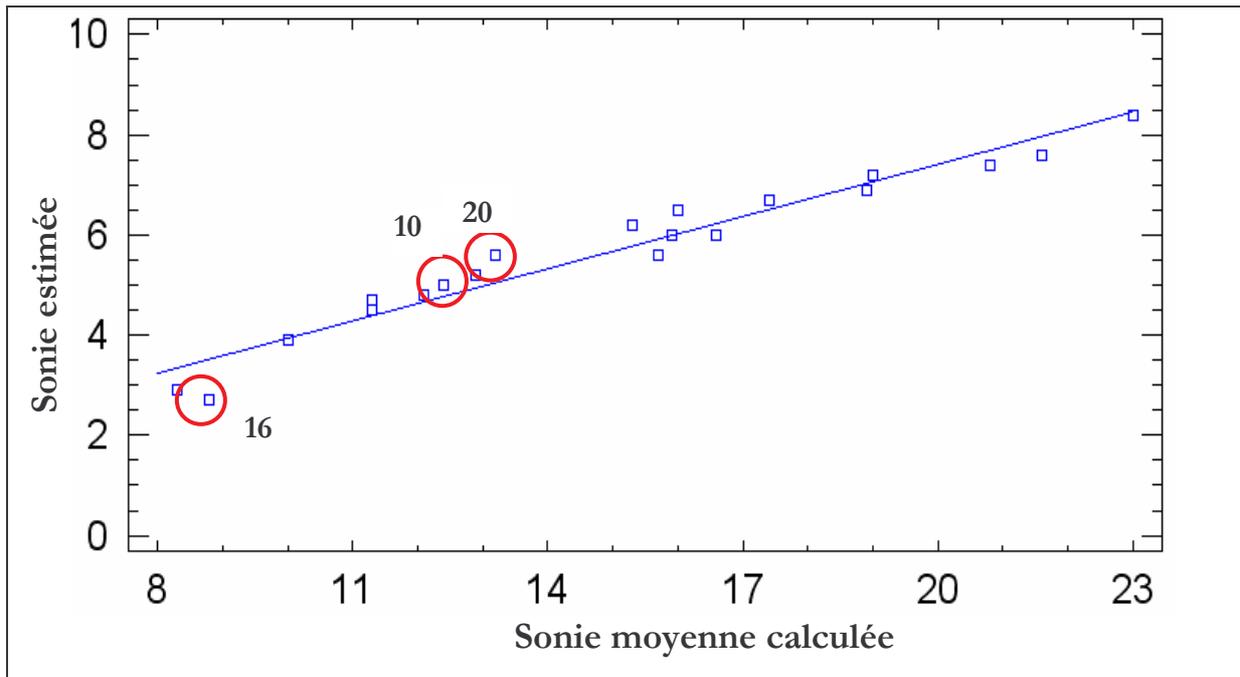


Figure 29 : Modèle linéaire (avec constante) de la force sonore estimée en fonction de la sonie moyenne calculée pour les vingt stimuli ($R^2 = 95,0\%$), (vingt stimuli et moyenne sur vingt-six sujets.)

2.2.5. Relation entre le désagrément et la force sonore subjective

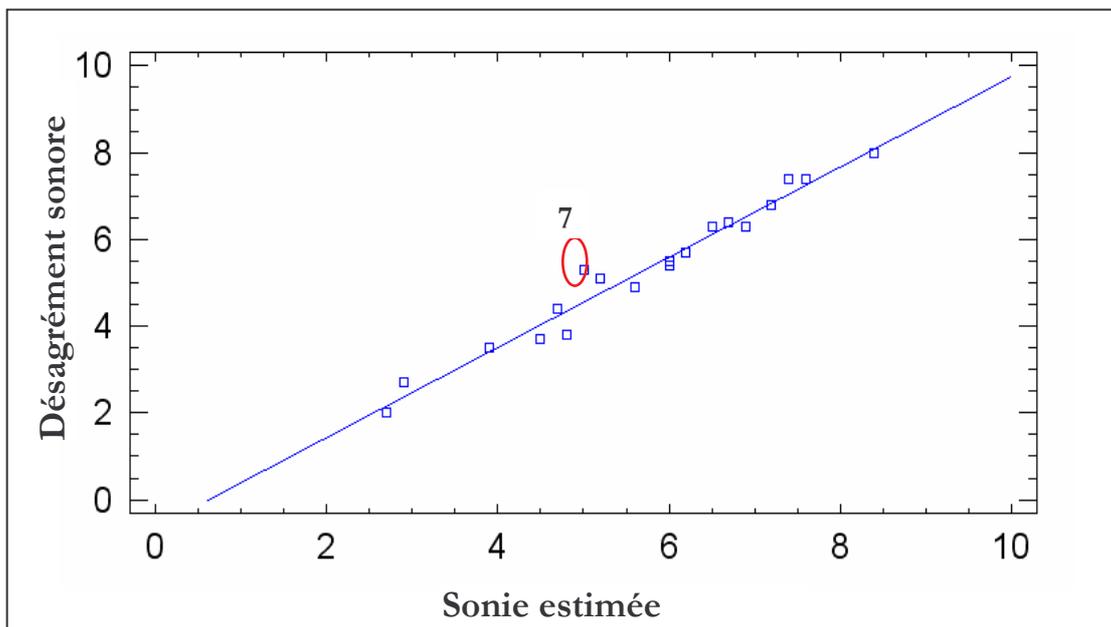


Figure 30 : Modèle linéaire (avec constante) du désagrément sonore estimée en fonction de la sonie estimée ($R^2 = 96,5\%$) (vingt stimuli et moyenne sur vingt-six sujets). Le stimulus sept se détache légèrement du modèle.

Le désagrément sonore est très fortement corrélé à la sonie [Figure 30]. Seul le stimulus numéro sept se détache légèrement du modèle. Comme nous l'avons déjà remarqué, ce stimulus contient deux cyclomoteurs. Par rapport à l'allure du modèle, le désagrément de ce stimulus est donc sous évalué par la mesure de la sonie. Ainsi, on peut suggérer, en s'appuyant sur la tendance générale de la relation présentée, que sur ce stimulus, l'effet négatif du cyclomoteur n'est pas entièrement pris en compte par la sonie estimée.

2.3. Force et désagrément sonore pour le test réalisé in-situ

2.3.1. Description du test

En complément des expériences de laboratoire, nous avons réalisé un travail similaire mais cette fois-ci in-situ et dans trois ambiances sonores distinctes [CAN-05]. Nous avons repris le même questionnaire utilisé dans ce mémoire lors d'un parcours réalisé au mois d'avril 2005. Il a ainsi été demandé aux auditeurs d'évaluer leur perception de la prégnance, de la présence, de la proximité des sources, de la force sonore ainsi que du désagrément sonore. Par rapport au test de laboratoire, nous avons étendu l'échelle du désagrément en tenant compte de l'aspect agréable de l'ambiance. En effet, les ambiances de type *parc* ou *marché* sont généralement qualifiées d'agréables [MAFIOLLO-99]. Ainsi, l'aspect agréable ou désagréable de l'environnement sonore a été évalué sur une échelle allant de « 0 » pour qualifier une ambiance ressentie comme « très désagréable » à « 10 » pour qualifier une ambiance qualifiée de « très agréable ». Une feuille d'explications type, fournie aux sujets avant le test est visible en annexe 4c.

Ce parcours consistait en une promenade dans Paris, au cours de laquelle les sujets ont été amenés à traverser, dans le même parcours, trois lieux représentant trois ambiances sonores distinctes :

- Un parc, (le parc Montsouris dans le 14^{ème} arrondissement),
- Une rue moyennement circulée, (La rue Reille dans le 14^{ème} arrondissement),
- Un marché, (Rue de la Glacière dans le 13^{ème} arrondissement).

Les cinquante neuf sujets (vingt-deux femmes et trente-sept hommes d'une moyenne d'âge de vingt-quatre ans) ont été répartis en deux groupes, effectuant le parcours en sens opposé :

- Un groupe se déplaçant du marché vers le parc (sens MP),
- Un groupe se déplaçant du parc vers le marché (sens PM).

Le parcours a été divisé en seize séquences, d'environ 1 min 30 s chacune donc légèrement plus longue que les 15 s des stimuli du travail en laboratoire (mais nous restons toujours dans le cadre d'évaluation à court terme). Chacun des deux groupes a été divisé en trois groupes homogènes, qui ont effectué le même parcours et ont donc entendu les mêmes scènes sonores :

- Un groupe notant chaque séquence (c'est-à-dire toutes les 1 min 30 s)
- Un groupe notant chaque ambiance (c'est-à-dire environ toutes les 15 min),
- Un groupe notant la globalité du parcours (c'est-à-dire à la fin).

La figure 31 illustre le parcours réalisé.

La répartition des sujets expliquée ci-dessus aboutit à la formation de six groupes (Tableau 16 b) :

	Sens de parcours	Type d'observation	Nombre de notation	Fréquence de notation
Groupe 1	marché-rue-parc	Séquences	16	1mn30
Groupe 2	marché-rue-parc	Ambiances	3	15mn
Groupe 3	marché-rue-parc	Global	1	45mn
Groupe 4	parc-rue-marché	Séquences	16	1mn30
Groupe 5	parc-rue-marché	Ambiances	3	15mn
Groupe 6	parc-rue-marché	Global	1	45mn

Tableau 16 b : Répartition par groupe des cinquante neuf sujets ayant participé au test.

Il a de plus été demandé, à la fin du parcours, aux sujets observant les séquences (groupes 1 et 4) et aux sujets observant les ambiances (groupes 2 et 5), de juger la globalité de la balade (soit l'équivalent des groupes 3 et 6).

Les sujets ont eu quatre séquences d'entraînement grâce auxquelles ils ont pu se familiariser avec les échelles et les notions demandées.

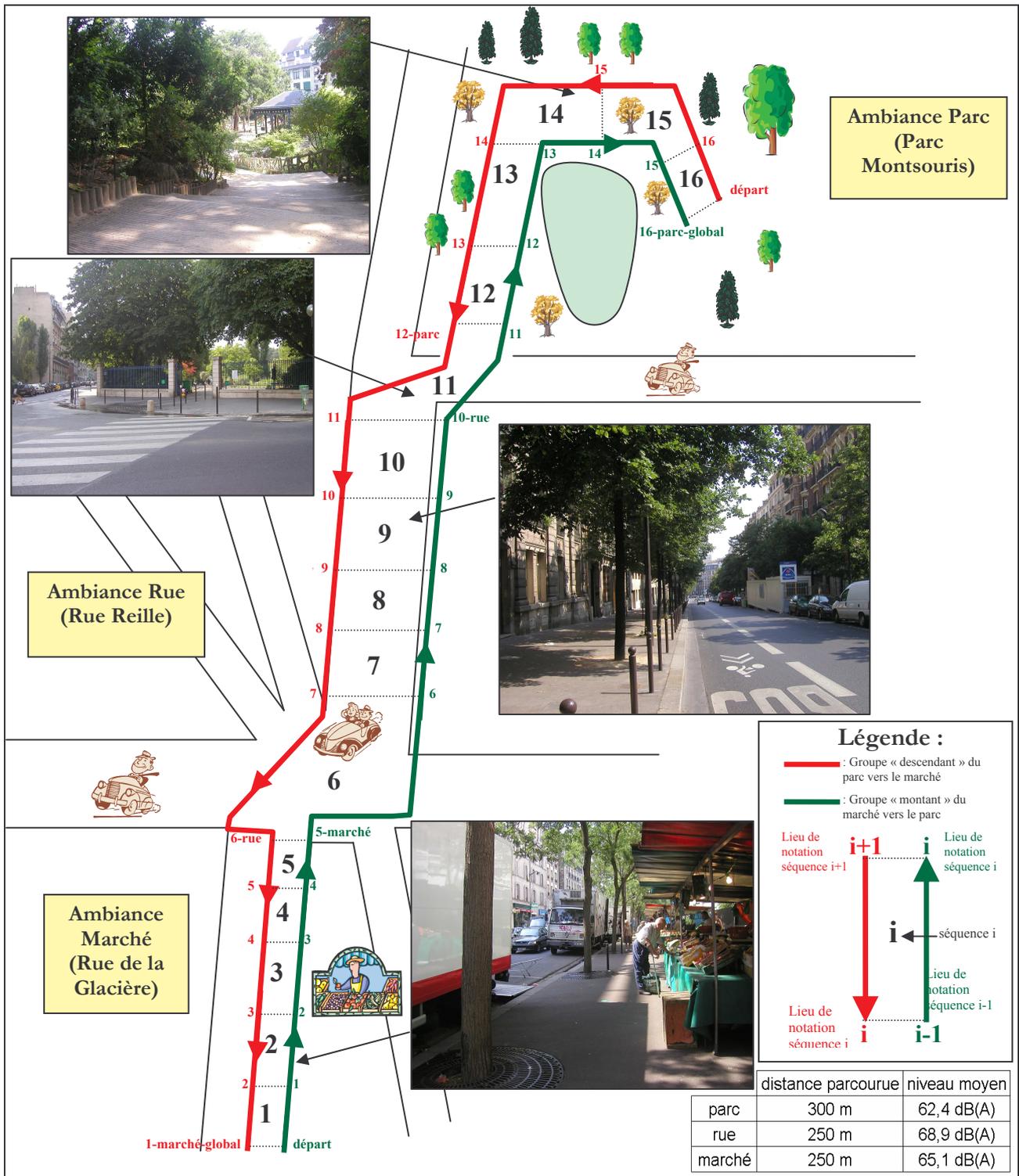


Figure 31 : Schéma du parcours du test in situ représentant les trois ambiances sonores.

Le but du travail réalisé est de comparer les résultats obtenus lors d'une expérience réalisée in-situ (donc dans des conditions naturelles où le critère esthétique faisant appel à des facteurs cognitifs et émotionnels entre en jeu [BLAUERT-97]) à ceux obtenus en laboratoire. L'une des finalités étant l'extraction de modèles de désagrément sous forme de régression linéaire multiple.

Notons que c'est également dans la rue Reille, qui représente ici l'ambiance « rue circulée », qu'ont été extraites quatorze des vingt séquences des tests effectués en laboratoire.

Dans la suite de ce mémoire nous ferons, lorsqu'il est opportun, référence aux travaux de cette recherche pour en comparer les résultats avec les tests réalisés en laboratoire. Ainsi, le paragraphe 2.3.2. étudie les relations entre force sonore et désagrément pour les trois ambiances. Le paragraphe 4.3 de ce chapitre est, quant à lui, entièrement dédié aux observations faites durant ce test sur la manière dont l'information sonore est agrégée par les sujets sur une période plus longue. Enfin, dans le paragraphe 3, consacré aux modèles, nous montrerons quelles sources sonores ont eu une importance lors du parcours perceptif.

2.3.2. Force et désagrément sonore

2.3.2.1. Réponses des sujets

Le graphique ci-dessous indique les réponses moyennées (avec les écarts types) des sujets ayant noté toutes les 1min 30s (Groupes 1 et 4, composés de dix personnes chacun) le caractère agréable ou désagréable des séquences [Figure 32].

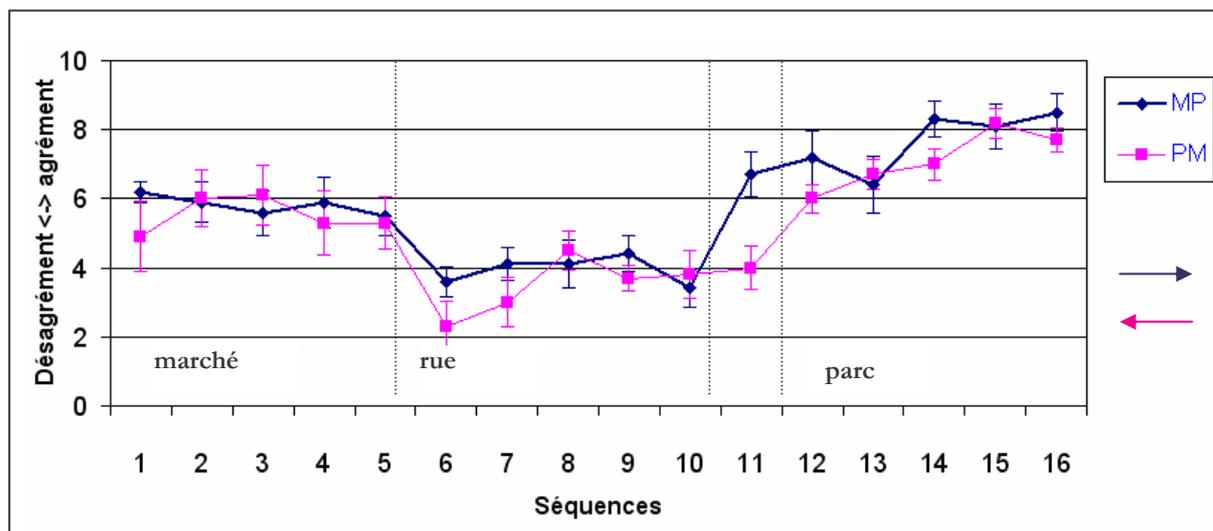


Figure 32 : Caractère agréable ou désagréable des séquences entendues pour les groupes 1 et 4 (dix sujets par groupe). La note « 0 » signifie une ambiance « très désagréable », une note « 10 » signifie une ambiance « très agréable ».

Les réponses sont cohérentes entre personnes du même groupe. En effet, les écarts types sont compris entre « 1 » et « 1,5 ». La figure précédente montre que les sujets ont trouvé que l'ambiance la plus agréable était celle du parc et que l'ambiance la plus désagréable était celle de la rue. L'ambiance du marché a été qualifiée d'agréable.

De plus, les estimations comparées des deux groupes sont cohérentes. En effet, les deux courbes ont la même allure.

Le graphique ci-dessous nous traduit l'intensité sonore perçue pour les mêmes séquences [Figure 33].

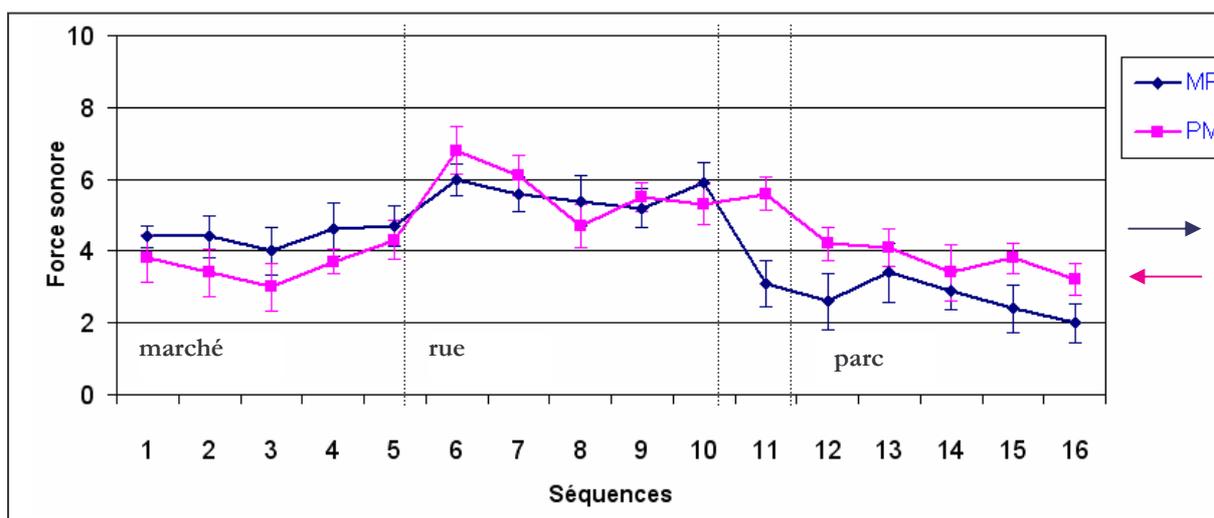


Figure 33 : Intensité sonore perçue pour les groupes 1 et 4 (dix sujets par groupe).

Nous pouvons formuler les mêmes remarques que celle émises concernant le désagrément sonore.

En analysant les deux figures précédentes, nous constatons que seules les séquences de transition sont perçues différemment. Elles dépendent en effet de l'endroit duquel arrivent les auditeurs. La séquence six, marquant l'entrée dans le rue, et la séquence onze, à cheval sur la rue et le parc, sont perçues de manière plus positive par le groupe se déplaçant vers le parc. Il y a probablement eu une influence due au changement d'ambiance sonore sur le caractère hédonique des séquences du parcours.

2.3.2.2. Force sonore estimée et désagrément sonore

La figure 34 représente les graphiques de la force sonore en fonction du coté agréable ou désagréable pour chacune des trois ambiances ainsi que pour les trois ambiances réunies.

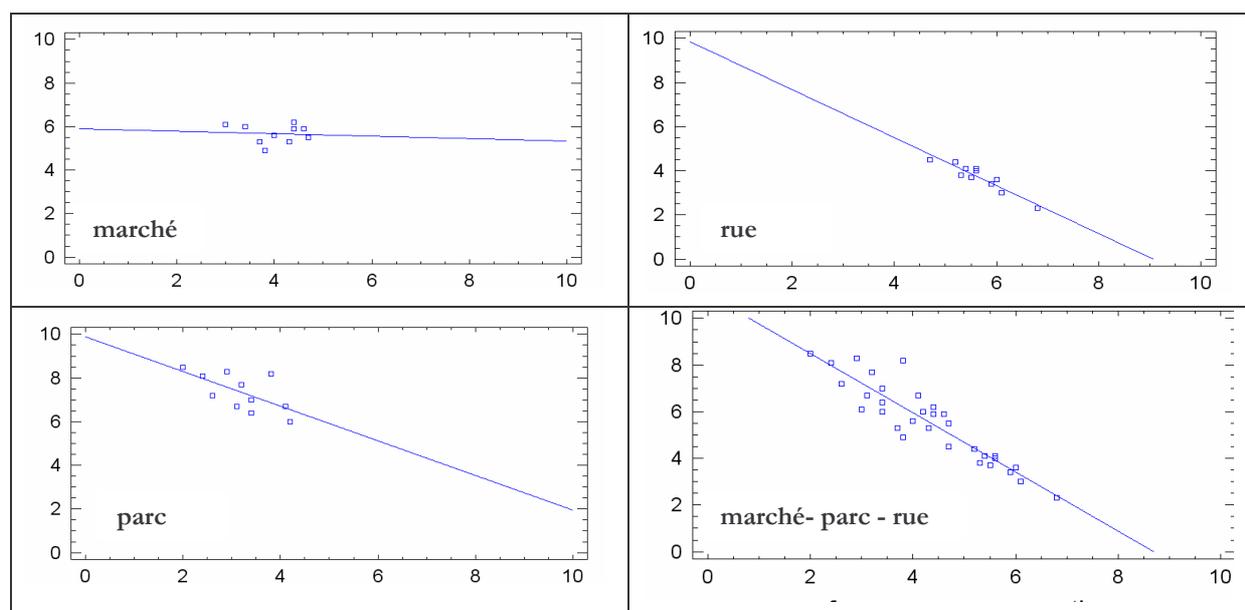


Figure 34 : Relations entre le désagrément sonore (en ordonnée) et la force sonore estimée (en abscisse) pour chacune des trois ambiances séparément puis pour les trois ambiances considérées ensemble. Vingt sujets ont pris part aux tests. La note « 0 » signifie une ambiance « très désagréable », une note « 10 » signifie une ambiance « très agréable ».

Nous observons que,

- Dans l'ambiance de type marché, la force sonore et le désagrément ne sont pas du tout corrélés avec un $R^2_{aj} = 0.0\%$
- Dans l'ambiance de type rue, la force sonore et le désagrément sont bien corrélés avec un $R^2_{aj} = 85.2\%$
- Dans l'ambiance de type parc, la force sonore et le désagrément sont faiblement corrélés avec un $R^2_{aj} = 34.4\%$

Pour les trois ambiances, les variabilités des notes de désagrément, comme celles des niveaux sonores estimés, sont sensiblement les mêmes. Les notes de désagrément (respectivement de force sonore) varient de « 2.3 » à « 4.5 » (de « 4.7 » à « 6.8 ») pour l'ambiance rue, de « 6.0 » à « 8.5 » (de « 2.0 » à « 4.2 ») pour l'ambiance parc et de « 3.0 » à « 4.4 » (« 4.9 » à « 6.2 ») pour l'ambiance marché.

Ces expériences montrent que le niveau sonore perçu n'est pas toujours bien adapté pour suivre l'évolution de la qualité sonore au sein d'une même ambiance. En effet, le niveau sonore n'est correctement corrélé qu'à l'ambiance rue. Celle-ci a été la seule qualifiée de désagréable.

Ces résultats sont à rapprocher d'une étude réalisée par Brigitta Berglund [BERGLUND-04] qui montre l'insuffisance de l'emploi de la force sonore pour caractériser le côté agréable ou désagréable des endroits qualifiés de calmes. Cependant, le terme de « calme » employé pour désigner des parcs ou des marchés n'est pas adapté ici. En effet, les niveaux sonores mesurés ne présentent pas un écart important. Le niveau moyen mesuré lors du test dans le parc était $L_{Aeq} = 62.4$ dB(A), sur le marché $L_{Aeq} = 65.1$ dB(A) et dans la rue $L_{Aeq} = 68.9$ dB(A).

Enfin, comme nous le montre la dernière régression de la figure 34, le niveau sonore perçu permet de comparer la qualité sonore d'ambiances sonores différentes ($R^2_{aj} = 86.6\%$).

Il semble que, plus les sources ont un sens élevé pour les auditeurs (les voix sont par exemple reconnues comme étant porteuses d'informations), moins la mesure de la sonie permet d'expliquer le désagrément sonore.

La confrontation faite sur observations obtenues lors des tests réalisés en laboratoires et in situ renforcent notre hypothèse de départ d'une définition de la qualité d'un environnement sonore qui stipule qu'elle est dépendante de sa composition. Dans ce sens, la qualité sonore constitue un jugement perceptif résultant d'une évaluation de la représentation mentale de la source d'une part et de l'adéquation entre des sources sonores et leur lieu d'occurrence d'autre part.

III. Etablissement d'un modèle de désagrément

3.1. Introduction

Ce chapitre est la partie « appliquée » de nos recherches. Les analyses factorielles entreprises précédemment ont permis de dégager des tendances concernant la sensibilité des auditeurs à la perception de certaines sources sonores dans un environnement urbain. Pour un bureau d'étude acoustique, il est important de compléter ces travaux théoriques par la recherche d'un modèle prévisionnel. Le cycle logique de la méthode scientifique de recherche peut se décomposer en trois phases, la *formulation de l'hypothèse*, la *vérification de l'hypothèse* et la *validation de l'hypothèse*. Si nous nous tenons à cette démarche générale, l'hypothèse, selon laquelle la prise en compte des sources sonores par un auditeur pour l'appréciation hédonique d'une ambiance, a tout d'abord été formulée en introduction. Puis, dans le chapitre suivant, la formulation s'est faite plus précise grâce à la définition d'un espace subjectif tridimensionnel. Sur cet espace sont projetées les appréciations auditives des auditeurs sur les catégories de sources prédéfinies. Ce nouveau chapitre constitue donc en quelque sorte une vérification de l'hypothèse.

Pour cela, nous utilisons les vingt séquences tests et leur recombinaison en une séquence globale de 5 min. Les jugements recueillis au travers de ces deux tests vont être ici utiles pour établir deux modèles de désagrément sonore. Le type de modèle choisi est la régression linéaire multiple.

Ce choix du type de modèle s'est effectué sur deux points essentiels. Tout d'abord c'est un classificateur couramment utilisé dans la littérature pour effectuer des prédictions, mais aussi car nous disposons des moyens assez « simples » et informatisés pour vérifier la pertinence des modèles obtenus. Le premier modèle, dit subjectif, pourra intégrer comme variables les jugements de prégnance, de présence et de proximité obtenus précédemment. Nous avons également besoin de l'intensité (ou force sonore) car de nombreux travaux ont montré que le désagrément était lié à la force sonore. Le second, dit objectif, pourra intégrer comme variables les mesures ayant obtenu une bonne correspondance avec les dimensions subjectives.

3.2. Les modèles et leur validité

3.2.1. Le modèle de régression linéaire

L'analyse de régression est un outil permettant de mettre en forme un problème caractérisé par un déterminisme non absolu. L'analyse de régression simple ou multiple, qui peut être définie comme la recherche de la relation stochastique qui lie plusieurs variables, est donc une méthode adaptée à notre problématique. Toutes les correspondances obtenues entre variables sont observables selon des relations linéaires.

Les descriptions et explications fournies dans la suite de ce mémoire et concernant les régressions (modèles, estimations, intervalles, tests...) sont largement inspirées de l'ouvrage «Analyse de Régression Avancée» [DODGE-04]. Ce sont des techniques couramment utilisées de nos jours et dont la littérature explicative est très abondante. De ce fait nous ne reprenons pas en détails les démonstrations mathématiques s'y rapportant. Seuls les principes fondamentaux qui ont fait le succès et la robustesse de ces types de modèles sont expliqués.

3.2.2. Méthode d'estimation des modèles : la méthode des moindres carrés

Les modèles obtenus expriment des relations linéaires entre les différentes observations. Deux types de modèles seront utilisés par la suite. Dans le cas où il y a une seule variable explicative dans le modèle, nous parlons d'un modèle de régression simple. Ce modèle mathématique décrivant une droite est appelé *droite de régression* et représente la description de la relation linéaire. Dans le cas où le modèle nécessite plusieurs variables explicatives, nous parlons de régression linéaire multiple. L'équation décrivant la droite de régression est alors remplacée par l'équation d'un hyperplan appelé *hyperplan de régression*. Dans tous les cas, la fonction première de cette équation ou « formule » mathématique est d'être la plus représentative possible de l'expérience. L'écart individuel entre une donnée observable et la prédiction (ou valeur estimée) du modèle est appelé *résidu*. La méthode d'estimation retenue dans notre mémoire (qui s'avère être la plus connue) pour obtenir le meilleur modèle consiste à minimiser les écarts entre les résidus ou plus exactement la somme des carrés des résidus. Cette méthode est appelée la *méthode des moindres carrés*.

Un développement mathématique montre que le problème établi (minimisation des résidus) revient également à poser une équation où la moyenne des valeurs observées est égale à la moyenne des valeurs prévues. De ce fait, la somme des résidus est nulle.

3.2.3. Coefficient de détermination (R^2) et R^2 ajusté (R^2_{aj})

Le critère objectif permettant d'évaluer la pertinence d'un modèle se définit comme étant un pourcentage de la variation totale qui est expliquée par le modèle. On définit ainsi le *coefficient de détermination* noté R^2 . Il représente le rapport entre la variation expliquée par le modèle et la variation totale.

Le R^2 ajusté (au nombre de variables), noté R^2_{aj} est également un estimateur de la pertinence d'un modèle. Son calcul se fait selon l'équation :

$$R^2_{aj} = \frac{(n-1).R^2 - (p-1)}{n-p}$$

p = nombre de variables explicatives,

n = nombre d'observations,

R^2 = coefficient de détermination = pourcentage de variation des résultats expliqué par le modèle de régression,

R^2_{aj} est donc fonction du R^2 (mais toujours inférieur) à la différence qu'il n'augmente pas forcément (et peut même régresser) lors de l'introduction de variables supplémentaires dans un modèle. En effet, lorsque « p » tend vers « n », R^2_{aj} tend vers « 0 », alors que R^2 tend vers « 1 ».

Dès lors, il est possible de comparer la performance de modèles n'ayant pas le même nombre de variables explicatives. Le R^2_{aj} est la statistique retenue dans ce mémoire pour nous permettre de sélectionner nos modèles. Nous exprimons les valeurs du R^2_{aj} directement en pourcentage.

3.2.4. Coefficient de corrélation

Bien que très peu employée par la suite (car très proche du R^2), la mesure de corrélation utilisée est le *coefficient de corrélation* de Bravais-Pearson. Cet indice numérique mesure le degré de liaison ou l'intensité de la relation entre deux variables. Ici, la première variable est représentée par l'ensemble des estimations du modèle, la seconde par l'ensemble des observations du désagrément sonore.

Le coefficient de détermination, décrit précédemment, noté R^2 , est le carré du coefficient de corrélation d'un modèle (avec constante) noté r . De ce fait, il peut prendre des valeurs négatives et a donc l'avantage d'indiquer le type de relation linéaire existant.

3.2.5. Test de validité des modèles

Pour tout modèle extrait et ayant une bonne description du phénomène étudié, il existe des « tests » permettant de s'assurer de leur validité ou de leur pertinence. L'« information » importante à tester porte sur la signification statistique des variables explicatives présentes dans le modèle.

Ainsi, nous vérifions tout d'abord les liens entre les variables indépendantes et la variable dépendante avec *l'intervalle de confiance*. Puis, nous traitons ensuite des relations des variables indépendantes entre elle, d'une part avec l'analyse graphique des résidus, d'autre part avec le *test de Durbin-Watson*.

3.2.6. Intervalle de confiance sur les paramètres du modèle (tests d'hypothèses)

La raison d'être d'un modèle est de pouvoir prédire le désagrément pour des situations futures. Or les modèles obtenus dans ce mémoire ont été extraits à partir d'un nombre réduit de stimuli qui constituent notre échantillon. Ainsi, il y a lieu d'appliquer un test pour optimiser leur prédictibilité dans des situations futures et attester que les résultats obtenus à partir de notre échantillon sont généralisables à une autre population. Ainsi, pour chaque modèle établi, nous testons si les variables explicatives y sont significatives, c'est-à-dire si la présence d'une variable n'est pas due au « hasard » (par exemple des variations aléatoires dans les réponses des sujets). Les fondements de ces tests reposent sur des tests d'hypothèses qui sont détaillés dans l'ouvrage de Dodge [DODGE-04].

Ces tests permettent également de s'assurer que l'on ne puisse pas considérer un modèle plus simple. L'intervalle de confiance pour les tests d'hypothèses est de 95%, ceci pour chaque variable du modèle. Ces tests statistiques ne constituent pas une « validité » absolue des modèles mais représentent plutôt une probabilité signifiant une minimisation des possibilités d'erreurs. Cependant, ces tests d'hypothèses supposent une distribution parfaitement normale (ou gaussienne) des variables du modèle.

3.2.7. Analyse graphique des résidus.

Pour chaque modèle présenté dans le mémoire, nous avons visualisé le graphique des résidus en fonction des valeurs estimées. Les modèles vérifient tous la condition suivante : les résidus sont répartis uniformément dans une bande horizontale du graphique. Cela montre à la fois que le modèle linéaire est adéquat pour décrire notre problème, que la variance est sensiblement la

même quelle que soit l'observation et enfin qu'il n'y a pas de valeur aberrante parmi les observations.

3.2.8. Test de Durbin-Watson.

Le test de Durbin-Watson [DURBIN-51] est réalisé sur le coefficient de corrélation du modèle. Il stipule l'absence de *corrélation sérielle* (ou *autocorrélation*) entre les résidus, c'est-à-dire qu'il vérifie l'indépendance des résidus. Par exemple, si dans un modèle nous avons le nombre de cyclomoteurs et la sonie, il teste alors le fait qu'une augmentation de la sonie entraîne de manière systématique une augmentation du nombre de cyclomoteurs ou inversement, auquel cas l'une des deux variables est inutile.

3.3. Le désagrément sonore

Le problème que nous avons à traiter est d'établir un lien entre une variable dite *dépendante* ou *expliquée*, qui est le désagrément sonore, au moyen d'une quantité relativement importante de variables dites *indépendantes* ou *explicatives*. Les variables explicatives sont des observations qui sont soit subjectives (donc obtenues lors des tests psychoacoustiques) soit objectives (obtenues après analyses physiques du signal sonore). Nous avons, pour chaque variable, vingt observations différentes ou individus (car nous avons utilisé vingt stimuli différents). Pour les observations subjectives ainsi que pour la variable descriptive, les vingt observations sont des moyennes des réponses des vingt sujets alors que les variables objectives sont uniques. Elles ne sont cependant pas absolues car elles sont le résultat d'un codage temporel qui a fait appel à l'écoute et à ses contraintes.

3.3.1. Données des modèles

3.3.1.1. Variables de départ :

Le tableau [Tableau 17a] expose les variables explicatives (relatives aux sources sonores) à tester qui servent à construire les modèles de désagrément sonore.

		Variables explicatives	
		Modèle objectif	Modèle subjectif
Voitures		L_{Aeq}	Prégnance Présence Proximité
		L_{A10}	
		L_{A5}	
		L_{AMax}	
		%T	
		Nombre	
Bus		L_{Aeq}	Prégnance Présence Proximité
		L_{A10}	
		L_{A5}	
		L_{AMax}	
		%T	
		Nombre	
Cyclomoteurs		L_{Aeq}	Prégnance Présence Proximité
		L_{A10}	
		L_{A5}	
		L_{AMax}	
		%T	
		Nombre	
Motos		L_{Aeq}	Prégnance Présence Proximité
		L_{A10}	
		L_{A5}	
		L_{AMax}	
		%T	
		Nombre	
Oiseaux		L_{Aeq}	Prégnance Présence Proximité
		L_{A10}	
		L_{A5}	
		L_{AMax}	
		%T	
Voix d'enfants		L_{Aeq}	Prégnance Présence Proximité
		L_{A10}	
		L_{A5}	
		L_{AMax}	
		%T	
Voix d'adultes		L_{Aeq}	Prégnance Présence Proximité
		L_{A10}	
		L_{A5}	
		L_{AMax}	
		%T	
Total de variables		39	21

Tableau 17a : Liste exhaustive des variables explicatives relatives aux sources susceptibles de rentrer dans les modèles objectifs et subjectifs.

Notre population se compose de vingt individus qui sont les vingt stimuli du test perceptif. Comme nous l'avons déjà souligné, le but de ce travail est d'obtenir une double caractérisation du désagrément sonore. L'une perceptive dite subjective et l'autre physique dite objective.

Le tableau 17b indique les autres variables explicatives qui sont utilisées pour construire les modèles. Ce sont des variables énergétiques calculées sur la durée totale d'un stimulus.

Sonie calculée (Zwicker)	Sonie 10%	L_{Aeq} moyen	L_{Aeq} 10%
--------------------------	-----------	-----------------	---------------

Tableau 17b : Liste des variables explicatives relatives calculées sur la durée totale d'un stimulus.

3.3.1.2. Variables objectives supplémentaires

D'autres critères, habituellement utilisés dans la littérature, ont été calculés pour être « testés » dans les modèles extraits. Ces critères, calculés avec le logiciel dBFa32 sur les modèles de Zwicker [ZWICKER-00], sont les suivants :

La force de fluctuation, l'acuité, la rugosité et l'Unbiased Annoyance ou « Gêne non biaisée » qui est une combinaison de critères psychoacoustiques.

3.3.2. Technique

Il existe plusieurs procédures de sélection de variables pour arriver à un modèle satisfaisant. Ici, l'idée est de tester un maximum de configurations possibles et d'obtenir un compromis satisfaisant entre un maximum de variance des résultats expliqués et un nombre minimum de variables présentes dans le modèle.

Notons que la technique habituellement utilisée pour extraire des modèles est la méthode *stepwise*. Cette procédure considère un modèle faisant intervenir un certain nombre de variables explicatives. On procède ensuite par élimination ou ajout successif de variables. Cependant, cette méthode ne restitue pas forcément le meilleur modèle parmi les 2^k (k = nombre de variables) modèles possibles car certains modèles ne sont pas examinés. Le modèle final dépend de l'ordre dans lequel les variables ont été introduites et des phénomènes de *multi-colinéarité* peuvent masquer l'extraction d'un modèle significatif [DODGE-04].

Une technique possible aurait été de réaliser une recherche exhaustive (avec cinq variables maximum par exemple), c'est-à-dire une recherche de tous les modèles possibles en rentrant dès le départ toutes les variables, et puis d'en extraire les meilleurs et retenir ceux qui passent les « tests » de vérifications. Cependant nous avons plus de variables explicatives que de stimuli. En effet, pour une population de vingt individus nous avons d'un côté trente-neuf variables pour le

modèle objectif et vingt et une variables de l'autre côté pour le modèle subjectif. C'est donc techniquement impossible de lancer une telle recherche.

La procédure retenue est finalement la suivante :

Nous examinons les données sur deux groupes différents, c'est-à-dire que nous recherchons les meilleures corrélations entre le désagrément sonore et les variables de ces deux groupes, afin d'en retenir les meilleures pour composer le modèle final. Le premier groupe a trait à la force sonore (les variables sont les critères énergétiques globaux du tableau 17b), le second aux valeurs relatives aux sources sonores (les variables sont indiquées dans le tableau 17a). Dans ce deuxième groupe, formons également deux sous-groupes : un avec les données relatives aux véhicules, un autre avec les données relatives aux voix et aux oiseaux. Pour chacun de ces sous-groupes formés, nous effectuons une recherche exhaustive sur toutes les variables avec cinq variables maximum présentes dans les modèles. Ainsi nous obtenons une sélection de données pertinentes que nous réinjectons dans les modèles finaux.

Enfin, nous testons chacun des modèles avec les variables présélectionnées. Les modèles définitifs sont ceux qui ont le meilleur R^2_{aj} avec quatre variables au maximum et qui ont été validés par les tests décrits précédemment.

3.3.3. Modèles avec la force sonore comme variable

3.3.3.1. Modèle subjectif

Le modèle liant le désagrément à la sonie estimée ou force sonore (N_{es}) seule est très bien corrélée avec $R^2_{aj} = 96.3\%$.

Nous précisons ici que les deux tests (force sonore et désagrément) ont été réalisés séparément, de manière à ne pas entraîner de confusions entre les deux notions de la part des sujets. Cependant, nous voyons que les réponses sont tout de même très semblables. Cela provient probablement du fait que les stimuli sont assez similaires : ils proviennent de deux rues de même type (donc de la même ambiance sonore) ; les sources sont assez équivalentes (une majorité de véhicules, quelques voix et oiseaux).

En complément à cette variable, la procédure de sélection effectuée avec les données perceptives nous permet d'augmenter le pourcentage de variance des résultats des tests¹ (Tableau 17c).

Nombre variables	Variables du modèle subjectif				R ² aj
1	Sonie estimée				96.3%
	Prégnance VE				10.5%
	Présence bus				36.4%
	Présence mob				5.3%
2	Sonie estimée	Prégnance VE			97.2%
	Sonie estimée	Présence bus			97.1%
	Sonie estimée	Présence mob			93.2%
3	Sonie estimée	Prégnance VE	Présence bus		97.7%
	Sonie estimée	Présence bus	Présence mob		98,5%
	Sonie estimée	Prégnance VE	Présence mob		97.1%
4	Sonie estimée	Prégnance VE	Présence bus	Présence mob	98.7%

Tableau 17c : Modèles avec sonie estimée obtenus en fonction des variables subjectives

La variable la plus significative du groupe des voix et oiseaux est la prégnance des voix d'enfants ($PgVE$). Deux variables ressortent du groupe des véhicules : la présence des bus ($PsBus$) et la présence des cyclomoteurs ($PsMob$).

Ainsi, le modèle final du désagrément ($Dés$) inclut les quatre variables. Son R² ajusté est R²aj = 98.7%. Son équation est la suivante :

$$Dés = 0.04 + 0.86 \times Nes + 0.07 \times PsCyclo + 0.11 \times PsBus - 0.06 \times PgVE$$

Nous remarquons que les voitures, bien que très présentes dans les stimuli, n'apparaissent pas comme une variable dans le modèle (cf. paragraphe 3.4.4.).

La figure 35 présente schématiquement le modèle de régression.

¹ Notons que les valeurs des coefficients de déterminations ajustés du tableau 17c sans sonie sont biaisées par les stimuli ne contenant pas le véhicule en considération. Nous avons le même effet pour les figures 18b, 21a et 21b.

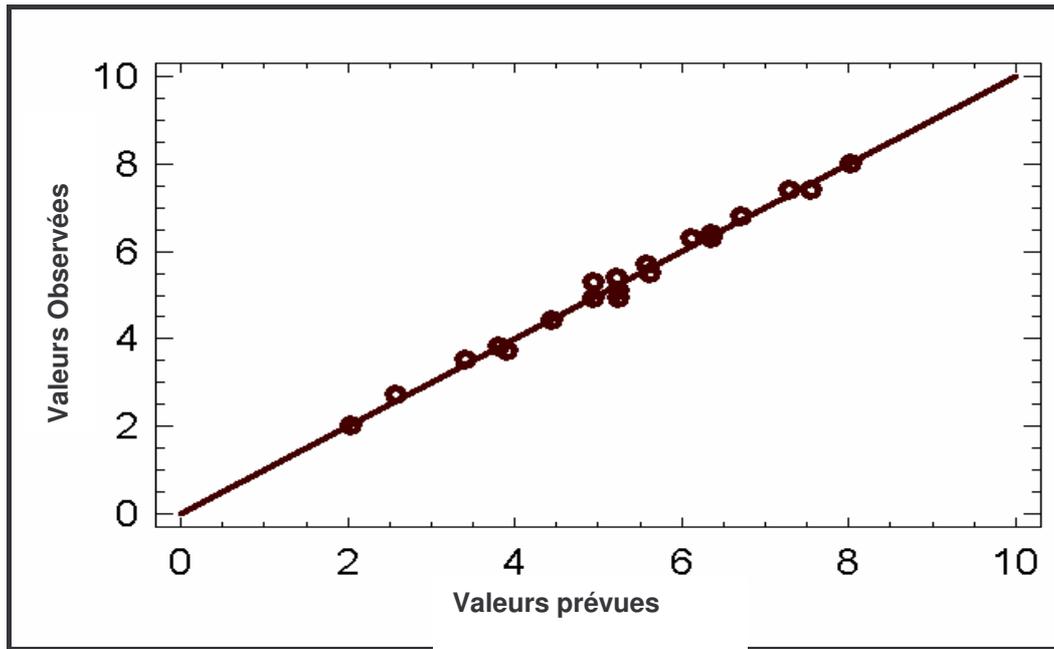


Figure 35 : Correspondance entre les valeurs du modèle (subjectif et avec sonie) de régression linéaire et les valeurs expérimentales. $R^2_{aj} = 98.7\%$.

3.3.3.2. Modèle objectif

La meilleure valeur rendant compte du désagrément parmi les critères calculés sur la durée totale (15 secondes) des stimuli est la sonie moyenne (calculée avec le modèle de Zwicker) N avec $R^2_{aj} = 90.6\%$. Le tableau 18a affiche les coefficients de détermination ajustés (R_{aj}^2) entre le désagrément et des valeurs physiques.

	Désagrément
Sonie calculée (Zwicker)	90.6%
Sonie 10%	80.3%
L_{Aeq} moyen	86.5%
L_{Aeq} 10%	84.1%

Tableau 18a : Tableau des coefficients de déterminations ajustés R^2_{aj} entre le désagrément sonore et des grandeurs physiques pour les vingt stimuli de quinze secondes. Toutes les corrélations ont une probabilité avec un niveau de confiance $p < 0.001$

En reprenant le même modus operandi que précédemment, nous pouvons améliorer de façon significative le pourcentage de variance expliquée. Trois variables sont ainsi extraites : le nombre de bus (NbBus), le nombre de cyclomoteurs (NbMob) ainsi que le pourcentage de temps de présence des voix d'enfants (%TVE).

Nombre variables	Variables du modèle objectif				R ² aj
	Sonie calculée				
1	Sonie calculée				90.6%
	Nombre bus				29.1%
	Nombre mob				6.7%
	%T VE				19.5%
2	Sonie calculée	Nombre bus			90.6%
	Sonie calculée	Nombre mob			90.5%
	Sonie calculée	%T VE			93.7%
3	Sonie calculée	Nombre bus	Nombre mob		91.7%
	Sonie calculée	Nombre bus	%T VE		93.7%
	Sonie calculée	Nombre mob	%T VE		93.7%
4	Sonie calculée	Nombre bus	Nombre mob	%T VE	94.7%

Figure 18b : Modèles avec sonie calculée obtenus en fonctions des variables objectives.

Le modèle final qui prend en compte la sonie moyenne et les trois variables précitées justifie d'un R²aj = 94.7% [Figure 46]. Son équation est la suivante :

$$Dès = 0.73 + 0.29 \times N + 0.29 \times NbMob + 0.46 \times NbBus - 0.02 \times \%TVE$$

La figure 36 présente schématiquement le modèle de régression.

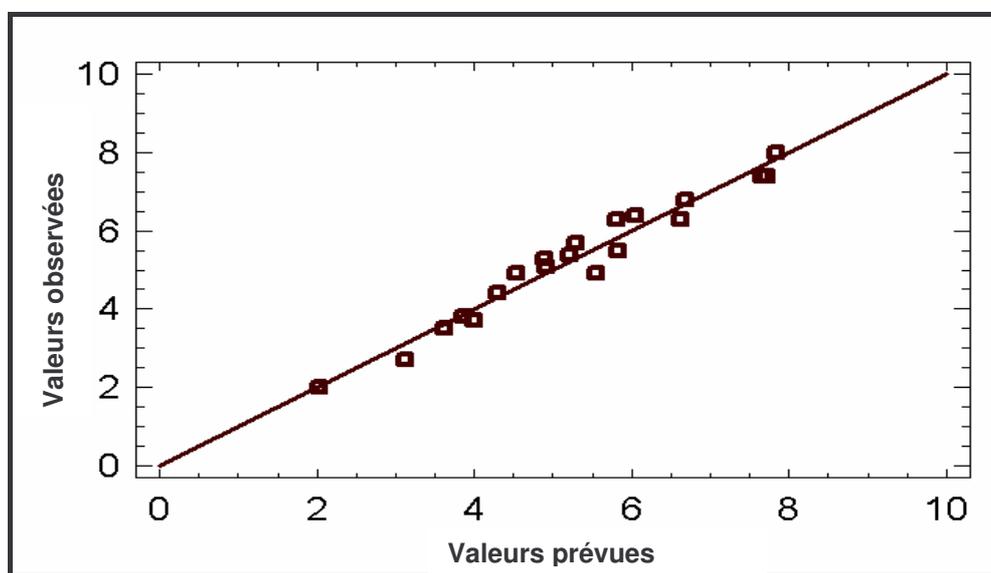


Figure 36 : Correspondance entre les valeurs du modèle (objectif et avec sonie) de régression linéaire et les valeurs expérimentales. R²aj = 94.7%.

3.3.3.3. Interprétations

Dans le premier modèle présenté, le rajout de valeurs subjectives à la sonie perçue fait passer le R^2_{aj} de 96.3% à 98.7%. De façon similaire, dans le second modèle présenté, l'ajout de valeurs objectives améliore le R^2_{aj} en le faisant passer de 90.6% à 94.7%. Le modèle de Zwicker calculant la sonie N ne prend évidemment pas en compte la représentation affective de la source, contrairement aux valeurs subjectives N_{ev} des sujets. En conséquence, il est prévisible que la valeur de l'ajustement fait à partir du modèle avec la sonie calculée selon le modèle de Zwicker soit inférieure à la valeur de l'ajustement fait avec la sonie estimée.

Les voitures n'apparaissent dans aucun des modèles, subjectif ou objectif. Leur influence est déjà prise dans l'évaluation de la force sonore (cf. paragraphe 3.4.4.).

Il est étonnant de ne pas voir apparaître l'influence des oiseaux dans les deux modèles présentés, alors qu'ils sont généralement reconnus comme ayant une influence positive sur l'environnement sonore [SCHAFER-77], [VIOLLON-00]. Bien que l'analyse faite au paragraphe 2.2.3. de ce chapitre et concernant les stimuli un, quatre, cinq et quinze, montre une possible influence des oiseaux (couplés à des voix) sur le caractère désagréable des séquences, sous un certain seuil d'émergence par rapport aux véhicules leur influence positive n'est pas statistiquement significative.

Les sons en rapport avec une activité humaine sont reconnus pour attirer l'attention car ils sont signes de vie [VIOLLON-00]. Pour les stimuli présentés, il semble que ce sont les voix d'enfants qui ont attiré particulièrement l'attention des auditeurs. Leur connotation positive, déjà suggérée dans les travaux de Maffiolo & al. [MAFFIOLO-00], est confirmée par le signe négatif devant leur coefficient dans les modèles objectifs et subjectifs.

De plus, on remarque que les coefficients de la régression sont d'autant plus élevés que l'objet sonore lui-même est plus émergent. En effet, les bus sont les sources qui ont les niveaux maximum L_{Amax} les plus élevés, suivi par les cyclomoteurs, les motos et les voix.

3.3.3.4. Critères psychoacoustiques

L'évolution des différents critères psychoacoustiques calculés sur les vingt stimuli est représentée sur la figure 37.

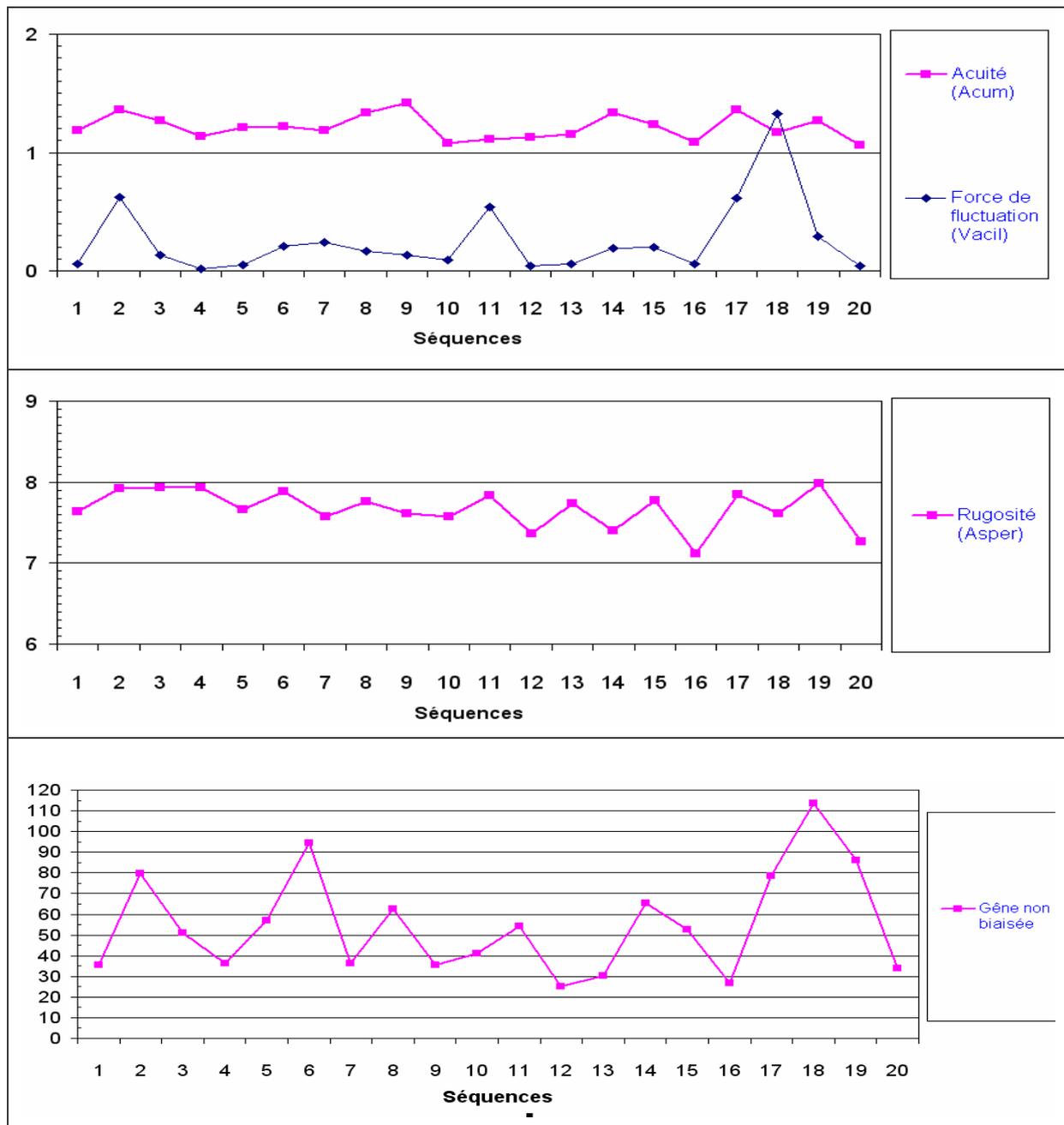


Figure 37 : Evolution des critères psychoacoustiques calculés sur les vingt stimuli.

Les valeurs de l'acuité et de la rugosité présentent peu de variabilité. Cela peut provenir du fait que les sons sont assez ressemblants car les bruits de moteur y sont prédominants.

La force de fluctuation présente une variabilité prononcée pour les stimuli 2, 11, 17 et 18. La composition en termes de source sonore [Tableau 2] de ces stimuli ne permet pas d'obtenir d'explications quant à l'augmentation constatée de ce critère.

La gêne non biaisée présente une variabilité engendrée par les valeurs de la sonie.

Pour évaluer si les critères psychoacoustiques calculés ont une influence sur la dispersion des valeurs du désagrément sonore, nous proposons de visualiser la figure suivante [Figure 38] :

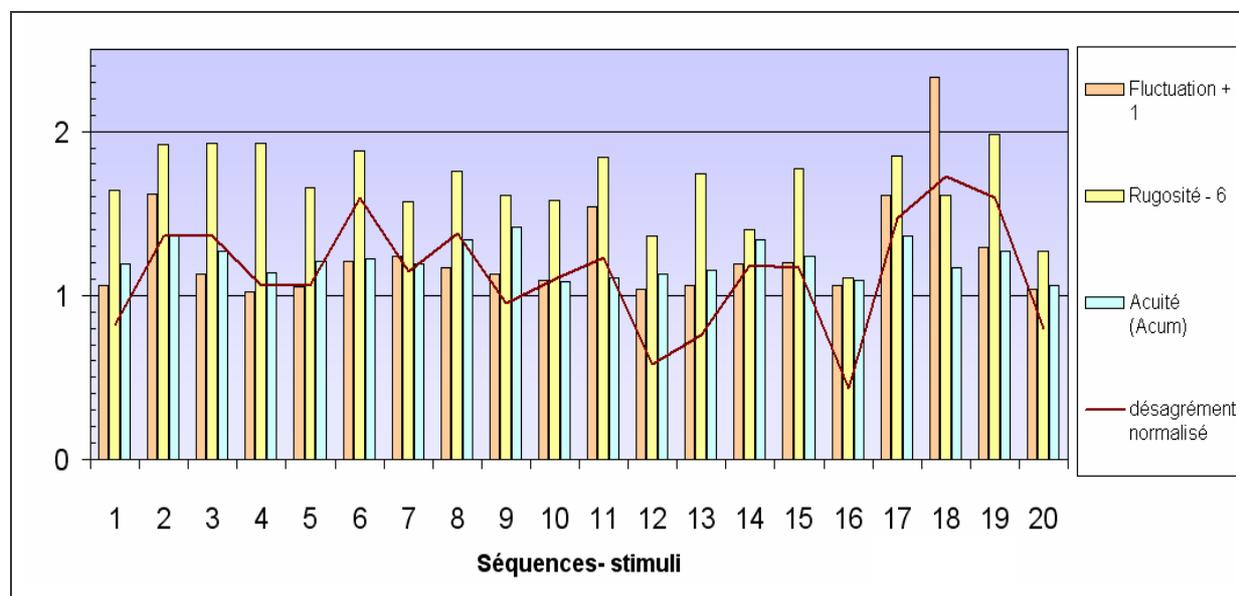


Figure 38 : Analyse de l'influence des critères psychoacoustiques sur les valeurs du désagrément sonore pour les vingt stimuli. Les valeurs des critères psychoacoustiques sont translatées pour être compris entre un et trois. Les valeurs du désagrément sont normalisées¹ pour être distribuées autour de un. On ne remarque pas de variation forte entre l'un des critères et le désagrément.

Cette figure ne montre pas d'évolution significative entre l'un des critères psychoacoustiques et le désagrément sonore.

Pour qualifier le caractère hédonique d'un environnement sonore, les critères psychoacoustiques ne semblent pas être bien adaptés à cause de leur faible variabilité comparée aux différences de désagrément perçu. Néanmoins, lorsqu'ils présentent des variabilités, nous avons constaté que celles-ci ne sont pas représentatives des évolutions du désagrément sonore.

Cette observation rejoint une étude précédente [LAVANDIER-02]. Les auteurs ont sélectionné plusieurs ambiances urbaines (une rue piétonne, une rue moyennement circulée, une rue fortement circulée, un parc, une place). Pour chacun des lieux et pour une durée de 24 h, ils ont calculé des critères psychoacoustiques toutes les 15 min, 30 min et 1 h. Les auteurs ne constatent qu'une faible variabilité de l'acuité, de la rugosité et de la force de fluctuation sur ces domaines temporels. Leur variabilité est également très faible d'un lieu à l'autre.

¹ La normalisation consiste à diviser chaque estimation par la moyenne géométrique de l'ensemble des notes de désagrément.

Dans une étude récente, Ellermeier & al. [ELLERMEIER-04] ont extrait un modèle de désagrément également sous forme de régression multiple linéaire. Sur des sons dits « techniques » comme un marteau piqueur, un camion, un train ou encore un mixeur, ils obtiennent un modèle ($R^2 = 98.0\%$) faisant entrer la rugosité et l'acuité comme variables indépendantes. Cependant, un de leur stimulus ne rentre pas dans leur modèle, c'est « de l'eau qui coule d'un robinet ». Cela tend à montrer que les modèles utilisant uniquement les propriétés physiques des sons peuvent fonctionner sur des ensembles de sons de la même classe perceptive. Mais les propriétés physiques des sons ne permettent pas d'évaluer le sens que les sources portent pour les auditeurs.

Le tableau 19 expose les coefficients de détermination ajustés R^2_{aj} entre le désagrément et les critères psychoacoustiques.

Variables	R^2_{aj}
Acuité	16.4%
Rugosité	44.5%
Force de fluctuation	38.4%
Gêne non biaisée	79.5%

Tableau 19 : Coefficients de détermination ajustés entre le désagrément et des critères psychoacoustiques.

Seul le critère Unbiased Annoyance, qui est un descripteur de qualité du son, explique près de 80% de la variance des réponses.

3.3.4. Modèles sans sonie

Nous avons vu que la sonie est la variable qui a le plus d'influence dans les deux modèles. De plus, aucune variable liée aux voitures n'apparaît alors que ce sont les sources les plus nombreuses. Une explication plausible à cette observation viendrait du fait que les voitures ne sont pas perçues comme des événements isolés à l'instar des cyclomoteurs ou des bus par exemple. Elles font partie intégrante de l'environnement sonore urbain et sont plutôt intégrées dans l'appréciation de l'ambiance globale. Ainsi, les jugements de sonie incluraient la perception des voitures passantes.

Comparaison avec l'étude in situ [CAN-05].

Pour l'ambiance de type rue de quartier, les voitures ne sont pas présentes de manière significative dans les modèles subjectifs. La force sonore est cependant un bon indicateur du désagrément. [Tableau 20a]

nombre variables	variables du modèle			R ² adj
1	force sonore			85,2%
	prégnance bus			50,5%
	proximité voitures			38,3%
2	force sonore	présence bus		89,1%
	force sonore	prégnance bus		86,0%
	présence bus	proximité voitures		51,7%
procédure stepwise ascendante :				
	force sonore	présence bus		89,1%
procédure stepwise descendante :				
	force sonore	présence bus		89,1%

Tableau 20a : Valeurs du R²aj pour différents modèles subjectifs dans l'ambiance de type rue de quartier.

Par contre, les voitures sont directement intégrées dans les modèles de désagrément pour les ambiances de type parc et de type marché [Tableau 20b]. En effet, alors que la force sonore seule ne représente plus un indicateur satisfaisant pour expliquer le désagrément, la perception de la présence des voitures semble la remplacer. La force sonore seule explique uniquement 34.4% de la variance des réponses pour l'ambiance de type parc. Toutefois, elle rentre comme variable explicative dans le meilleur modèle final. Pour l'ambiance de type marché, seule, elle n'explique aucune variabilité des réponses et n'est pas intégrée dans le modèle final. Par contre, on observe que la présence des voitures, de même que la proximité des voix sont les variables du modèle final.

Nous remarquons également que le niveau des voitures n'intervient pas dans les modèles. Pour les auditeurs, leur présence apparaît comme un critère plus significatif.

nombre variables	variables du modèle			R ² adj
1	force sonore			34,4%
	présence voitures			60,2%
	présence oiseaux			53,7%
2	présence oiseaux	force sonore		88,4%
	présence voitures	présence oiseaux		83,3%
3	présence voitures	présence oiseaux	force sonore	95,5%
procédure stepwise ascendante :				
	prégnance voitures	prégnance oiseaux	force sonore	94,4%
procédure stepwise descendante :				
	présence voitures	présence oiseaux	force sonore	95,5%

nb variables	variables du modèle			R ² adj
1	force sonore			0,0%
	présence voitures			47,4%
	proximité voix			25,5%
2	présence voitures	proximité voix		74,7%
3	présence voitures	proximité voix	présence voix	76,7%
	présence voitures	proximité voix	prégnance voix	80,0%
procédure stepwise ascendante :				
	présence voitures	proximité voix		74,7%
procédure stepwise descendante :				
	présence voitures	proximité voix		74,7%

Tableau 20b : Valeurs du R²aj pour différents modèles subjectifs dans l'ambiance parc (tableau du dessus) et l'ambiance de type marché ouvert.

Ainsi, dans un lieu autre qu'une route, les voitures deviennent aussi des « objets » et prennent un caractère événementiel.

En conséquence, il devient intéressant d'extraire un nouveau modèle qui ne prend pas la sonie comme variable pour tester si les voitures intègrent les modèles.

3.3.4.1. Modèle subjectif

Dans ce cas de figure, nous constatons que le modèle sans sonie fait ressortir comme variables tous les véhicules présents.

Nombre variables	Variables du modèle subjectif				R ² aj
1	Présence bus				36.4%
	Présence mob				5.3%
	Présence mot				0.0%
	Présence voit				0.0%
2	Présence bus	Présence mob			67.4%
	Présence bus	Présence mot			34.5%
	Présence bus	Présence voit			40.2%
3	Présence bus	Présence mob	Présence mot		66.2%
	Présence bus	Présence mob	Présence voit		80.1%
	Présence bus	Présence mot	Présence voit		44.0%
	Présence mob	Présence mot	Présence voit		0.0%
4	Présence bus	Présence mob	Présence mot	Présence voit	85.9%

Tableau 21a : Modèles sans sonie obtenus en fonction des variables subjectives.

Le meilleur modèle subjectif ainsi extrait donne un R²aj = 85.9%. Celui-ci prend en compte la présence de chaque catégorie de véhicule : présence voiture (*PsVoit*), présence Bus (*PsBus*), présence Moto (*PsMot*) et présence cyclomoteurs (*PsMob*) [Figure 21a].

Son équation est la suivante :

$$Dés = 1.50 + 0.27 \times PsVoit + 0.62 \times PsBus + 0.24 \times PsMot + 0.34 \times PsMob$$

La figure 39 présente schématiquement le modèle de régression.

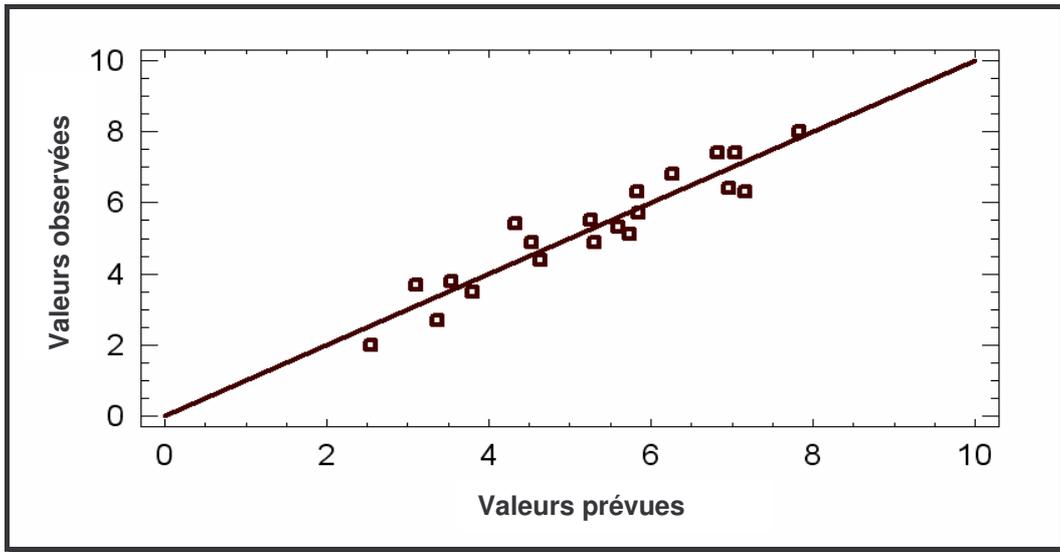


Figure 39 : Correspondance entre les valeurs du modèle (subjectif et sans sonie) de régression linéaire et les valeurs expérimentales. $R^2_{aj} = 85.9\%$.

3.3.4.2. Modèle objectif

Le meilleur modèle objectif fait ressortir le nombre de véhicules pour chaque catégorie : le nombre de voitures (NbVoit), le nombre de bus (NbBus), le nombre de motos (NbMot) et le nombre de cyclomoteurs (NbMob), pour un $R^2_{aj} = 81.2\%$ [Tableau 21b].

Nombre variables	Variables du modèle objectif				R^2_{aj}
1	Nombre bus				29.1%
	Nombre mob				6.7%
	Nombre mot				0.0%
	Nombre voit				0.0%
2	Nombre bus	Nombre mob			58.8%
	Nombre bus	Nombre mot			25.8%
	Nombre bus	Nombre voit			28.9%
3	Nombre bus	Nombre mob	Nombre mot		58.6%
	Nombre bus	Nombre mob	Nombre voit		73.8%
	Nombre bus	Nombre mot	Nombre voit		26.9%
	Nombre mot	Nombre voit	Nombre mob		0.0%
4	Nombre bus	Nombre mob	Nombre mot	Nombre voit	81.2%

Tableau 21b : Modèles sans sonie obtenus en fonction des variables objectives.

Son équation est la suivante :

$$Dès = 2.27 + 0.54 \times NbVoit + 2.95 \times NbBus + 1.11 \times NbMot + 1.64 \times NbMob$$

La figure 40 présente schématiquement le modèle de régression.

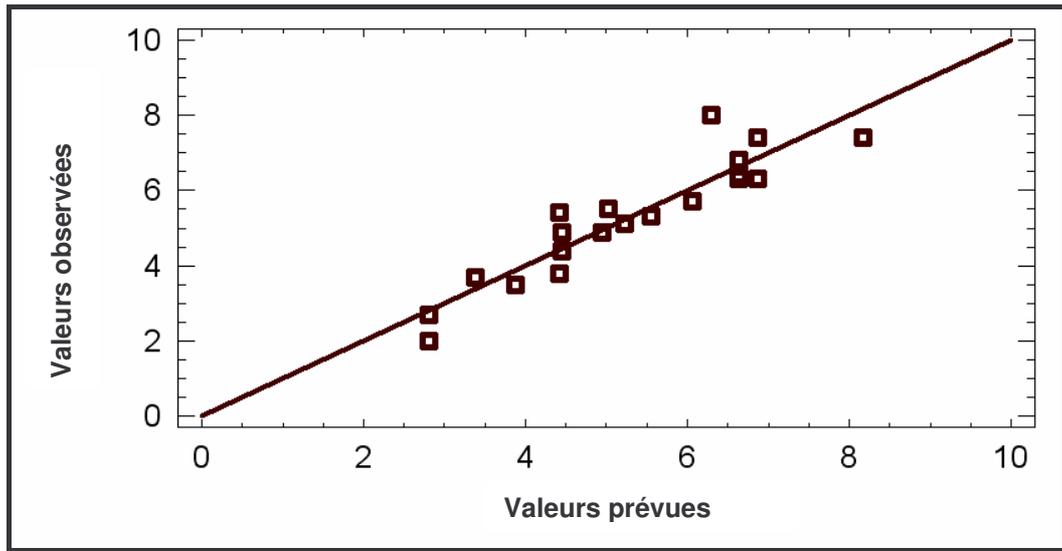


Figure 40 : Correspondance entre les valeurs du modèle (objectif et sans sonie) de régression linéaire et les valeurs expérimentales. $R^2_{aj} = 81.2\%$.

3.4. Domaines de validité des modèles objectifs – intervalles de confiance

Une fois les modèles obtenus, il est important d'évaluer leurs domaines de validité en vue de nouvelles prédictions. Nous exposons une procédure qui peut être employée pour estimer le domaine de validité selon un niveau de confiance désiré. Nous l'utilisons sur le modèle objectif avec sonie.

Tout d'abord, le modèle testé a été établi à partir des données suivantes, accompagnées de leur plage de donnée. Ainsi, sur une donnée de 15 secondes :

- Sonie moyenne calculée (N) en sones : $8.3 \leq N \leq 23$
- Nombre de bus (NbBus) : $0 \leq \text{NbBus} \leq 2$
- Nombre de cyclomoteurs (NbMob) : $0 \leq \text{NbMob} \leq 2$
- Pourcentage de temps d'apparition des voix d'enfants (%TVE) : $0\% \leq \%TVE \leq 55\%$

Ainsi, nous pouvons d'ores et déjà affirmer que la validité du modèle sera limitée à cette plage de données des variables explicatives. En effet, si la sonie est égale à cinquante et le nombre de bus à deux, alors le modèle prédit des valeurs supérieures à dix.

Néanmoins, la robustesse du modèle repose sur des plages de données relativement larges pour des séquences de 15 s.

Ensuite, comme le but ici est de construire des intervalles de confiance en vue de nouvelles observations, le principe retenu est de faire varier les variables (si possible dans la plage sur laquelle a été construit le modèle) et d'extraire l'intervalle de confiance à 95% qui en découle.

En d'autres termes, on teste les domaines de validité en regardant, en fonction des variables du modèle, la taille de l'intervalle de confiance à 95%. Cela ne garantit pas une réponse du modèle juste, mais nous éclaire sur l'intervalle de confiance de la réponse du modèle. On peut par exemple estimer qu'un intervalle est satisfaisant quand on a 95% de chances d'avoir la réponse comprise dans un intervalle de deux (ce qui correspond à une erreur de plus ou moins un).

Ci-dessous nous exposons des intervalles de confiance à 95% obtenus en fonction des quatre variables : sonie calculée, pourcentage de présence des voix d'enfants, nombre de bus et nombre de cyclomoteurs. [Figure 41].

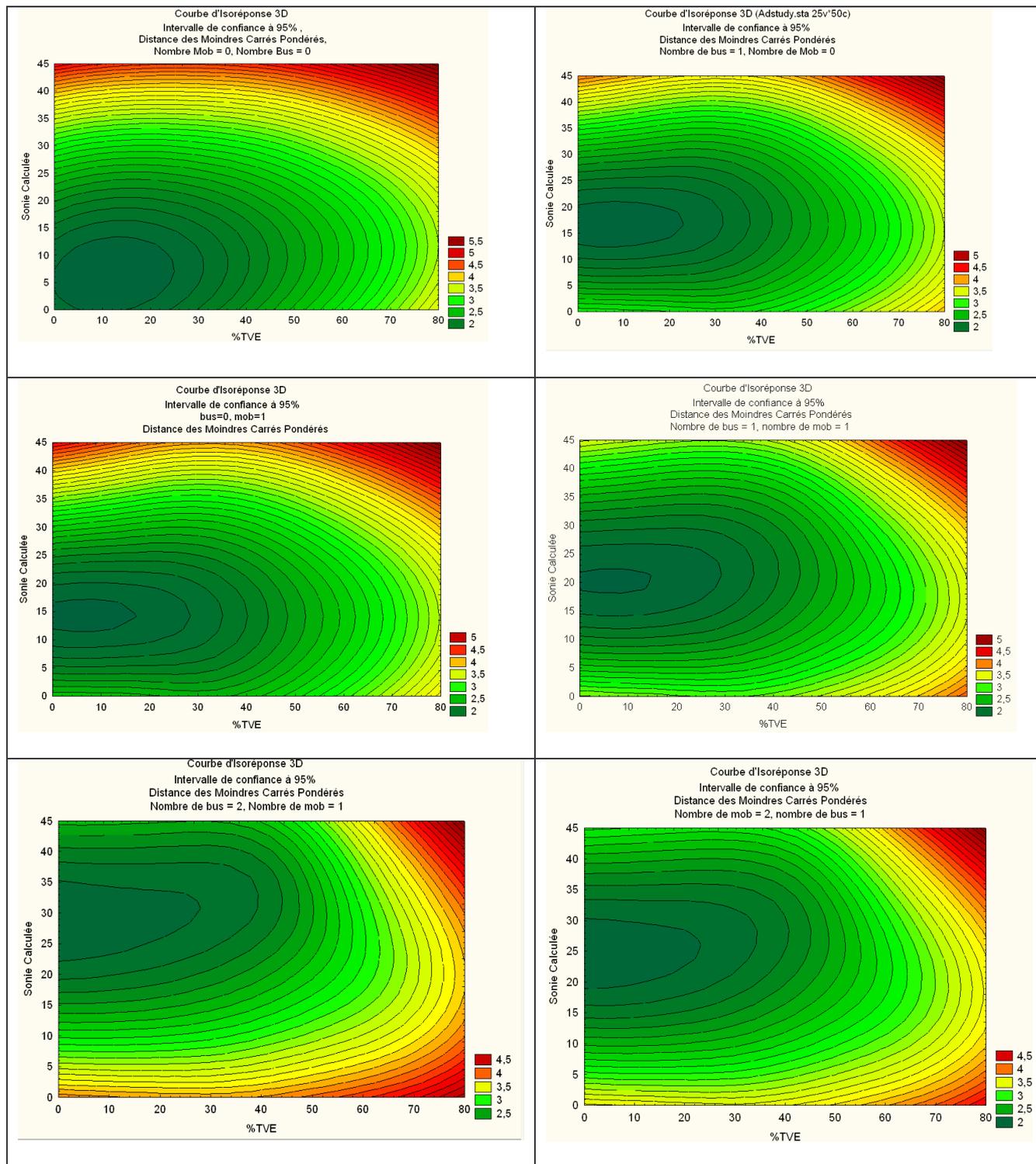


Figure 41 : Intervalles de confiances à 95% en fonction des quatre variables du modèle objectif de désagrément avec sonie. Le nombre de bus varie de « 0 » à « 2 », de même que le nombre de cyclomoteurs (mob).

Lors de futures utilisations des modèles de désagrément établis dans ce mémoire en tant que modèles prédictifs, il est possible d'exposer en complément de la mesure, un écart ou une erreur de prévision.

Derrière la représentation graphique de ces intervalles de confiance des modèles, on remarque que la prédiction du modèle perd en justesse lorsque l'on sort du cadre dans lequel il a été construit. Ainsi, nous relevons l'utilité de construire des modèles sur des observations qui soient les plus larges possibles.

3.5. Synthèse sur les modèles de régression

Les modèles extraits représentent la formalisation de l'hypothèse de départ. Cette hypothèse repose sur le fait que, dans un environnement sonore qui est varié en termes de sources sonores et complexe en termes d'analyse physique, l'emploi de critères purement énergétiques pour expliquer le désagrément peut être amélioré en tenant compte de la composition de cet environnement.

Nous avons tout d'abord confirmé par rapport à d'autres études que l'emploi de la sonie calculée sur le modèle de Zwicker pour estimer le désagrément sonore est mieux adapté que le niveau de pression acoustique pondéré A. A la différence près que nous trouvons une meilleure corrélation avec le calcul de la sonie moyenne alors que d'autres études parlent de la sonie dépassée pendant 4% du temps. Seulement, bien que le calcul de la sonie soit normalisé, sa complexité est la raison essentielle de sa non-popularité. De ce fait, elle n'apparaît dans aucune législation. De plus, son modèle a été déterminé pour des sons stationnaires, or les sons de l'environnement sont, dans la très grande majorité des cas, variables en niveau. Bien que des études montrent que l'on peut être proche du désagrément par l'emploi d'indices acoustiques associés comme les indices percentiles ou encore en utilisant la sonie moyenne, la rationalité de ces critères n'est pas encore approuvée.

En second lieu, nous avons montré que l'on pouvait améliorer la mesure du désagrément par la prise en compte de l'influence de certaines sources sonores.

En effet, nous avons trouvé des modèles sous forme de régression linéaire multiple ayant une très bonne correspondance avec les résultats expérimentaux. Ces modèles intègrent des variables qui sont relatives aux « objets bruits » identifiés comme les bus et les cyclomoteurs ou à des variables ayant une influence sur l'attention des personnes comme les voix d'enfants. Les critères utiles à ces modèles sont plutôt relatifs au temps (pour les voix et pour les véhicules) ou au nombre de sources (pour les véhicules) que basés sur leur intensité.

La figure 42 montre que lorsque certaines sources (en l'occurrence celles qui entrent en compte dans les modèles) apparaissent dans les stimuli, l'écart entre la sonie et le désagrément a

tendance à augmenter, dans un sens lorsqu'il y a des cyclomoteurs ou/et des bus, dans un autre sens lorsqu'il y a des voix d'enfants.

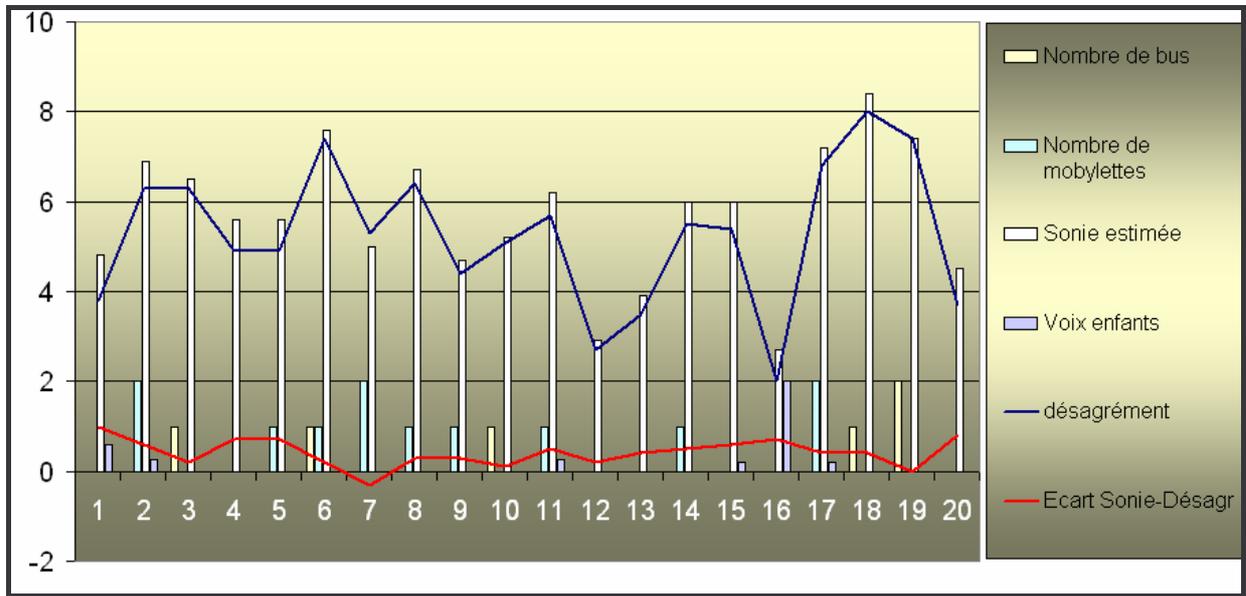


Figure 42 : Analyse Sonie estimée/Désagrément estimé en fonction des sources sonores influentes : influence des données objectives sur les données subjectives. En abscisse nous avons les numéros des stimuli. On remarque que l'écart entre la sonie et le désagrément augmente quand les variables reliées aux cyclomoteurs, aux bus et aux voix d'enfants sont présentes.

D'un autre côté, nous avons constaté que les oiseaux étaient littéralement « noyés » dans les bruits des véhicules. De ce fait les auditeurs n'ont pas tenu compte de leur influence généralement positive observée dans d'autres recherches ([SCHAFER-77], [VIOLLON-00]). Ainsi ces derniers n'apparaissent pas dans les modèles et n'ont donc pas d'effet significatif sur le désagrément sonore. A travers l'analyse détaillée de la composition des stimuli, nous avons cependant vu que les oiseaux ont bien une influence positive dans le cas d'une ambiance de parc.

Enfin, les voitures, bien qu'étant la source sonore la plus fréquente, n'apparaissent pas dans les modèles. Elles sont directement intégrées dans l'évaluation de la sonie.

Les modèles indiquent qu'il est possible de prévoir un indice subjectif (le désagrément sonore) pour des séquences de 15 s représentant une ambiance sonore urbaine dite « événementielle ». Cette modélisation est rendue possible par l'utilisation, non seulement de la sonie, c'est-à-dire de l'intensité sonore subjective, mais également par la prise en compte de valeurs subjectives propres à certaines sources sonores présentes dans les stimuli. De plus, de bonnes correspondances ont été trouvées entre les variables subjectives des modèles et des mesures physiques. Ainsi, il nous est possible de modéliser le désagrément sonore à partir de l'analyse du signal physique. Nous avons ainsi créé un « pont » entre le domaine subjectif et le domaine objectif. En dernier lieu,

nous avons également montré que le fait de compter les véhicules présents dans les stimuli nous donne des résultats similaires à la mesure de la sonie.

Les tests réalisés in-situ ont montré que chaque ambiance sonore caractéristique d'un lieu (parc, rue, marché) doit être analysée selon la composition de son environnement. La rue circulée est caractérisée de façon importante par la force sonore mais peut également être caractérisée par des critères temporels liés aux véhicules. Dans des ambiances qualifiées de calmes et/ou agréables, la force sonore n'est plus un bon descripteur si l'on veut suivre la qualité sonore perçue. Il faut alors tenir compte des temps d'apparition des sources (voix, oiseaux) et des véhicules.

Au final ces observations tendent à démontrer que, pour caractériser une ambiance sonore, il suffit de connaître sa composition en termes de sources et de critères temporels associés (nombre ou temps d'apparition).

Nous avons vu en introduction, chapitre 2, que Björk [BJÖRK-84] a montré que la perception des sons naturels pouvait être décrite à travers deux émotions : le plaisir et l'éveil. Russell [RUSSELL-80] suggère aussi que les émotions plus complexes (comme la peur, la colère, la joie, la tristesse...) peuvent être appréhendées sur ces deux dimensions bipolaires non indépendantes. Ainsi, nous proposons de construire un espace émotionnel, réunion de ces deux dimensions (un axe « déplaisir/plaisir » et un autre « discrétion/éveil ») afin d'y placer les noms des sources sonores étudiées dans ce mémoire telles que nos analyses perceptives nous les suggèrent [Figure 43]. La position des sources dans cet espace est bien sûr relative uniquement à l'environnement sonore présenté aux auditeurs. Ces positions ne se veulent pas être descriptives d'un cas général car elles doivent être adaptées au type d'ambiance sonore dépendant du lieu (leur position résulterait de l'adéquation pour l'auditeur entre les sources sonores et le lieu d'occurrence).

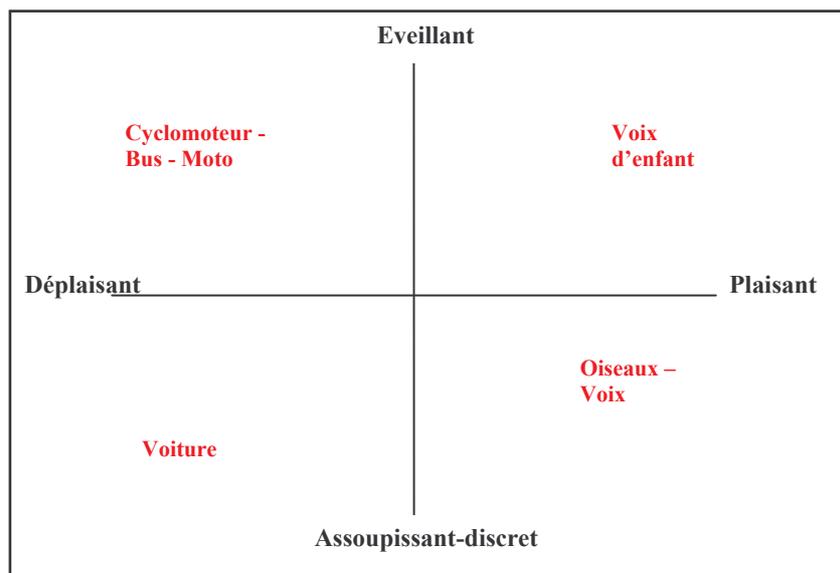


Figure 43 : Sources sonores dans un plan émotionnel, disposées telles qu'elles ont été perçues durant les tests auditifs.

Dans les modèles de régressions multiples du désagrément sonore présentés, seules les sources qualifiées d'« éveillantes » dans ce plan entrent comme variables. Bien que l'influence des autres sources ait été suggérée, elle n'est pas statistiquement significative et peut être directement intégrée dans la mesure du niveau sonore (sonie).

Cette classification des sources selon leur « importance perceptive » rejoint une hypothèse couramment admise concernant la fonction fondamentale de l'émotion. En préparant l'organisme à l'action et en réduisant la complexité des problèmes rencontrés, elle constitue un prétraitement (dans le sens où elle agit avant la formation de la perception) permettant à la raison de fonctionner de manière efficace [SCHERER-89]. En d'autres termes, elle permet le découplage entre la réception d'une stimulation et la réaction induite.

Dans une ambiance sonore telle que nous l'avons étudiée, ce sont les voix d'enfants et certaines sources (bus, cyclomoteurs, motos) qui ont la plus grande importance émotionnelle, dans le sens où elles provoquent un éveil et entrent comme variables dans l'appréciation du désagrément sonore.

Nous suggérons qu'un objet sonore peut également être caractérisé selon une dimension qui traduit son importance émotionnelle. Cette caractérisation permet de décrire une de ses propriétés que nos expériences ont suggérées ici : fournir des informations nécessaires pour déclencher une réponse comportementale adaptée à la situation [DAMASIO-03]. Cette importance émotionnelle peut être représentée dans l'espace émotionnel constitué par la réunion de l'axe « déplaisir/plaisir » et de l'axe « discrétion/éveil ».

IV. Caractérisation de la séquence de 5 min

L'objectif de ces tests est d'obtenir des éléments nous permettant de décrire la caractérisation subjective d'une séquence de durée moyenne (5 min), à partir de la caractérisation des éléments la composant. Cela peut nous permettre d'émettre des hypothèses quant aux processus perceptifs qui permettent aux sujets d'apprécier de façon globale les variables étudiées dans ce mémoire.

4.1. Test n° 4 : Prégnance, présence et proximité pour la séquence globale de 5 min

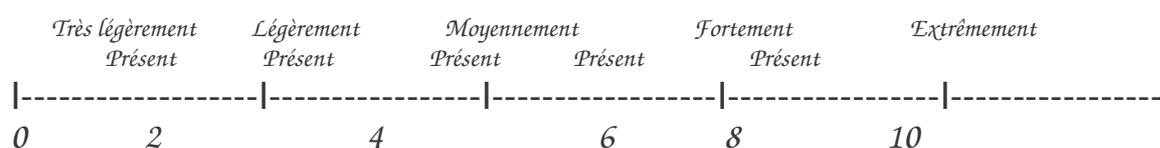
4.1.1. Questionnaire et procédure

La séquence de 5 min est formée à partir du regroupement des vingt stimuli qui ont fait l'objet du test n° 1. Sur une interface [Figure 44] similaire à celle du test n° 1, les auditeurs sont invités à noter, à la fin de la séquence, la prégnance, la présence et la proximité globales perçues des sources qui leur sont indiquées, sur la durée totale du test. Comme la tâche demandée peut s'avérer difficile, la séquence est présentée deux fois. C'est seulement à la fin de la deuxième écoute que les auditeurs rentrent les notes. Le protocole expérimental est le même que pour le test n° 1. Vingt-deux sujets ont pris part à cette expérience.

Les instructions sont inspirées de l'échelle de Borg [BORG-02] (Annexe 5). L'échelle va de « 0 » (pour une absence de la source considérée) à « 10 ». La note « 10 » est présentée aux sujets comme correspondant à une sensation maximale par rapport à son expérience antérieure. Les sujets n'ayant pas d'idée de l'échelle, on leur demande de se baser sur leur expérience personnelle (d'où l'intérêt de prendre un échantillon homogène de sujets).

Aux valeurs numériques correspond une échelle sémantique qui permet de guider le sujet dans ses réponses (les chiffres doivent être des entiers). L'échelle pour la sensation de présence est visible ci-dessous :

Echelle de présence



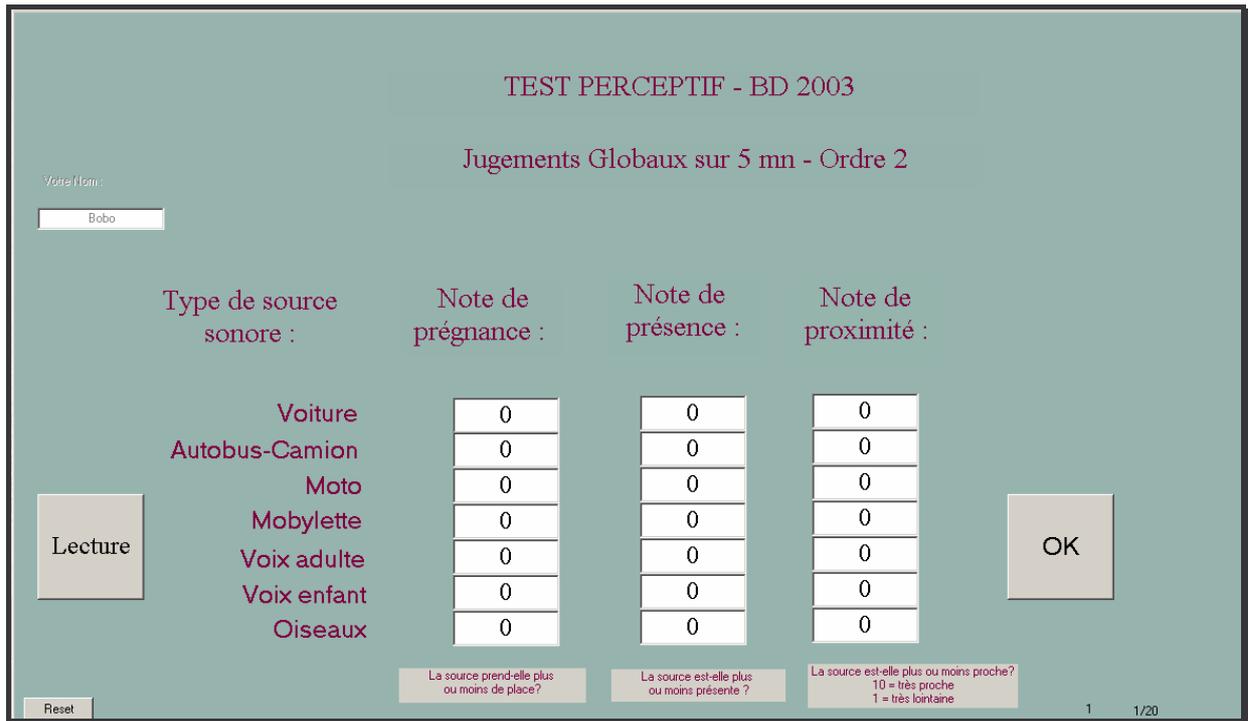


Figure 44 : Interface du test de prégnance, présence et proximité des sources sonores pour la séquence de 5 min.

4.1.2. Normalisation des données

Bien que les réponses des sujets soient notées sur les mêmes échelles, elles sont tout de même normalisées pour que la variance des réponses de chaque personne soit constante. La normalisation est la suivante :

$$X_{norm,j,p} = X_{i,j,p} \sqrt{\frac{\sum_{j,p} X_{j,p}^2}{\sum_j X_{j,p}^2}}$$

Où

$p = 1$ à P personnes

$j =$ questions 1 à 21 (trois questions pour sept sources différentes),

$X_{j,p} =$ réponse initiale de la personne p pour le stimulus i et la question j ,

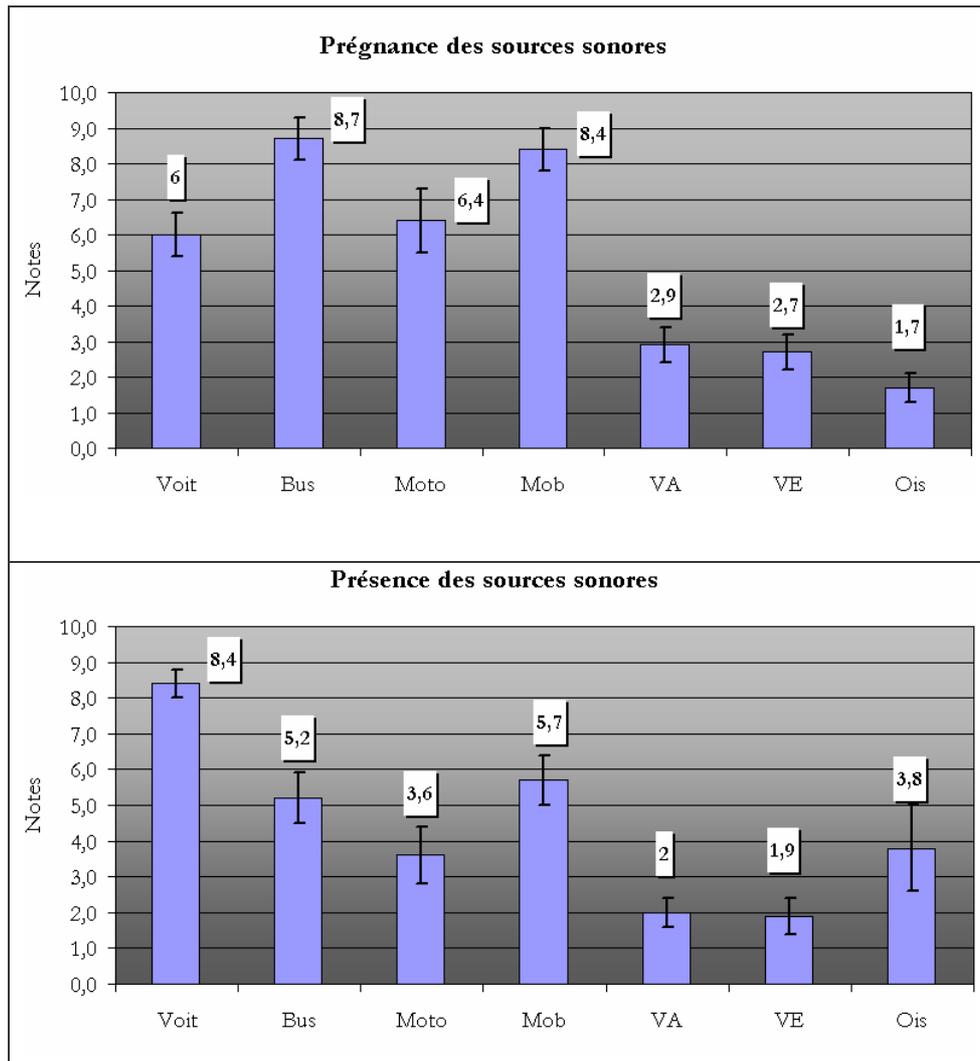
$X_{norm,j,p} =$ réponse normalisée de la personne p pour le stimulus i et la question j ,

$\sum_j X_{j,p}^2 =$ somme des carrés de toutes les réponses de la personne p ,

$$\overline{\sum_{j,p} X_{j,p}^2} = \left(\sum_{p=1}^P \sum X_{j,p}^2 \right) / P = \text{moyenne de la somme des carrés de tous les sujets.}$$

4.1.3 Présentation des Résultats

Les graphes ci-dessous [Figure 45] montrent les moyennes et écart types des réponses des sujets.



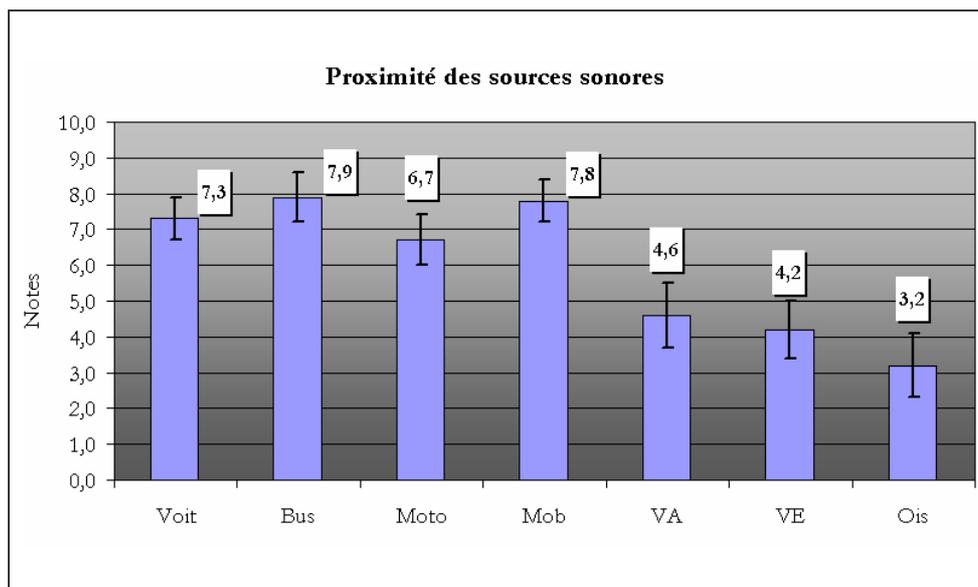


Figure 45 : Moyennes et écarts types des réponses des sujets pour la prégnance, la présence et la proximité de l'ensemble des sources sonores (séquence de 5 min, moyenne sur vingt-deux sujets).

Les écarts types sont relativement faibles [Figure 45]. Cela signifie que les notions ont été comprises de manière similaire par l'ensemble des auditeurs.

Les valeurs de prégnance et de présence sont nettement plus élevées pour les véhicules que pour les sources naturelles. Ce résultat s'explique par le fait que l'enregistrement représente une ambiance sonore de rue circulée, les sons mécaniques y sont dominants. Le tableau ci-dessous expose les pourcentages de temps d'apparition des différentes sources sur les 5 min.

Voiture	Bus	Moto	Mob	VA	VE	Ois
42,5%	16,5%	6,6%	21,2%	7,0%	4,8%	26,7%

On constate que les oiseaux (Ois) ont un pourcentage supérieur aux bus et aux cyclomoteurs alors qu'ils ont été jugés comme étant moins présents. Ici encore, on remarque que les oiseaux ont été noyés par les autres sons et leur présence a été sous-évaluée.

Le tableau ci-dessous indique les trois valeurs les plus importantes des L_{Aeqmax} pour chacun des véhicules.

Voitures			Bus			Moto			Mob		
69,1	65,8	64,9	74,4	74,1	71,4	73,3	66,7	64,9	70,5	70,4	68,6

La prégnance est proportionnelle aux niveaux maximums des véhicules.

L'écart entre les sources mécaniques et non mécaniques est plus faible pour les notes de proximité. Le peu de variation dans la séquence de la distance entre les sources et l'auditeur a probablement fait ressortir cette dimension de manière plus réaliste.

4.1.4. Analyses statistiques

La prégnance, la présence et la proximité des sources sur une écoute globale sont comparées aux indices statistiques calculés sur les réponses de présence, prégnance et proximité par source et par stimulus, de façon à comprendre comment les sujets globalisent les réponses.

Nous supposons que nous pouvons comparer de façon objective les réponses faites sur des stimuli aux réponses faites sur la séquence globale. En effet, les valeurs maximales de prégnance, de présence et de proximité concernant les stimuli ont été choisies pour être représentatives de sensations maximales correspondant à l'expérience « normale » d'un sujet en situation urbaine.

Les statistiques sont les suivantes : Maximum, Minimum, Médiane (centile 50%), Moyenne, Centile 10% (moyenne de 10% des valeurs maximales ce qui revient à dire que 10% des réponses sont au-dessus de cette valeur), Centile 90% (moyenne de 90% des valeurs maximales, ou bien 90% des réponses sont au-dessus de cette valeur).

Par exemple, voici, sur la figure 46, la moyenne des réponses des sujets sur la prégnance des voitures par stimulus :

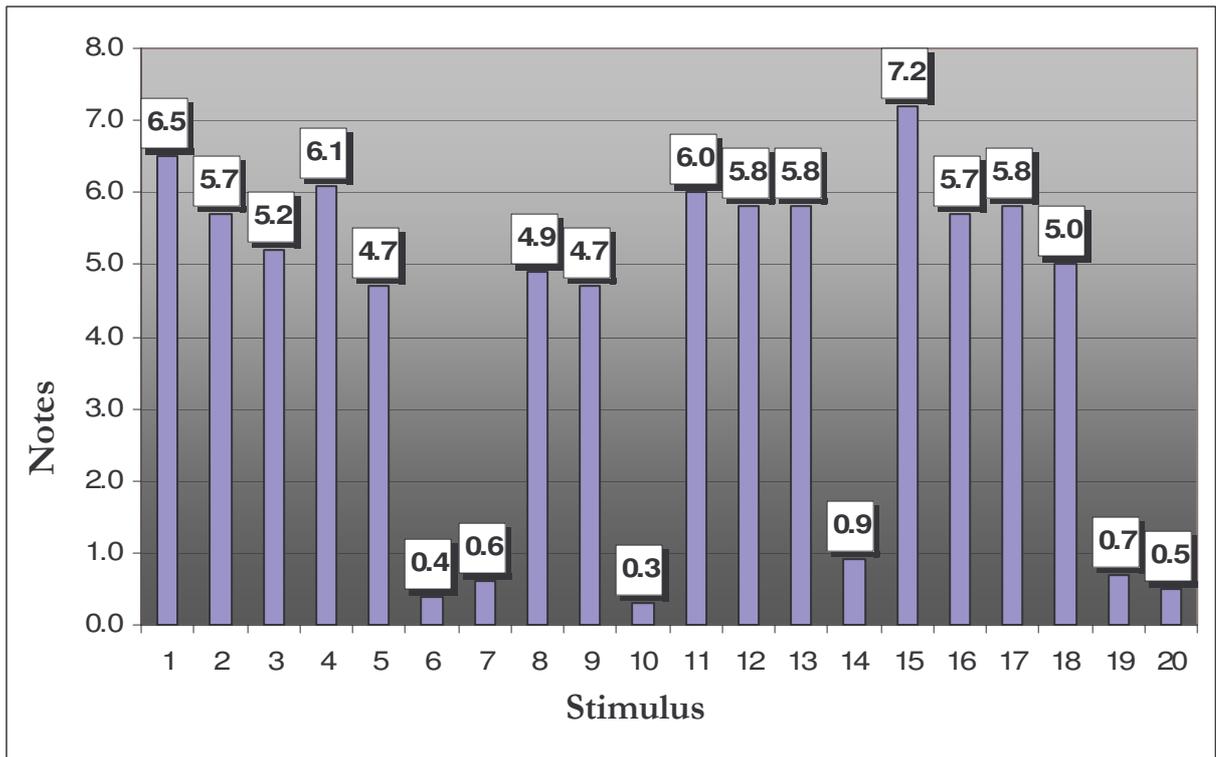


Figure 46 : Moyenne des réponses des sujets pour la prégnance des voitures pour les vingt stimuli.

Nous en déduisons les indices statistiques suivants [Figure 47] :

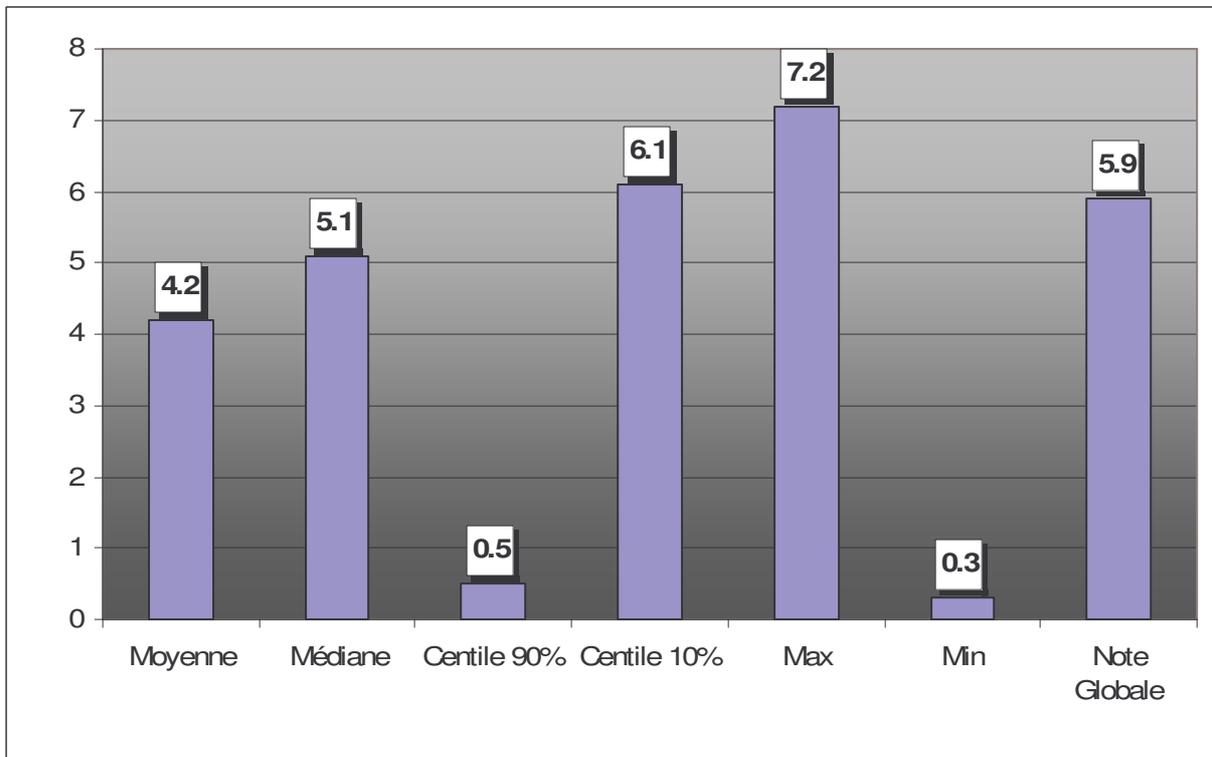


Figure 47 : Graphique représentant les valeurs statistiques (moyenne, médiane, centile 90%, centile 10%, maximum, minimum) calculées sur les vingt stimuli ainsi que la note globale (pour la séquence de 5 min) pour la prégnance des voitures.

La moyenne normalisée de la « prégnance voiture » sur la séquence de 5 min est : « 5.9 ».

Ainsi, la valeur la plus proche de la prégnance globale des voitures est la valeur centile 10%. Cette valeur correspond aux valeurs maximales qu'a atteint la prégnance des voitures sur les vingt stimuli. On peut en déduire que, pour noter une séquence globale, les sujets se sont appuyés sur les événements marquant de cette séquence.

Remarque importante : pour comparer des jugements entre eux, il est important que les auditeurs aient la même dynamique d'échelle, en particulier qu'ils aient une même référence pour les sensations de maximum (note « 10 »). Or, entre les expériences sur les stimuli de 15 s et l'expérience sur la séquence de 5 min, il est évident qu'il est difficile de comparer les valeurs maximales. Cependant, on peut supposer que les « 10 » pour les stimuli construits sur 15 s sont aussi les « 10 » de l'expérience urbaine de Paris dans une rue de quartier car certains stimuli contiennent jusqu'à cinq voitures (stimulus « 4 ») ou deux bus (stimulus « 19 ») ou deux cyclomoteurs et deux voitures (stimulus « 2 ») en 15 s. Les maxima de ces trois stimuli sont à rapprocher des maxima de l'expérience personnelle des sujets qu'il leur a été demandé d'imaginer pour le test portant sur la séquence de 5 min.

Sur le tableau 22 nous pouvons voir que globalement, pour toutes les autres variables, les auditeurs suivent la même logique : ils ont fait des estimations globales en s'appuyant sur les événements marquants de la séquence. Ces observations recourent les conclusions déjà obtenues par plusieurs études faites sur la sonie de séquences sonores extraites d'ambiances urbaines [FASLT-91] d'une part, et d'accélération de voitures [SUSINI-02] d'autre part. De même, Kuwano et al ([KUWANO-02]) ainsi que Gottschling [GOTTSCHLING-99] ont étudié les relations entre des jugements continus et des jugements globaux effectués immédiatement après la fin de séquence. Ils observent que les jugements globaux sont toujours supérieurs à la moyenne des jugements continus. Les auteurs concluent que le jugement global dépend principalement des événements dont les niveaux acoustiques crêtes sont les plus importants.

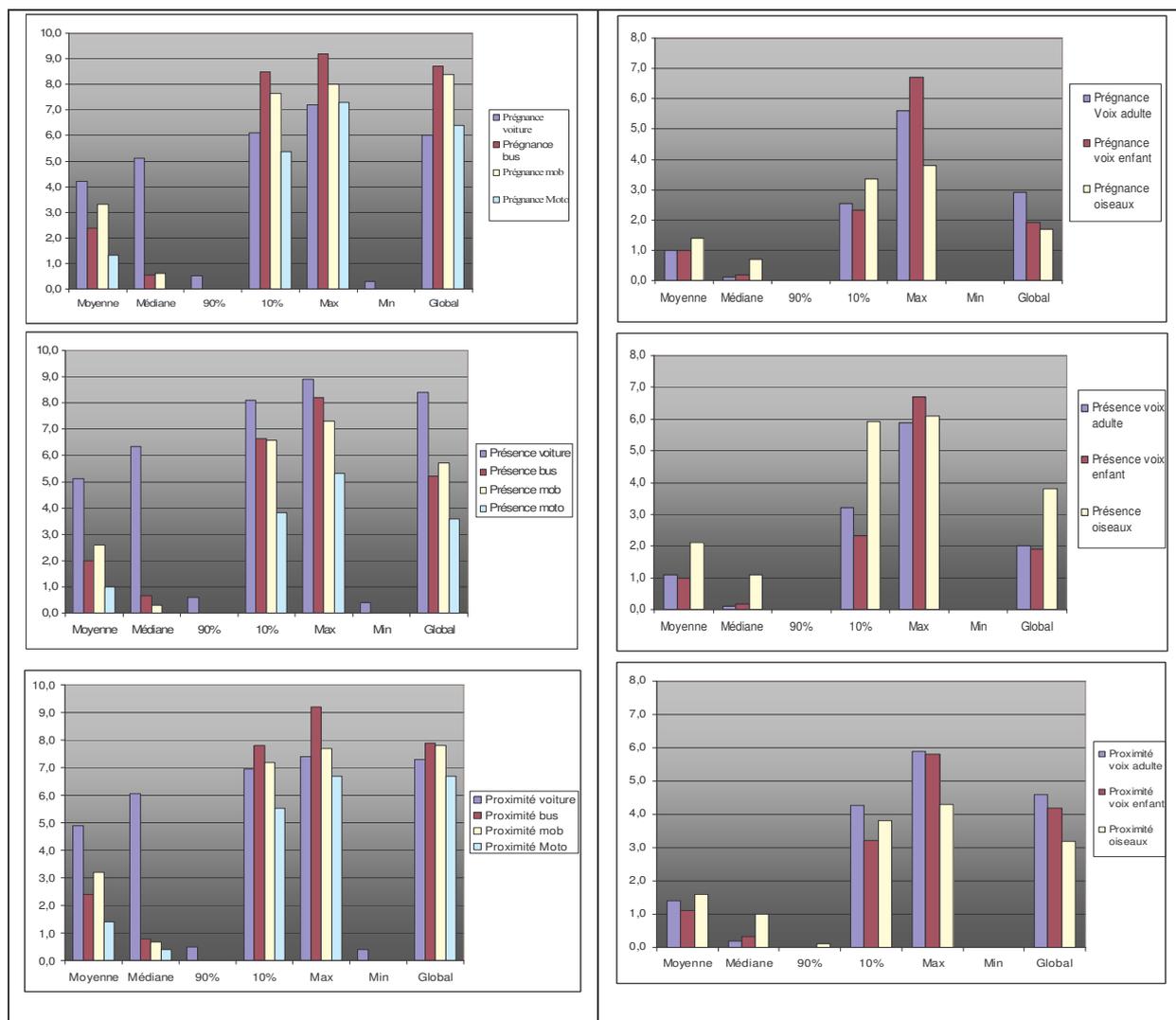


Tableau 22 : Graphiques représentant les valeurs statistiques (moyenne, médiane, centile 90%, centile 10%, maximum, minimum) calculées sur les vingt stimuli ainsi que les notes globales (à droite des graphiques). Ces histogrammes montrent que les notes globales sont proches du 10% des valeurs par stimuli montrant que les estimations globales dépendent ici des événements dominants.

4.1.5. Effet de récence

Les jugements globaux peuvent être soumis également à une autre influence : l'effet de récence. Cette autre « stratégie » de mémorisation à long terme des événements a été mise en évidence dans le domaine sonore par Susini [SUSINI-99], [SUSINI-02]. Elle correspond à « un processus qui favorise le stockage à court terme des derniers événements entendus » [SUSINI-05].

Nous allons donc vérifier si la composition de la séquence globale a pu faire que les jugements des auditeurs soient influencés par les événements qui se produisent à la fin de la séquence. Si tel était le cas, étant donné que les jugements des auditeurs sont proches des maxima, il faudrait que les « événements marquants » se produisent à la fin de la séquence. Or nous savons déjà que c'est

peu probable que, pour l'ensemble des sources sonores, ces dernières soient apparues de manière franche uniquement à la fin de la séquence.

De façon plus détaillée, le tableau 23 contient les numéros des stimuli pour les valeurs maximales attribuées aux différents adjectifs subjectifs, les temps écoulés lors de ces maxima et le ratio de ces temps sur le temps total de la séquence (300 s).

Variables	Numéro de stimuli pour lequel le maximum est apparu	Temps écoulé depuis le début de la séquence	Ratio temporel par rapport à la longueur totale de la séquence
Prégnance voiture	15	120	0,40
Présence voiture	4	240	0,80
Proximité voiture	15	120	0,40
Prégnance bus	19	180	0,60
Présence bus	19	180	0,60
Proximité bus	6	270	0,90
Prégnance cyclomoteur	17	150	0,50
Présence cyclomoteur	7	285	0,95
Proximité cyclomoteur	14	105	0,35
Prégnance Moto	20	75	0,25
Présence moto	20	75	0,25
Proximité Moto	20	75	0,25
Prégnance Voix adulte	14	105	0,35
Présence voix adulte	1	195	0,65
Proximité voix adulte	14	105	0,35
Prégnance voix enfant	1	195	0,65
Présence voix enfant	16	135	0,45
Proximité voix enfant	1	195	0,65
Prégnance oiseaux	20	75	0,25
Présence oiseaux	20	75	0,25
Proximité oiseaux	20	75	0,25

Tableau 23 : Illustration des « moments » où se sont produits les maximums parmi les variables subjectives.

Ce tableau permet de construire un histogramme illustrant, pour l'ensemble des variables, le nombre de maxima obtenu en fonction du ratio temporel. [Figure 48].

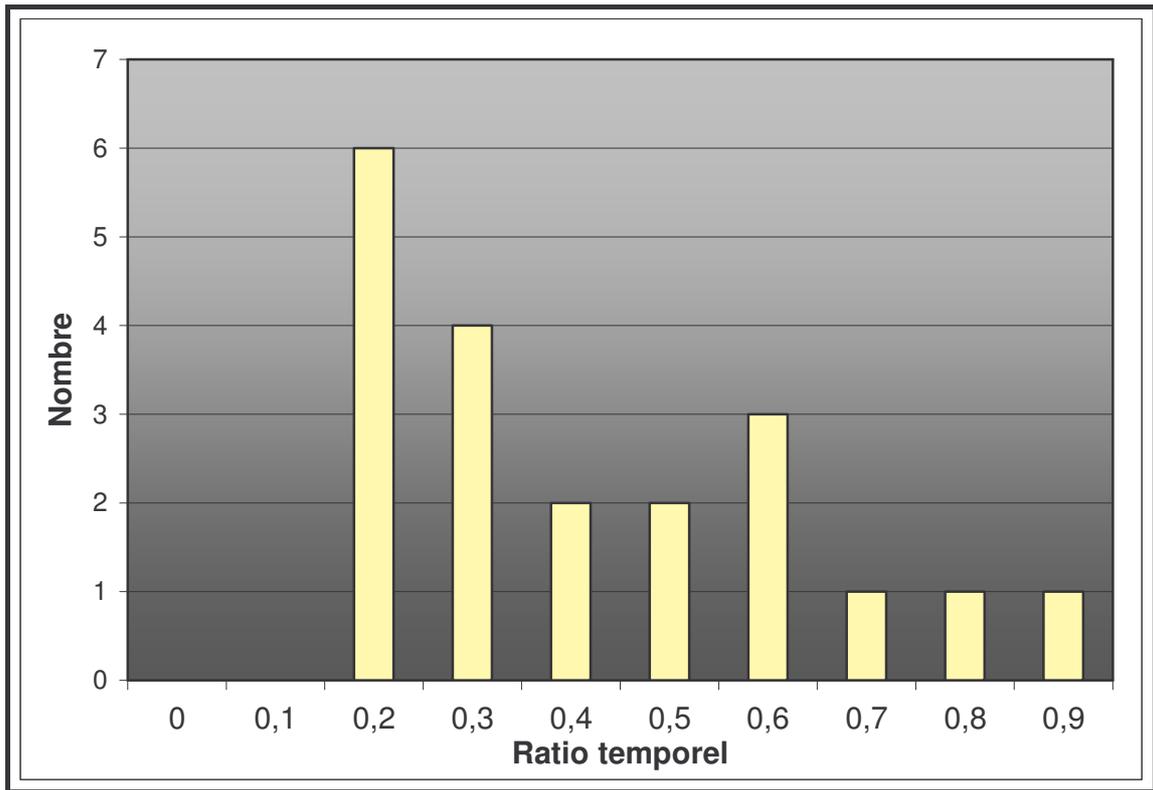


Figure 48 : Distribution temporelle par rapport à la longueur totale de la séquence globale, de l'ensemble des variables prégnance, présence, proximité de chaque catégorie de source sonore. Par exemple, il y a eu six maxima qui se sont produits à un ratio entre 0.2 et 0.3 c'est à dire à une durée d'environ $300 \times 0.25 = 75$ s.

Nous voyons ainsi que globalement, les maximanes sont pas apparus à la fin de la séquence.

Nous allons voir par la suite (paragraphe 4.3.) que, lors du test réalisé in situ, la fin des séquences n'a pas eu d'effet significatif sur les jugements globaux des auditeurs.

4.2. Tests n° 5 et n° 6 : Sonie et désagrément sonore pour la séquence globale de 5 min

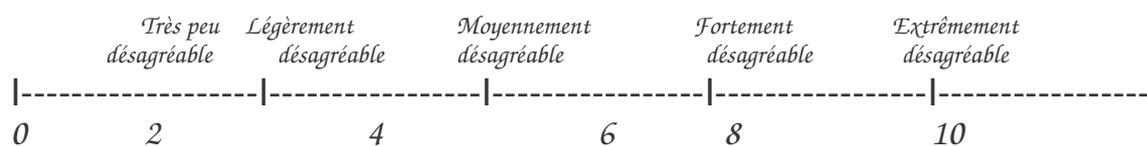
4.2.1. Questionnaire et procédure

La séquence de 5 min est formée à partir du regroupement des vingt stimuli qui ont fait l'objet du test n° 1. Sur une interface ergonomique [Figure 49] similaire à celle du test n° 2, les auditeurs sont invités à noter, à la fin de la séquence, la sonie ou le désagrément sonore global perçu sur la durée totale de la séquence. Les deux tests ont été réalisés séparément pour rendre la signification des deux termes la plus indépendante possible. La séquence est présentée une seule fois. C'est seulement à la fin de l'écoute que les auditeurs rentrent les notes. Le protocole expérimental est le même que pour le test 1. Vingt-six sujets ont participé au test sur le désagrément, dix-sept au test sur la force sonore.

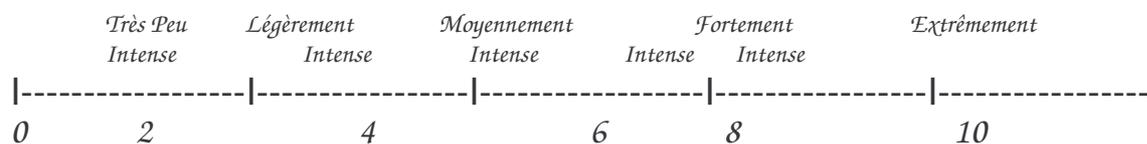
De la même manière que sur l'échelle de Borg [BORG-02], des échelles sémantiques sont présentées pour faciliter la réponse des sujets sur l'échelle numérique. L'échelle va de « 0 » (pour une absence de la source considérée) à « 10 ». La note « 10 » est présentée au sujet comme correspondant à une sensation maximale par rapport à son expérience antérieure.

Les échelles respectives sont présentées ci-dessous :

Echelle de désagrément sonore



Echelle d'intensité sonore :



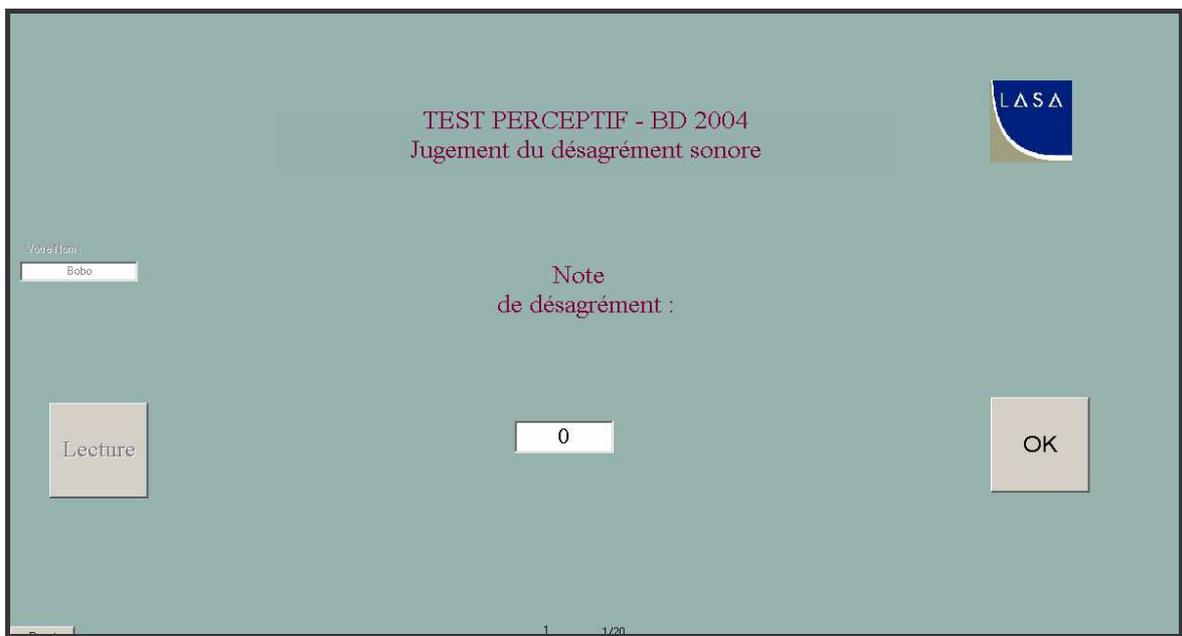
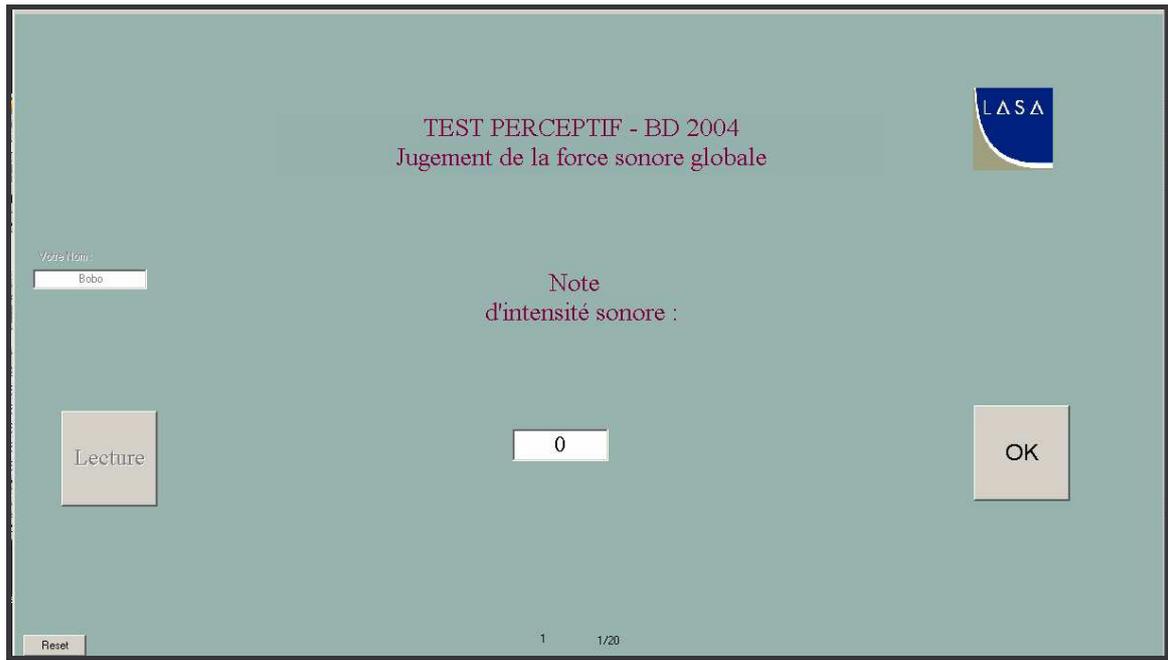


Figure 49 : Interface des tests de sonie et de désagrément sonore pour la séquence de 5 min.

4.2.2. Analyses statistiques

4.2.2.1. Résultats et Moyennes

Les résultats sont donnés sous forme d'histogrammes [Figure 50] des réponses.

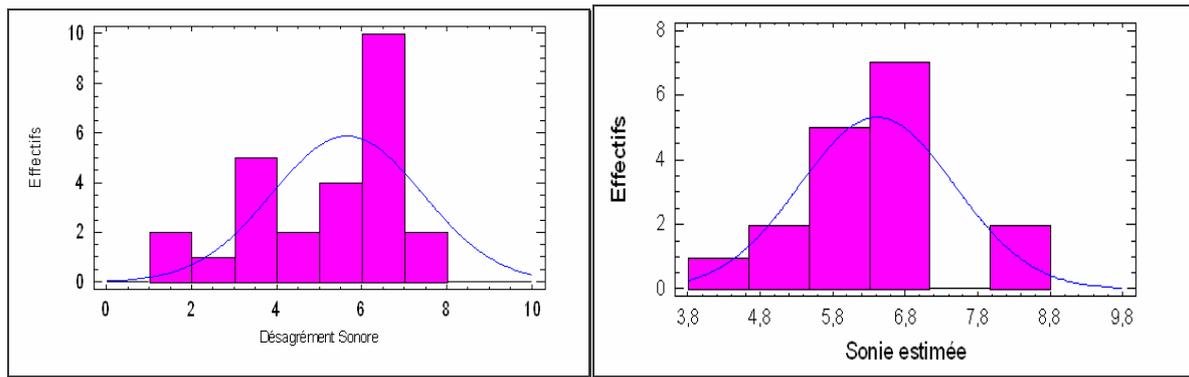


Figure 50 : Histogramme des effectifs pour les jugements de désagrément et de sonie.

Des tests statistiques sont réalisés sur les valeurs de désagréments et de sonie afin de vérifier la « cohérence » dans les réponses des sujets. Ces différentes analyses, *l'histogramme des effectifs* et la *boîte à moustaches*, vérifient en priorité la distribution en allure de gaussienne des résultats et s'assurent de l'absence de valeur aberrante. Les conclusions de ces tests nous permettent de faire des moyennes sur l'ensemble des réponses recueillies.

Histogramme des effectifs et loi normale :

Désagrément sonore	Sonie subjective
Asymétrie standardisée = -1,46069	Asymétrie standardisée = -1,04565
Aplatissement standardisé = -0,604082	Aplatissement standardisé = 0,238834

Les coefficients d'aplatissement standardisé et d'asymétrie standardisée sont utilisés pour déterminer si l'échantillon suit une loi normale. Des valeurs de ces statistiques en dehors de la plage « -2 » à « +2 » indiquent un écart significatif de la normalité. Pour les valeurs de désagrément et de sonie, les valeurs de l'asymétrie standardisée ainsi que de l'aplatissement standardisé sont dans la plage attendue pour des données suivant une loi normale [Figure 50].

Boîte à moustaches :

La boîte à moustaches est un outil permettant de représenter schématiquement la distribution d'une variable [TUKEY-77]. La partie rectangulaire du graphique s'étend du premier au troisième quartile, couvrant la moitié centrale des données. La ligne centrale à l'intérieur de la boîte indique la position de la médiane. Le signe plus indique la position de la moyenne. Les moustaches s'étendent de la boîte jusqu'au minimum et au maximum des données de l'échantillon, sauf pour les points suspects et extrêmes, qui sont affichés séparément. Les points suspects sont des points à plus de une fois et demi l'étendue inter-quartiles au-dessus ou au-dessous de la boîte et ils sont affichés comme des petits carrés. Les points extrêmes sont des points à plus de trois fois

l'étendue inter-quartiles au-dessus ou au-dessous de la boîte et ils sont affichés comme des petits carrés contenant un signe plus.

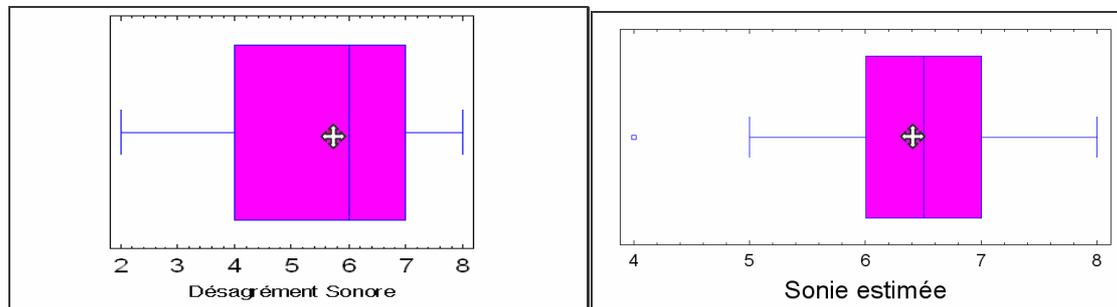


Figure 51 : Boîtes à moustaches pour le désagrément sonore et la sonie estimée. Il y a un « point suspect » pour la sonie estimée qui est la note 4. Elle est supérieure à une fois et demi mais inférieure à trois fois l'étendue interquartile en dessous de la boîte.

Dans le cas du désagrément, il n'y a pas de points suspects ni de points extrêmes [Figure 51].

Dans le cas de la sonie estimée nous avons un point suspect correspondant à la note 4. Comme ce n'est pas un point extrême et qu'il n'exclut pas une allure de loi normale, nous décidons de le garder pour effectuer la moyenne.

4.2.2.2. Comparaisons avec les jugements des vingt stimuli

La procédure d'analyse est la même que celle réalisée pour le test n° 2 : nous comparons les valeurs des jugements globaux à des statistiques réalisées sur les jugements des vingt stimuli. Les statistiques sont les suivantes : Maximum, Minimum, Médiane (centile 50%), Moyenne, Centile 10% (moyenne de 10% des valeurs maximales ce qui revient à dire que 10% des réponses sont au-dessus de cette valeur), Centile 90% (moyenne de 90% des valeurs maximales, ou bien 90% des réponses sont au dessus de cette valeur).

La figure 52 montre que pour la sonie et le désagrément, les valeurs globales sont proches des valeurs moyennes ou médianes par stimuli tout en étant légèrement supérieures. Ces résultats diffèrent de ceux trouvés pour la prégnance, la présence et la proximité des sources ainsi que des résultats issus de la littérature en général. Nous faisons ici référence au paragraphe 4.1.3.

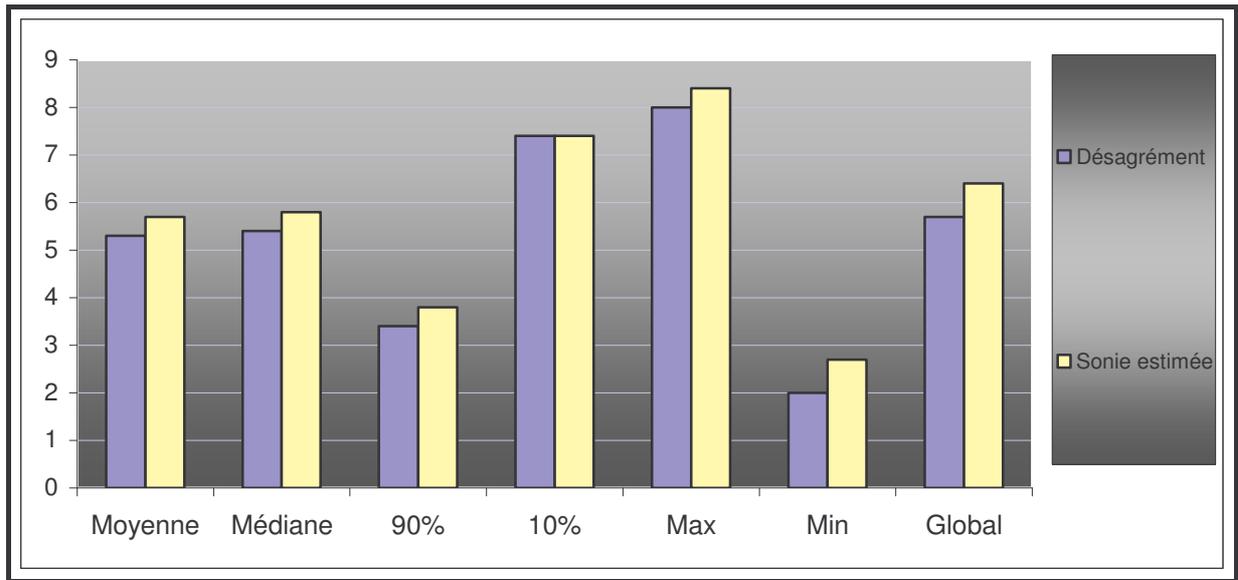


Figure 52 : Graphique représentant les valeurs statistiques (moyenne, médiane, centile 90%, centile 10%, maximum, minimum) calculées sur les vingt stimuli ainsi que les notes globales (à droite des graphiques) pour la sonie et le désagrément sonore. Cet histogramme montre que les notes globales sont proches de la moyenne des valeurs par stimulus.

4.2.3. Effet de récence :

La figure 53 montre les valeurs de la sonie et du désagrément par stimulus ainsi que les jugements globaux. Comme nous l'avons déjà indiqué au paragraphe précédent, l'effet de récence correspond à « un processus qui favorise le stockage à court terme des derniers événements entendus » [SUSINI-05].

A l'observation du graphique, il y a effectivement une coïncidence entre les estimations du stimulus « 11 » (qui est le dernier à apparaître sur la séquence globale) et les estimations de la séquence de 5 min. Il est toutefois difficile d'affirmer, à partir d'une unique séquence, qu'il y a eu un effet de récence qui aurait significativement influencé les jugements.

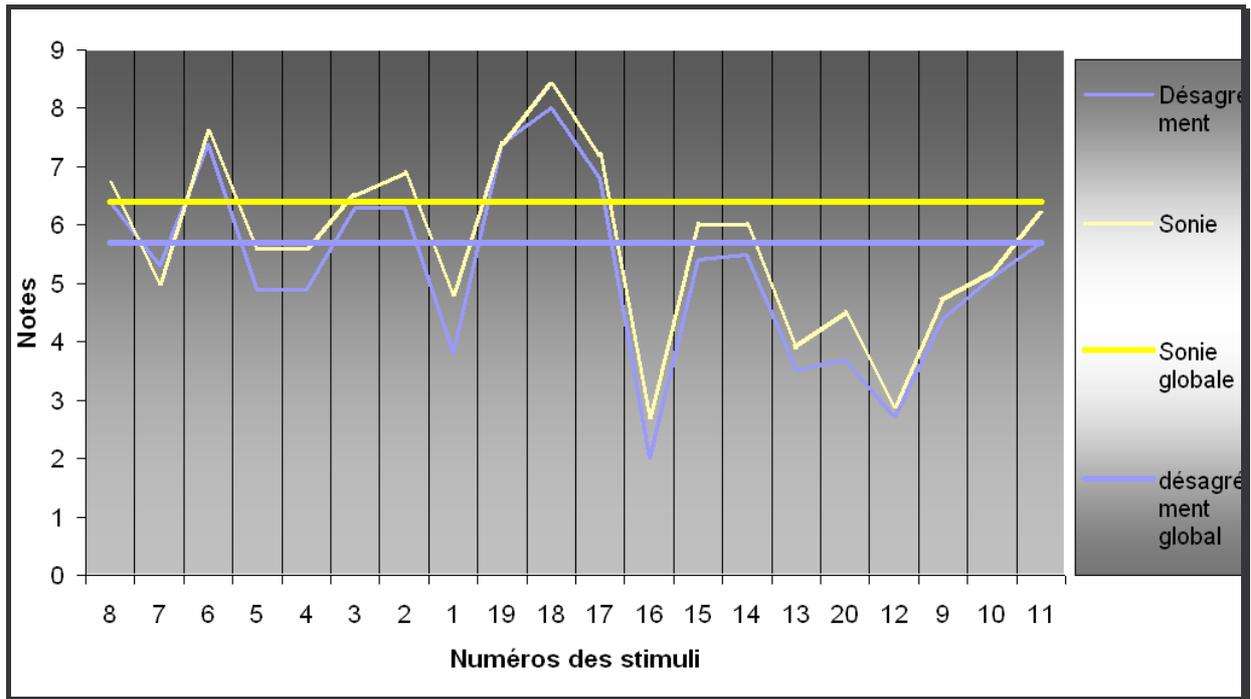


Figure 53 : Graphique représentant les valeurs de la sonie et du désagrément pour les vingt stimuli (dans l'ordre dans lesquels ils apparaissent dans la séquence globale) ainsi que pour les valeurs globales. Les notes globales sont quasiment identiques aux notes du dernier stimulus (11) indiquant un possible effet de récence.

Plusieurs remarques et hypothèses peuvent être formulées pour tenter d'expliquer ces résultats :

Tout d'abord, les notes de sonie sont, dans tous les cas, plus importantes que les notes de désagrément. Cela renforce la cohérence des résultats.

Etant donné que les jugements ont été effectués sur une unique séquence, nous pouvons suggérer que les auditeurs sont restés « prudents » dans leurs réponses et ont ainsi mis des valeurs autour de la moyenne, c'est-à-dire cinq ou six.

La figure 53 semble indiquer un possible effet de récence dans le jugement des auditeurs. Il faut donc faire d'autres tests pour le vérifier.

Nous remarquons également que ces résultats diffèrent de ceux obtenus avec la prégnance, la présence et la proximité des sources. En effet, les valeurs globales étaient plutôt reliées aux valeurs maximales des stimuli. Il semble donc que les jugements globaux d'événements reliés à des sources sur des séquences de plusieurs minutes soient plus enclins à être dépendants des

événements marquants que des jugements reliés à des impressions de force sonore ou de désagrément.

Bien sur, il convient de réaliser d'autres tests sur des séquences de plusieurs minutes pour répondre aux questions soulevées par ce test.

4.3. Jugement global d'une séquence réalisée in-situ

Lors de l'étude réalisée in situ (voir paragraphe 3.2), les sujets ont également eu à évaluer leur environnement sonore sur plusieurs séquences (d'une durée d'environ 1 min 45) ainsi que sur la globalité du parcours (d'une durée d'environ 45 min).

La même analyse que celle décrite au paragraphe 4.1.3 a été réalisée, à savoir la comparaison entre les notes globales et une étude centile sur les notes mises par séquences.

Ainsi, les relations entre l'évaluation globale de chaque variable en fonction de celle à court terme montre que :

- Pour les sources sonores de type véhicule, les auditeurs ont fait leurs estimations globales (présence, prégnance, proximité) en fonction des événements marquants du parcours. L'évaluation globale des variables s'orientent vers les notes maximales, autour de la note dépassée pendant 10% du temps.
- Pour les sources sonores naturelles (voix et oiseaux), l'évaluation globale des variables s'oriente vers une note plus médiane, autour de la note dépassée pendant 30% du temps.
- Pour la force sonore, l'évaluation globale s'oriente vers une note dépassée de 30% du temps.
- L'évaluation du caractère agréable de la globalité du parcours s'oriente vers une valeur supérieure à la médiane (environ vers la note dépassée pendant 30% du temps).

De plus, l'étude montre que les dernières séquences n'ont pas influencé de manière significative les notations globales de désagrément. Bien que la notation globale ait été jugée de manière générale, plus agréable par le groupe finissant par le parc que par le groupe finissant par le marché (ce qui pourrait indiquer un possible effet de récence dû au sens du parcours) [Figure 54], cette influence n'est pas détectée comme statistiquement significative au niveau de confiance de 95%.

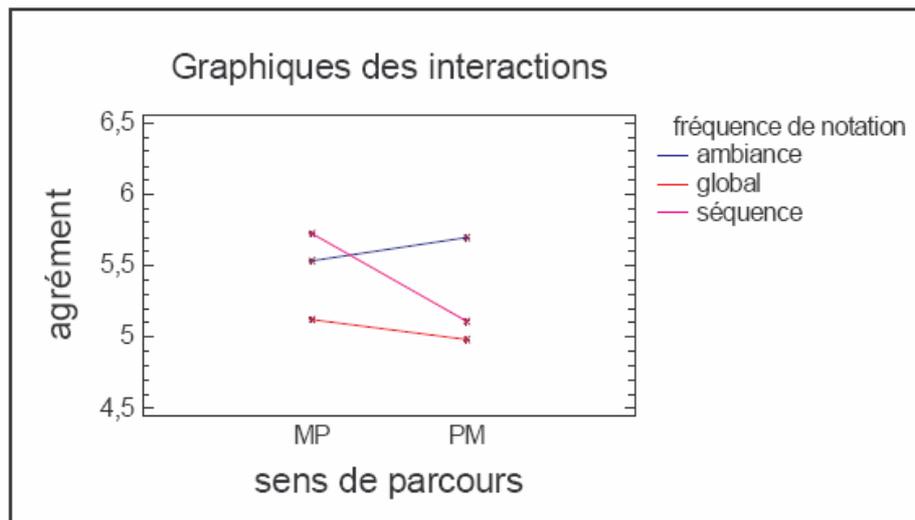


Figure 54 : Analyse de variance étudiant l'influence du sens du parcours sur les notes de désagrément (MP = du marché au parc, PM = du parc au marché). L'influence n'est pas statistiquement significative au niveau de confiance de 95%.

Ces résultats renforcent les tendances observées lors des tests en laboratoires : pour les sources sonores, l'évaluation globale se fait sur les éléments marquants alors que pour la force sonore et le désagrément, la notation globale est plus proche des valeurs moyennes.

De plus, la composition des derniers instants n'est pas détectée comme étant un facteur qui influence de manière significative la perception globale des séquences.

V. Conclusions du chapitre 2

L'objet du chapitre II est l'étude de la perception sonore de l'environnement urbain du point de vue de sa composition, c'est-à-dire des sources sonores occurrentes.

Nous avons donc réalisé un ensemble de tests psychoacoustiques sur des stimuli caractéristiques d'une ambiance urbaine : rue moyennement circulée, présence, par intermittence, de voitures, de bus, de cyclomoteurs, de motos, de voix d'enfants, de voix d'adultes et de chants d'oiseaux.

D'un point de vue perceptif, nous avons déterminé un espace tridimensionnel à partir duquel nous avons recueilli des jugements auditifs. Ces jugements portent sur les sources sonores entendues et précitées. L'espace subjectif tridimensionnel permet d'analyser la perception que les auditeurs ont :

- De la prégnance, présentée comme étant l'importance qu'une source prend dans son environnement,
- De la présence, présentée comme ayant attrait à la durée d'apparition de la source,
- De la proximité qui est en quelque sorte la distance relative de la source à l'auditeur.

Une analyse physique du L_{Aeq} codé par source sonore est également réalisée. Nous obtenons ainsi des critères temporels et énergétiques pour chacune des sources présentes dans les stimuli.

Les analyses statistiques par analyses factorielles ainsi que le croisement des données objectives et subjectives nous ont permis d'obtenir des résultats dont les principaux points sont résumés ci-dessous :

- La sensation de prégnance des voitures est corrélée aux mesures temporelles alors que celle des autres véhicules (bus, motos et cyclomoteurs) est corrélée à des mesures de niveaux. De plus, le meilleur des trois indicateurs permettant de différencier les stimuli contenant des voitures est la proximité. Ainsi en établissant une comparaison entre les vingt stimuli qui leur sont présentés, les auditeurs ne perçoivent qu'une différence toute relative sur la durée d'apparition des voitures ou sur leur puissance. Ces derniers seraient donc sensibles à une représentation spatiale des voitures plutôt que temporelle ou énergétique.

- Ces observations tendent à montrer que, contrairement aux autres véhicules précités, considérés comme des « objets bruits » car émergeant clairement du bruit de fond, les voitures sont perçues comme faisant partie intégrante de la route.
- Pour l'ensemble des catégories de véhicules établies (voitures, motos, bus et cyclomoteurs), la sensation de présence est proportionnellement liée au nombre de véhicules. Nous obtenons ainsi une bonne correspondance entre un paramètre subjectif et son corollaire objectif.
- La sensation de présence des oiseaux et des voix est correctement liée au critère temporel qu'est la durée d'apparition. Cependant aucune mesure objective ne décrit correctement leur prégnance. En effet, si les auditeurs ont pu correctement évaluer leur « importance » dans l'ensemble des stimuli, il n'est pas possible d'extraire une mesure énergétique propre à ces sources, à partir d'un codage manuel des courbes de L_{Aeq} . Cette étude met donc en évidence l'insuffisance de l'utilisation du L_{Aeq} pour mesurer l'énergie de sources dites « intermittentes » car souvent superposées à des véhicules.
- Nous avons également mis en évidence une « distorsion » entre des observations objectives et des appréciations subjectives. En effet, si l'on constate que dans l'ensemble des stimuli, les voitures sont les sources qui obtiennent la plus grande variabilité en termes de nombre, d'intensité et de temps de présence, il n'est pas de même dans les jugements recueillis. Les axes principaux formés par les analyses factorielles reflètent une variabilité importante des jugements concernant la prégnance, la présence et la proximité des bus, des cyclomoteurs, des motos et des voix d'enfants. Les auditeurs se sont donc basés sur ces sources pour différencier les vingt stimuli qui leur ont été présentés. Une conséquence pratique que montrent ces observations est que les auditeurs sont plus sensibles à une variation entre « 0 » et « 2 » bus qu'à une variation entre « 0 » et « 5 » voitures.

Les conclusions formulées ci-dessus sont traduites dans la construction de modèles de régression linéaire. Ces modèles représentent des outils ayant pour but de prédire le désagrément en milieu urbain.

Nos expériences ont montré et confirmé que le calcul de la sonie est un bon indicateur pour expliquer le désagrément lorsque nous sommes dans une situation où le bruit de trafic est

dominant et où il s'apparente à un bruit de fond. Seulement, lorsqu'il y a des événements marquants, il y a un écart qui se produit entre le désagrément et la sonie. Nous avons montré que l'on peut « combler » ces écarts en introduisant des variables (objectives ou subjectives) liées à ces événements marquants. Pour le modèle subjectif les variables sont : la présence des bus, la présence des motos et la prégnance des voix d'enfants. Pour le modèle objectif, les variables sont : le nombre de bus, le nombre de cyclomoteurs et le pourcentage de temps de présence des voix d'enfants. Nous remarquons que ces sources, qui interviennent donc comme variables dans les modèles, ressortaient déjà lors des analyses factorielles. En effet, elles obtenaient le plus de consensus et étaient le reflet d'une sensibilité particulière des auditeurs à leur écoute.

L'étude menée in situ apporte des éléments complémentaires à ces observations.

- Dans une rue circulée, les voitures sont perçues comme du bruit de fond. Le désagrément est alors modélisable par l'évaluation de la sonie moyenne à laquelle on rajoute les effets des autres sources.
- Dans d'autres types d'ambiances sonores comme un parc ou un marché, les voitures deviennent des objets et ainsi la sonie n'est pas adaptée pour évaluer le désagrément à court terme. Dans de tels environnements, le désagrément est lié uniquement aux sources, les voitures étant prises en compte comme des événements.

Il est également important de signaler que nos modèles sont les reflets de la composition des stimuli présentés. En conséquence, ils tiennent compte des variables qui y sont apparues. Ces dernières ne sont donc pas représentatives de l'ensemble des configurations (en termes de sources sonores) que nous puissions rencontrer dans une rue moyennement circulée. C'est pourquoi il est important de tenir compte des domaines de validité des modèles. Pour cela, nous avons montré qu'il est possible de construire des marges d'erreurs respectables (à un écart de plus ou moins un de la valeur prédite) à des niveaux de confiance de 95%.

Enfin, l'étude sur la perception des séquences plus longues, en laboratoire comme in-situ, a également apporté des éléments utiles quant à la prédiction de la qualité de l'environnement sonore. Il a été montré que l'évaluation de la sonie et du désagrément sonore sur des séquences longues pouvait être rapportée à une moyenne alors que la perception globale des dimensions relatives aux sources était plutôt liée aux événements marquants de cette séquence. Ainsi, il est possible, sur ces principes, d'évaluer le désagrément sonore sur des séquences de 5 min à partir de nos modèles extraits pour des séquences de 15 s.

CHAPITRE 3. ETUDES POUR L'IDENTIFICATION AUTOMATIQUE DES SOURCES SONORES

I. Pourquoi l'identification ?

Dans l'étude du chapitre précédent, à travers une série de tests psychoacoustiques sur un ensemble de séquences représentant une ambiance urbaine de quartier, deux éléments ont été mis en évidence :

- La perception hédonique d'un environnement sonore urbain de ce type peut être évaluée selon deux aspects complémentaires.
- Ces deux principaux aspects permettent d'utiliser de façon optimale un modèle de type régression linéaire multiple pour expliquer la qualité de cet environnement sonore.

Le **premier aspect** est lié à la force sonore.

Le **second aspect** traduit l'influence des sources sonores occurrentes et émergentes.

Une étude in situ a également montré que ces deux aspects n'ont pas le même poids selon l'ambiance sonore. Pour illustration, le tableau 24 indique les pourcentages de variance des jugements de désagrément expliqués par les différentes variables selon les trois lieux étudiés dans ce travail : parc, rue et marché.

		Variables subjectives		
		Force sonore	Sources sonores	Force sonore + sources sonores
Ambiance	Rue	85 à 90%	86%	89.1% à 98.7%
	Parc	34.4%	83.3%	95.5%
	Marché	0.0%	80.0%	80.0%

Tableau 24 : Pourcentages de variance des jugements de désagrément expliqués par les différentes variables selon les trois lieux étudiés dans ce travail : parc, rue et marché.

L'évaluation chiffrée du désagrément à partir d'un environnement sonore existant nécessite de transcrire ces variables objectives à partir de mesures physiques.

La transcription objective de la force sonore est réalisée de manière satisfaisante par la mesure physique de la sonie (N) selon le modèle de Zwicker, ou plus exactement de la sonie moyenne. De surcroît, il existe des logiciels qui réalisent ce calcul (par exemple dBFa32 de 01dB Acoustics&Vibration que nous avons utilisé).

Concernant les variables ayant trait aux sources sonores ; nous avons vu que nous pouvons les transcrire par l'extraction de données temporelles relatives aux sources. Ces données sont disponibles à partir d'un codage sur la courbe d'évolution du niveau de pression L_{Aeq} . Ce codage est cependant fastidieux pour l'expérimentateur car il nécessite la réécoute des stimuli et le codage par source.

Si l'on veut pouvoir utiliser les modèles extraits de façon automatique, il est nécessaire de permettre le codage des sources sans intervention manuelle et de façon rapide.

Ainsi, ce troisième et dernier chapitre expose deux pistes pour identifier automatiquement les principales sources sonores de l'environnement urbain.

Nous précisons que, même si ce n'est pas indiqué de façon systématique par la suite, le travail d'identification entrepris porte sur l'identification des *passages* de véhicules. Nous excluons l'identification des véhicules à l'arrêt ou au démarrage par exemple. Nous ne faisons évidemment pas ces distinctions pour les voix et les oiseaux. De même, nous désignons par cyclomoteur la catégorie de véhicules à deux roues motorisés ayant une cylindrée inférieure à 125 cm^3 : par exemple les scooters et les cyclomoteurs. La notation abrégée de la catégorie cyclomoteur est Mob. Les motos sont quant à elles caractérisées par une cylindrée supérieure à 125 cm^3 .

La première étude est centrée sur le concept d'*activité* d'une source. A partir de la mesure de la distance euclidienne entre une représentation spectrale type d'une source et des spectres mesurés, nous avons imaginé pouvoir évaluer la contribution énergétique de chaque source sonore. Cette routine est relativement simple. De ce fait, elle a l'avantage de pouvoir facilement fonctionner en temps réel mais sa robustesse aux changements d'états types n'est pas garantie.

La seconde étude est basée selon une démarche *heuristique* dans le sens où nous procédons par évaluations successives (bases de données sonores) et hypothèses provisoires (assignation arbitraire d'un nom de classe aux sources sonores). Cette recherche, ainsi fondée sur une approche progressive du problème donné, est le résultat de l'utilisation d'un outil informatique, EDS (Extractor Discovery System) [ZILS-04]. Celui-ci permet de caractériser un type d'événement sonore par un modèle faisant appel à des assemblages d'opérateurs de traitement du signal. Ce système fonctionne selon une recherche supervisée, c'est-à-dire que lui sont indiqués à l'avance (par étiquetage) les noms des classes correspondant aux événements sonores. La méthodologie employée est de partir d'une classification préétablie (cyclomoteurs, bus....) et d'interpréter les résultats d'EDS en vue d'établir une nouvelle typologie. Des regroupements et des séparations de classes vont donc être effectués tout au long de ce travail, au gré de la pertinence des résultats obtenus par le système.

II. L'identification des sources ou événements sonores

2.1. Théorie

Nous avons vu en introduction que l'une des fonctions primaires du système auditif est de nous renseigner sur notre environnement pour pouvoir agir en conséquence.

Une fois l'information sonore disponible, l'identification est donc l'une des premières tâches que nous accomplissons. Dans le cadre des études menées actuellement sur le sujet qui ont pour bases théoriques le traitement de l'information, nous allons voir quelles sont les différentes étapes de traitement qui permettent au système auditif d'identifier la nature d'un objet émetteur de son.

2.1.1. Les étapes de l'identification

La théorie cognitive décompose le traitement de l'information auditive en étapes et en explique les liens. Stephen McAdams [MCADAMS-94] dans un chapitre de l'ouvrage « Penser les Sons », expose une bonne partie des recherches ayant trait à cette théorie de l'information en la décomposant en « étapes hypothétiques ». Nous tentons ici d'en reprendre les grandes lignes.

On peut grouper les étapes d'une identification en deux groupes principaux. Chacun de ces groupes étant lui-même décomposé en processus élémentaires [Figure 55].

Tout d'abord, le groupe des processus dits ascendants, car essentiellement fondés sur une analyse des données entrantes. Celui-ci comprend une phase de transduction sensorielle où l'information auditive est transmise au cerveau. Elle passe par le système auditif périphérique (oreille externe, moyenne et interne) et est transmise au cerveau par les fibres nerveuses qui constituent le nerf auditif. Le processus de groupement [BREGMAN-90] permet, dans le même temps, de traiter l'information auditive en réalisant automatiquement un groupement ou une séparation des flux auditifs. Ces processus sont présumés automatiques. A ce stade il n'y a pas encore d'identification, uniquement un prétraitement ou un pré-conditionnement de l'information. Concrètement, l'information sonore nous parvenant de deux véhicules passant simultanément sera divisée en deux flux distincts. Ceci de manière présumée automatique dans la mesure où l'expérience personnelle a une influence minimale sur ce conditionnement.

Vient ensuite le groupe des processus descendants. Une fois l'information traitée et les flux auditifs séparés, il faut distinguer deux analyses qui vont permettre à une personne d'identifier une source sonore.

L'analyse des *propriétés microtemporelles* consiste à extraire des propriétés de l'événement sur une fenêtre temporelle de durée assez courte et de les comparer à des *invariants* qui seraient stockés en mémoire. Notre expérience nous a ainsi rendus sensibles aux *invariants structuraux* qui nous fournissent des informations sur la structure physique de la source sonore. Par exemple, des voitures sont identifiées comme telles, quel que soit leur régime moteur ou leur largeur de pneus, uniquement parce qu'elles émettent un bruit de moteur et/ou un bruit de contact pneumatique-chaussée. Par ailleurs, un bruit de moteur est reconnaissable quel que soit le nombre de cylindres disponibles. Cela fait appel à la détection d'un *invariant transformationnel* dans le signal sonore permettant l'identification de l'événement mécanique. L'auditeur peut par exemple utiliser cet invariant pour lui permettre d'établir un diagnostic sur le fonctionnement du moteur.

La seconde analyse qui permet à un auditeur l'identification d'une source considère une fenêtre temporelle plus large. En conséquence elle permet de s'attacher « au rythme et à la texture de l'événement sonore » [MCADAMS-94]. C'est l'analyse des *propriétés macrotemporelles* qui permettent d'indiquer une modification de la source. Dans le cas des sources mécaniques, la modification du niveau sonore d'un véhicule (conjuguée à l'effet Doppler, conséquence de son mouvement), est un indice qui peut s'avérer utile pour différencier, par exemple, le bruit d'un cyclomoteur de celle d'une tronçonneuse.

Après avoir structuré le signal sonore et extrait des propriétés pertinentes [Figure 55], l'étape suivante est l'*ajustement au lexique auditif*. À ce stade, le sujet va « comparer » les informations reçues et retenir la configuration la plus probable permettant ainsi la reconnaissance de l'événement sonore. Plusieurs thèses existent à ce point sur le déroulement de cette étape¹.

Enfin, le dernier traitement est celui qui nous fait passer du stade de la reconnaissance à celui de l'identification lorsque nous pouvons nommer la source ou l'événement sonore. L'activation lexicale entraîne dès lors l'activation d'autres effets qui ne sont plus purement auditifs.

¹ Il peut y avoir soit une comparaison avec des représentations stockées en mémoire avec retenue de la configuration la plus probable, soit l'activation directe de représentations mnésiques avec identification par dépassement d'un seuil d'activation.

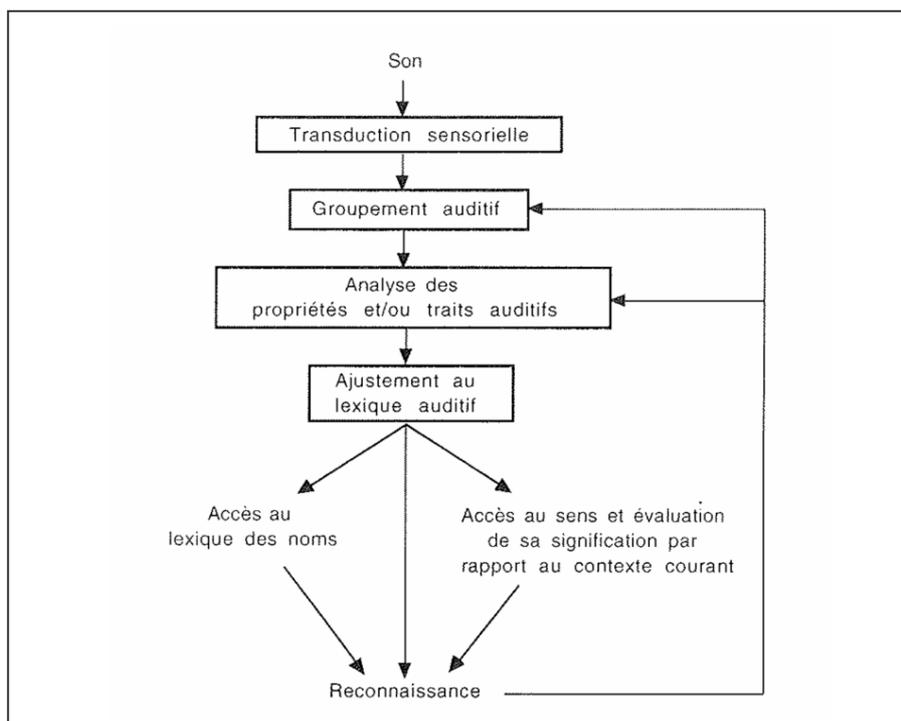


Figure 55 : Schéma des étapes hypothétiques du traitement qui interviennent dans l'identification ou la reconnaissance. La représentation sous forme d'organigramme n'implique pas forcément que le traitement de chaque type soit séquentiel. [MCADAMS-94].

2.1.2. Les facteurs cognitifs

En plus des facteurs acoustiques spectraux et temporels, des facteurs cognitifs ou « schémas mentaux » tiennent également une grande importance dans la tâche d'identification des sources. Ces facteurs ont été définis en faisant l'hypothèse que les schémas de la reconnaissance des sons de l'environnement sont comparables aux schémas de traitements de la reconnaissance vocale [HANDEL-89] (la reconnaissance d'un événement sonore dans son environnement se construit de la même façon que la reconnaissance d'un mot dans une phrase). Au cours des recherches effectuées dans les années quatre-vingt et quatre-vingt-dix, plusieurs concepts ont été mis en évidence.

L'*ambiguïté causale* (« causal ambiguity ») [BALLAS-93], correspond à une incertitude basée sur le nombre de différentes causes génératrices d'un même son (le miaulement d'un chat et le cri d'un bébé par exemple). C'est en quelque sorte le nombre d'homonymes sonores. Sur un corpus de sons assez courts (environ 600 ms), Ballas et ses collègues ont même trouvé que le temps nécessaire à l'identification des sons de l'environnement était en relation logarithmique avec le nombre d'alternatives causales de cette source [BALLAS-86].

Deux indices nous servent pour l'identification d'un son ayant une ambiguïté causale :

La *fréquence écologique* (« ecological frequency ») [BALLAS-87] est définie par la fréquence d'occurrence des sons. C'est un indice temporel qui nous permet, selon la période (journalière ou annuelle par exemple) d'apparition d'un événement sonore, de l'identifier.

L'*information contextuelle* [BALLAS-91] est un facteur qui nous permet, par exemple, d'identifier une séquence comme correspondant à l'explosion d'un pétard si elle est composée des sons d'une allumette qu'on gratte, de la mèche du pétard qui brûle et enfin de l'explosion ; de la même manière que l'identification correcte d'un mot dans une phrase peut dépendre des mots suivants et précédents.

Nous pouvons également ajouter à ces deux indices utiles la notion de *typicalité* et de *familiarité* du son (par exemple, il a été montré que les sons les mieux identifiés sont ceux de la catégorie « sons humains », suivis par les sons d'instruments, les sons d'objets et enfin les vocalisations animales [LASS-82]).

Brian Gygi [GYGI-01] a testé l'habilité du système auditif à identifier les sons environnementaux en réduisant leurs informations spectrales et temporelles (par filtrages et modulations qui conservent l'enveloppe temporelle d'un son mais en conservant un spectre quasi uniforme). Il apparaît que les capacités d'identification de ces sons sont très robustes à tous les changements précités. L'information spectrale utile se situe, pour la plupart des soixante-dix sons employés, dans les régions des moyennes et hautes fréquences (de 600 à 4800 Hz). Les sons traités les mieux reconnus sont ceux qui ont de fortes caractéristiques temporelles (hélicoptère, montre, galop) alors que ceux caractérisés par leur hauteur (oiseau, flûte) et les sons stationnaires (la scie électrique par exemple) sont plus difficiles à identifier. De plus, il ressort que les auditeurs utilisent les informations temporelles d'un son, en l'absence d'information spectrale. Il est ainsi difficile de tirer des lois générales quant aux informations spectrales et temporelles qui nous servent à l'identification. Une combinaison de variables descriptives des propriétés spectrales et temporelles des sons (nombres de pics, moyennes spectrales, énergie...) permet d'expliquer (seulement) jusqu'à 60% de la dimension principale résultant d'un test de comparaisons entre les sons présentés, et l'auteur de conclure que des facteurs cognitifs de « haut-niveau » ainsi que les interactions sensibles (signification des sons) ont également une très grande importance dans une tâche d'identification.

Ballas [BALLAS-93], quant à lui, a montré que les propriétés spectrales et temporelles associées à la mesure de la *fréquence écologique* permettaient d'expliquer 75% de la variance du temps d'identification d'un panel de sons environnementaux.

2.1.3. La classification des sons

Un système de reconnaissance parfait serait celui qui reproduirait à l'identique l'habilité du système auditif humain à identifier les sons. Cette modélisation nécessiterait une connaissance approfondie des caractéristiques utiles des signaux que nous extrayons et des processus cognitifs de décision, ainsi qu'une puissance de calcul sans doute très importante.

Une façon naturelle d'identifier les sources est de s'appuyer d'abord sur une catégorisation arbitraire des sons. Cette catégorisation découle de notre perception visuelle. Intuitivement, nous allons en effet classer les sources en fonction de leur appartenance visuelle, c'est-à-dire par l'objet producteur de son. Par exemple, concernant les bruits urbains, la première classification qui s'impose est de différencier les voitures, les motos, les cyclomoteurs, les bus, les camions, les oiseaux, les voix... Mais cette différenciation est délicate pour des sources identiques visuellement mais différentes d'un point de vue sonore. Nous verrons par la suite que c'est le cas pour les motos. De même, simplement pour une voiture, le son émis sera très différent si elle est à l'arrêt, au démarrage ou en passage.

Notre travail s'appuie sur une caractérisation acoustique des sources. Si les classes qui découlent de cette caractérisation sont trop nombreuses, un regroupement des sous-classes permet de retrouver la classification des sources telles que nous les voyons.

Ainsi, la dénomination des sons ne peut se restreindre à une seule méthode. Traditionnellement, on dénomme les sons par leur origine : le bruit de la porte, la parole du sage etc, mais les sources d'inspirations peuvent être multiples et variées : apparentées aux ressemblances animales, actions, mécanismes, formes...

Les auteurs peuvent avoir ainsi différentes approches, selon la finalité de l'étude sur laquelle se base cette classification :

L'approche musicale :

Pour établir une classification des objets sonores, Schaeffer [SCHAEFFER-66] a tenté d'établir tout d'abord une typologie basée sur la morphologie des sons. Celle-ci est définie sur sept critères de valeurs :

- 1) la masse : organisation du son dans la dimension spectrale ;
- 2) la dynamique : description de l'intensité des différentes composantes du son ;
- 3) le timbre harmonique : qualités particulières et « couleur » du son ;
- 4) le profil mélodique : évolution temporelle du spectre global du son ;
- 5) le profil de masse : évolution temporelle des composantes spectrales internes du son ;
- 6) le grain : analyse des irrégularités de surface du son ;
- 7) l'allure : analyse des vibratos (de hauteur et de dynamique) du son.

L'approche « écologique » : ainsi dénommée car elle tient compte de l'information que produisent les sons. En effet, de son côté, Gaver [GAVER-89] a choisi d'établir une taxinomie des sons de l'environnement selon le type d'interaction dont ils proviennent. Il classe ces événements sonores selon une hiérarchie (catégorielle) et relie ensuite ces interactions avec des descriptions acoustiques des ondes sonores. [Figure 56]

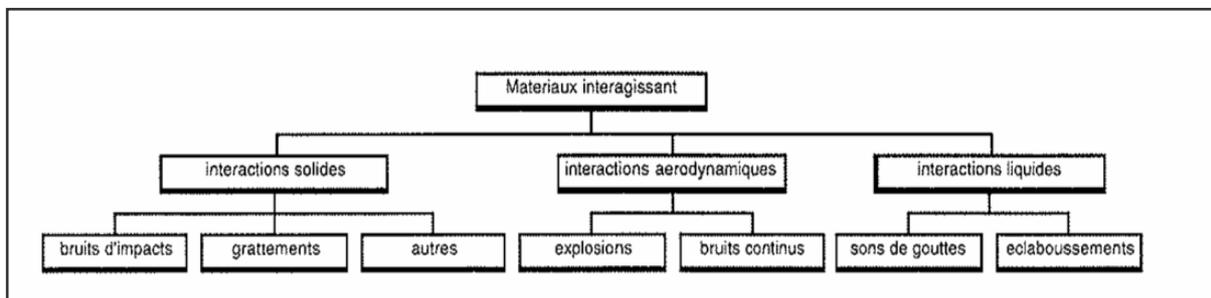


Figure 56 : Description hiérarchique des événements sonores selon Gaver.

Enfin, Léobon [LEOBON-95] a répertorié l'ensemble des sources acoustiques de l'environnement sonore urbain dans un nombre limité de catégories de sources ayant une importance perceptive « identique ». Six types de sources sont ainsi dégagés [MAFIOLLO-99] :

- Bruit de fond : quantifie indirectement les notions de calme et de faux silence.
- Activités mécaniques : révèle les bruits de moteurs, de circulation automobile, les bruits de transports et les bruits de chantiers de construction.

- Activités humaines : signe des bruits liés aux activités quotidiennes des individus, aux actes qui ne révèlent pas de simple fonction de passage, aux loisirs, etc.
- Bruits d'animaux et de la nature : se réfèrent généralement aux espaces paysagers (également au bruit du tonnerre ou de la pluie).
- Présence humaine : relative aux indices de passage humain (voix, pas, etc.).
- Langage et communication : rendent compte des messages sonores très interpellant (signalétiques, animations musicales, conversations intelligibles, etc.).

Les trois classifications décrites ci-dessus sont des classifications perceptives car elles découlent de différents types d'écoutes. Ces écoutes sont orientées selon les besoins de l'auditeur et de ce fait suivant les informations qui lui sont utiles.

La première approche présentée s'appuie sur une description précise des caractéristiques spectrales et temporelles des sons pour décrire les objets sonores. Bien que cette dernière ait été établie en vue d'une représentation structurelle des sons de l'environnement, l'objet est séparé de son contexte naturel de production et est placé dans un contexte d'écoute musicale. L'écoute diffère de notre propos, le mode de perception également ; la nature même des sources sonores ou leur identité découlant d'une telle classification risque d'être occultée.

La seconde approche est intéressante dans le sens où elle organise les sons de manière hiérarchique. De plus elle s'appuie sur des informations acoustiques présentes dans les sons, informations qualifiées de « bas-niveau » que nous pourrions extraire. Ainsi, nous pourrions imaginer d'orienter les recherches de sources en partant des interactions détectées. Cependant, dans l'environnement sonore urbain, de nombreux sons peuvent faire intervenir plusieurs de ces interactions (nous avons déjà cité le son d'une voiture).

La dernière classification ne s'appuie pas assez sur des considérations physiques des sons. Or ce sont les seuls paramètres que nous sommes en mesure d'évaluer ici.

Ainsi il n'y a pas, à notre connaissance, de classification déjà établie dont les caractéristiques pourraient nous servir de base de départ à l'étude d'identification.

Notre but ici est bien d'identifier les sources que nous percevons en milieu urbain et qui entrent en compte dans nos modèles. Ce sont donc celles-ci qui forment le point de départ de notre taxinomie. Cette taxinomie s'appuie sur les similitudes spectro-temporelle des sources sonores.

2.2. Applications aux sons de l'environnement urbain

2.2.1. Des méthodes aux applications

Les recherches sur la reconnaissance des sources sonores telles que la voix d'un locuteur ou des instruments de musique sont relativement abouties. Bien que les systèmes de reconnaissance ne modélisent pas encore les effets cognitifs précités, il est désormais possible d'identifier ces sources en raison d'une connaissance approfondie des traits pertinents de leur signature sonore. Les systèmes d'identification les plus performants se réfèrent aux caractéristiques des structures micro et macro temporelles des sons. Ces systèmes utilisent ces caractéristiques en entrée de modèles ou classificateurs (réseaux de neurones, chaînes de Markov...). Ils fonctionnent par apprentissage sur de larges bases de données (généralement plusieurs milliers de séquences sonores). Les conditions requises pour une bonne identification sont d'une part une caractérisation robuste des signaux et d'autre part la mise à disposition d'une base de données assez grande. Rappelons que tous les systèmes n'apprennent uniquement que ce qu'on leur donne à apprendre (bases de données) et ce qu'on leur demande de chercher (caractéristiques physiques).

Ce niveau d'expertise a été rendu possible grâce aux nombreuses applications commerciales : multimédias, robotique, téléphonie...

Peu de recherches ont été entreprises concernant les sons de l'environnement et encore moins sur des ambiances dans lesquelles les sons sont inévitablement mixés avec d'autres bruits. De plus, la tâche est rendue d'autant plus difficile qu'il existe une grande variété de signaux possibles pour une même catégorie d'événements sonores.

Techniquement, la quasi-totalité des modèles exposés dans la littérature [COWLING-03], [TOYODA-04], [ZHANG-98], [GOLDHOR-93], [ZILS-04], [DUFAUX-00] et ayant pour but l'identification des sons de l'environnement (qui sont des sons pris isolément) se font sur un schéma identique, déjà éprouvé pour la reconnaissance vocale et d'instruments de musique.

Ils s'appuient sur l'utilisation de classificateurs automatiques. Les *classificateurs supervisés* sont ceux couramment employés. Ces systèmes sont pré-orientés dans le sens où on leur indique au préalable les noms des sources recherchées, contrairement au *classificateur non –supervisé* dans lequel les systèmes assignent eux-mêmes au son sa classe d'appartenance.

Dans la littérature, les systèmes utilisés pour l'identification des sons de l'environnement s'appuient sur une classification supervisée.

2.2.2. La classification supervisée :

Elle se déroule en deux étapes principales. Il y a tout d'abord une recherche de traits pertinents des signaux classés grâce à l'analyse des propriétés spectrales et temporelles (soit par décomposition en séries de Fourier, soit par décomposition en ondelettes) suivie d'une phase de classification automatique utilisant les techniques d'intelligence artificielle.

Pour être fiables, les traits doivent désigner exclusivement la source recherchée. Dans ce sens, on utilise des opérateurs de traitement du signal qui manipulent les sons et permettent d'en dégager des caractéristiques propres. La phase de classification quant à elle, est utilisée pour reconnaître le son en cataloguant les traits des sons existants d'une certaine manière (entraînement ou apprentissage) pour ensuite confronter des sons tests à cette base de données de tests.

Il semble, selon les auteurs cités au paragraphe précédent, que les résultats obtenus pour l'identification des sons de l'environnement sont satisfaisants. Toutefois, ces conclusions sont issues de tests effectués sur des bases de sons isolés. Or nous savons que de tels sons ne se produisent que très rarement de façon individuelle. Si ces systèmes fonctionnent de manière satisfaisante sur de tels sons, leur utilité dans le cadre de nos recherches est donc limitée car les séquences que nous utilisons contiennent du *bruit*. L'absence de bibliographie sur ce sujet, ajoutée à l'inexistence de systèmes ou logiciels commercialisés, semble indiquer que les algorithmes pour l'instant employés pour la reconnaissance de source en environnement urbain ne sont pas satisfaisants.

Dans ce sens, notre travail consiste à proposer des pistes originales pour l'identification de sources en milieu naturel.

Dans un premier temps nous mettons au point un modèle simple qui ne s'appuie pas sur l'apprentissage et qui traduit l'activité des sources sonores.

Dans un second temps, nous exposons les résultats obtenus par l'utilisation d'opérateurs de traitement du signal réalisant une analyse fine des variations spectrales et temporelles des sources.

III. Activité des Sources Sonores – A2S

3.1. Principe

L'idée, derrière cette méthode, est d'obtenir un système simple pouvant indiquer la présence de sources sonores. Une fois cette présence détectée, cette méthode doit nous permettre d'évaluer la contribution de chaque source.

Le postulat, derrière cette recherche entreprise et concrétisée sous forme de logiciel (que nous avons appelé *A2S*), est que l'on peut trouver suffisamment de traits caractéristiques dans le spectre de chacune des sources urbaines pour qu'il soit possible de les différencier les unes des autres.

Ces traits caractéristiques nous permettraient alors d'effectuer des identifications mais également de pouvoir évaluer la contribution énergétique de chaque source : c'est le principe de l'activité des sources sonores. Cette contribution énergétique particulière pourrait de ce fait être reliée au caractère prégnant de la source qui, rappelons-le, a été défini au chapitre précédent comme étant l'« importance » que prend une source dans son environnement.

La partie développement informatique s'est faite en collaboration avec la société 01dB Acoustics&Vibration. La base du travail (acquisition, filtres numériques, interface graphique) a été un prototype dédié à l'identification des termites du bois [BARTHET-03]. Ce logiciel est utile pour caractériser l'activité d'une source de bruit unique, aux variations spectrales faibles et définies par son émergence par rapport à un bruit de fond.

Pour l'adapter à des sources de nature variable et de d'intensités non stationnaires, nous avons profondément transformé le logiciel en modifiant son principe de base. Nous avons observé les caractéristiques spectro-temporelles des sources sonores et imaginé en conséquence des opérations à effectuer pour identifier ces dernières.

Ces opérations consistent en deux routines principales: la *ressemblance spectrale* et la *détection de pics*. Nous pouvons ensuite en ajuster les paramètres dans des fenêtres de contrôles.

Pour des raisons de clarté de l'exposé, nous décrivons en premier lieu les routines implémentées pour ensuite décrire l'analyse spectrale et temporelle des sources sonores. Chronologiquement, notre travail a d'abord consisté à réaliser l'étude acoustique des sources puis, à travers leur analyse (différences, ressemblances), nous avons déduit les deux routines utilisant ces spécificités.

Pour aboutir à une représentation mentale d'une information sonore, l'oreille interne filtre le signal acoustique (en temps réel et avec une sélectivité fréquentielle considérable grâce à l'organe de Corti) et réalise une analyse fréquentielle et temporelle de façon à fournir au cortex auditif des informations telles que nous pouvons les observer sur un sonagramme [MOORE-97].

La première opération d'A2S repose sur un principe de modélisation de la tâche que le système auditif aurait à accomplir à partir des composantes fréquentielles obtenues en fonction du temps. Plus précisément, il réalise une mesure de la qualité d'ajustement entre un gabarit spectral type (qui est en quelque sorte une forme analytique de la représentation mnésique) et les spectres auditifs des entrées. Nous l'appelons la *ressemblance spectrale*. Lors du développement de cette routine, nous nous sommes aperçus que son principe (l'ajustement) était à rapprocher du fonctionnement simplifié du modèle de Klatt utilisé dans le cadre de la reconnaissance vocale [KLATT-89]. Selon Klatt, le système de traitement de la parole compare chaque spectre à court terme avec des patterns spectraux prototypiques et recherche celui qui y correspond le mieux.

Cette routine a été imaginée comme étant un moyen d'obtenir une indication chiffrée des différences spectrales observées entre véhicules. Pour les voix et les oiseaux, qui sont caractérisés sur des largeurs spectrales plus réduites (trois ou quatre bandes de tiers d'octave), cette routine n'est plus pertinente.

En second lieu, nous avons implémenté la *détection de pics* dont la fonction est de repérer des émergences spectrales caractéristiques de certaines sources. Cette seconde routine peut, soit détecter la présence d'une source si cette caractéristique est suffisante (pour les oiseaux en particulier), soit venir « valider » la tendance dégagée par la mesure de l'activité (pour les cyclomoteurs par exemple).

La méthode générale est explicitée ci-dessous.

3.2. Routines

3.2.1. Ressemblance spectrale

On détermine des gabarits spectraux types pour chaque catégorie de source prédéterminée. Ces gabarits spectraux ont comme particularité le fait d'être non seulement caractéristiques du plus grand nombre d'événements de leur catégorie, mais également assez différents des gabarits des autres catégories. Nous exposerons dans le chapitre suivant comment ces spectres ont été déterminés. A ce stade, nous avons déterminé la finesse optimale du spectre (le spectre en tiers d'octave est retenu), en tenant compte de contraintes liées aux vitesses de calcul et à la finesse des filtrages.

Une procédure de filtrage du spectre par le gabarit type permet d'obtenir les écarts aux gabarits par bandes de tiers d'octaves. Ce spectre est appelé *spectre filtré* dont on calcule le niveau global filtré : $L_{Aeq\text{ filtré}}$. Le logiciel calcule ce spectre sur une période T par pas glissants Δt , comme indiqué sur la figure 57.

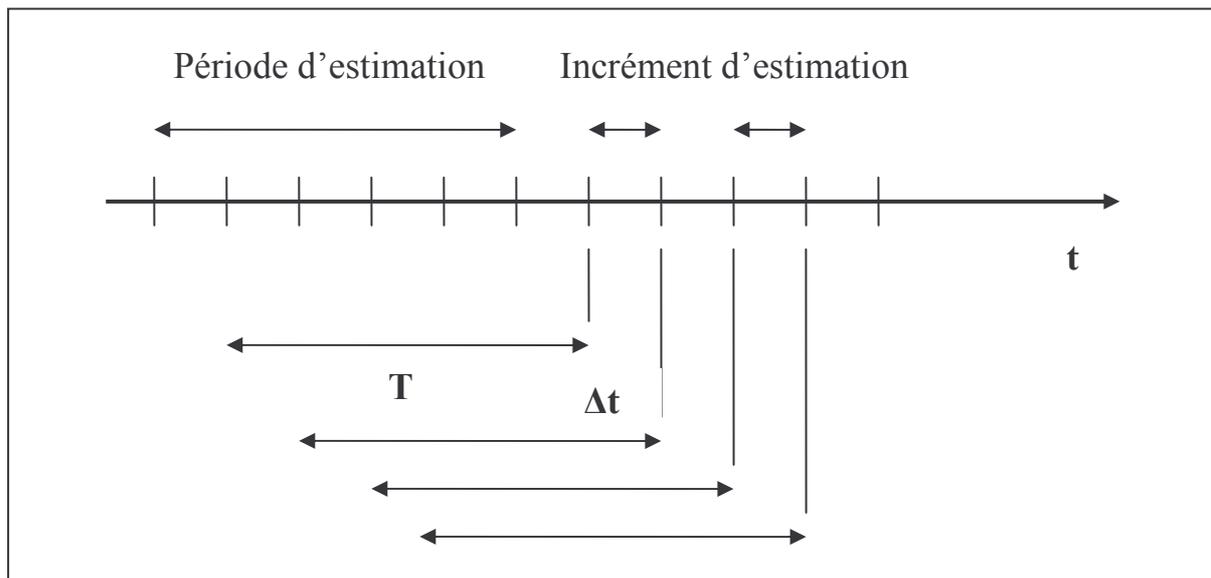


Figure 57 : Algorithme de calcul pour la procédure de filtrage (routine de ressemblance spectrale).

Ainsi la chance de rencontrer un spectre proche du gabarit est optimisée.

L'ensemble des écarts représente une mesure de la distance spectrale entre notre gabarit type recherché et la signature de l'événement occurrent. La valeur de la ressemblance spectrale brute

(rs) est déterminée comme étant l'écart type des valeurs obtenues (écarts par rapport au gabarit) par bandes de tiers d'octaves.

$$rs = \text{EcartType}O_i$$

i = bande de tiers d'octaves,

O_i = valeur de la bande de tiers d'octaves filtrée.

Cette valeur brute (rs) est ensuite inversement normalisée entre « 0 » et « 1 » de la façon suivante :

- Si $rs > 10$, alors $RS = 0$,
- Si $0 \leq rs \leq 10$, alors $RS = 1/(rs+1)$

RS = valeur de la ressemblance spectrale normalisée entre « 0 » et « 1 ».

Nous obtenons en quelque sorte la « platitude » du signal filtré. Si la valeur de ressemblance est proche de « 1 » (allure de bruit rose), cela signifie que les écarts entre le gabarit d'entrée et le gabarit mesuré sont minimales, traduisant la présence de la source recherchée, et inversement. Pour chaque catégorie de source recherchée, un seuil de confiance est fixé, au-delà duquel la présence d'une source est validée. Cette valeur est déterminée de manière empirique après la réalisation d'essais sur des séquences sonores tests.

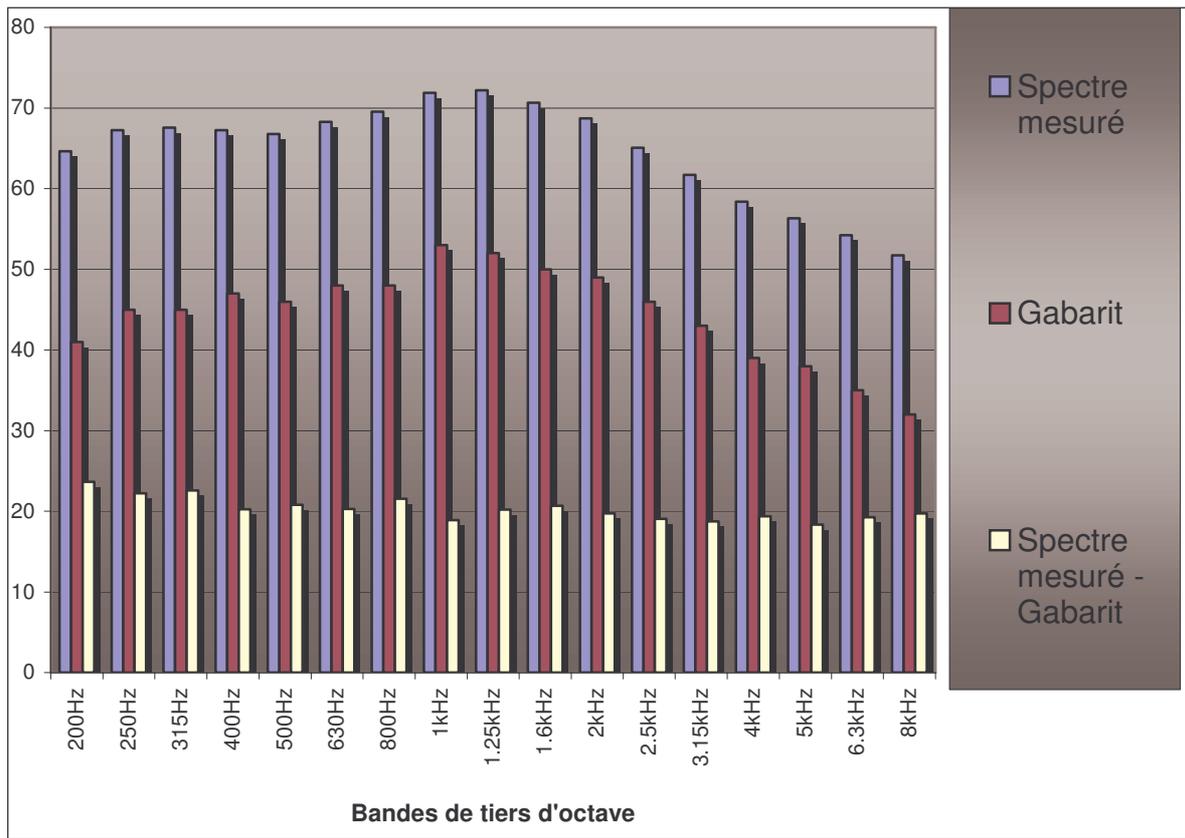


Figure 58 : Instantané du fonctionnement de la routine de ressemblance spectrale. En bleu, nous avons le spectre en bande de tiers d'octaves qui est une mesure sur une période déterminée (200 ms en général). En rouge, nous avons le spectre représentatif d'une source sonore. En jaune nous avons le résultat du filtrage du spectre mesuré par le gabarit (appelé également L_{Aeq} filtré et correspondant à l'activité de la source lorsque celle-ci est détectée).

Nous calculons ensuite la « platitude » du spectre jaune. Cette mesure nous informe sur la ressemblance entre une source caractérisée par son gabarit et l'événement sonore mesuré, elle est normalisée entre « 0 » et « 1 ».

On remarque que le niveau filtré (en jaune sur la figure 58) dépend du niveau émergent de la source identifiée.

3.2.2. Détection de pics

3.2.2.1. Principe

C'est en analysant les spectres en tiers d'octaves des cyclomoteurs et des oiseaux que nous avons décidé d'intégrer au logiciel la possibilité de détecter des émergences spectrales. Les cyclomoteurs sont caractérisés par des émergences spectrales créées par le moteur de l'engin [Figure 59].

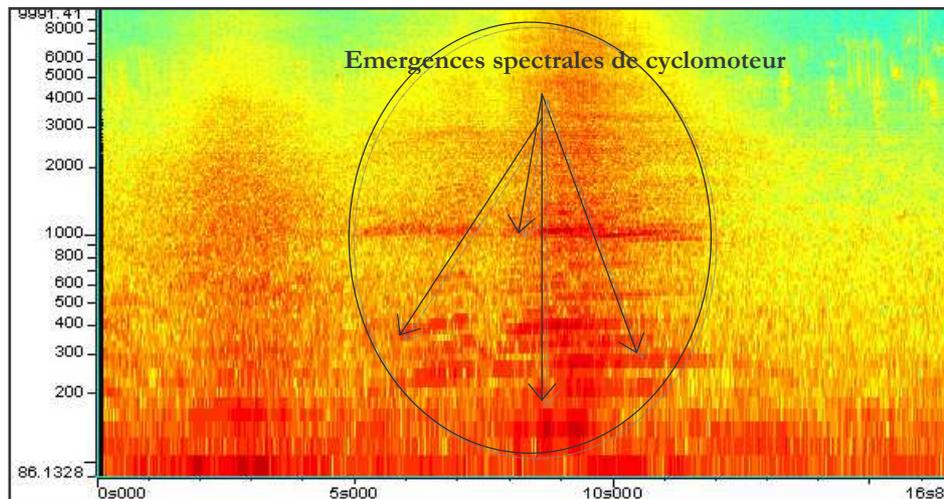


Figure 59 : Sonogramme représentant un cyclomoteur au centre ainsi que ses émergences spectrales caractéristiques.

3.2.2.2. Emergence à deux bandes

La plupart du temps, une bande de tiers d'octave, entre les bandes centrées sur les fréquences entre 100 et 250 Hz, émerge d'au moins 5 dB par rapport à ses bandes adjacentes [Figure 60].

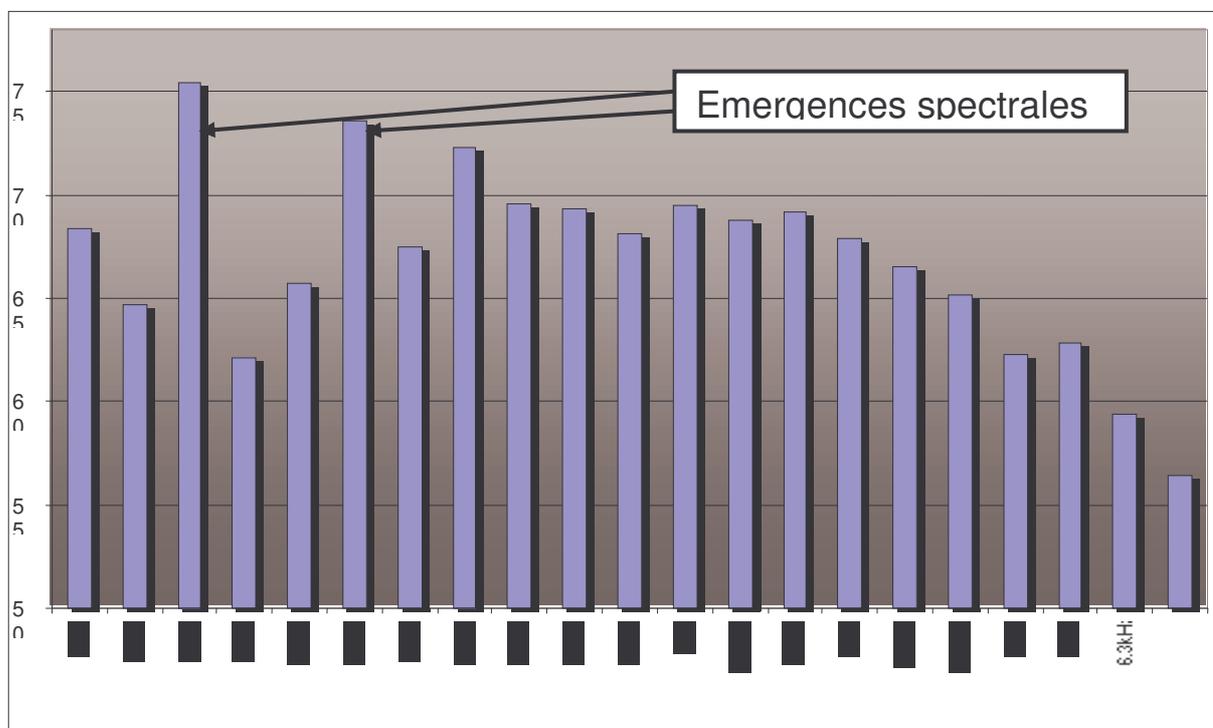


Figure 60 : Spectre moyen (0.2 s) de cyclomoteur par bande de tiers d'octave avec deux émergences spectrales. L'activité calculée par la ressemblance spectrale est augmentée en raison de la détection de l'une de ces émergences.

Nous avons implémenté une détection de pics qui multiplie la valeur de la ressemblance spectrale par un coefficient proportionnel à la valeur de l'émergence. Un contrôle permet de choisir la valeur minimum de dépassement. Un autre permet de fixer la largeur de bande à explorer pour détecter les émergences.

Cette routine peut être activée sur un seuil paramétrable. C'est-à-dire que, si on le désire, elle ne se déclenche que lorsque la courbe de ressemblance a dépassé un certain seuil.

Pour la détection des cyclomoteurs, cette routine peut être vue comme une sorte de validation de la ressemblance spectrale.

3.2.2.3. Emergence à une bande

Les oiseaux sont également caractérisés par une émergence spectrale entre les bandes de tiers d'octaves centrées sur les fréquences comprises entre 2.5 et 5 KHz [Figure 61].

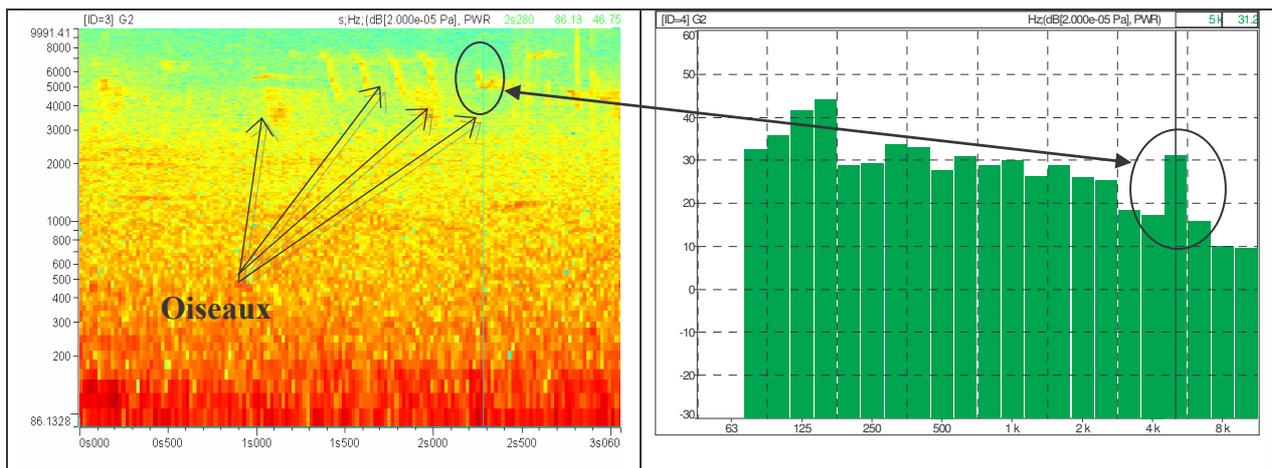


Figure 61 : Sonogramme d'une séquence d'ambiance urbaine avec présence d'oiseaux. (Durée 3 s, fenêtre de hanning, résolution 12,5 Hz, pas temporel = 20 ms). A droite une coupe correspondant au trait vertical du sonogramme. On observe une émergence de plus de 10 dB de la bande de tiers d'octave centrée sur la fréquence 5000 Hz.

Cependant, cette énergie peut parfois ne pas être distribuée uniquement dans une seule bande mais chevaucher deux bandes adjacentes [Figure 62].

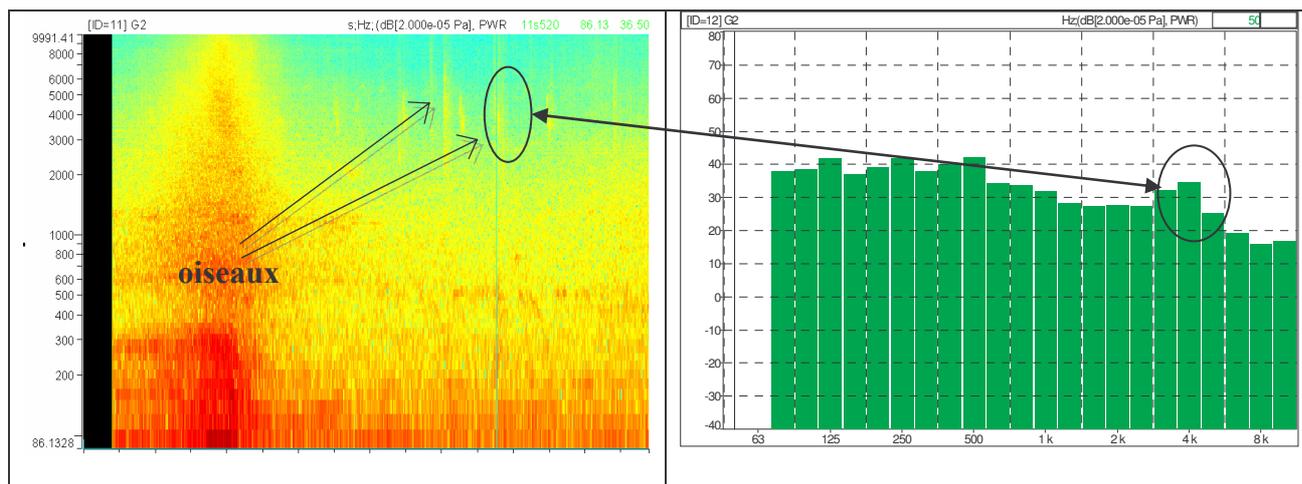


Figure 62 : Sonogramme d'une séquence d'ambiance urbaine avec présence d'oiseaux. (Durée 15 s 80, fenêtre hanning, résolution 12,5 Hz, pas temporel = 20 ms). A droite une coupe correspondant au trait vertical du sonogramme. On observe une émergence de 8 dB de la bande de tiers d'octave centrée sur la fréquence 4000 Hz par rapport à la bande de fréquence supérieure.

Nous avons ainsi implémenté une seconde *détection de pic* ayant pour objectif la recherche d'émergence significative (supérieure à 5 dB) par rapport à une seule bande de tiers d'octave adjacente.

3.3. Détermination des gabarits spectraux

3.3.1. Introduction

Nous exposons ici la démarche adoptée pour l'extraction des gabarits types servant à l'identification des bruits de véhicules.

a) Démarche empirique

La démarche consiste à établir une première série de spectres types à partir d'enregistrements réalisés dans une même rue. Dans un premier temps, nous analysons les correspondances spectrales et temporelles (par utilisation de sonogramme et par moyennes des spectres) et nous formons des groupes. Une fois ce travail réalisé, nous recommençons d'autres enregistrements mais cette fois-ci dans d'autres types d'artères urbaines et nous regardons s'il y a correspondance avec les premiers résultats trouvés, et ainsi de suite jusqu'à pouvoir trouver des similitudes assez significatives pour notre travail selon les situations [DEFREVILLE-03].

b) Démarche statistique

Une seconde analyse, complémentaire à la première, est réalisée par la suite. Elle consiste à extraire une série de spectres sur le passage d'un véhicule. L'idée est de réaliser une analyse statistique permettant d'extraire des attributs spectraux caractéristiques des sources et de les comparer, à la fois entre les individus d'un même groupe, et entre les individus de groupes différents.

3.3.2. Démarche empirique

3.3.2.1. Constitution d'une base de données

Des enregistrements ont été réalisés selon la norme 31-010 [NORME-96] dans plusieurs rues parisiennes. Les configurations des rues ont été choisies de manière à être les plus larges possibles : un ou plusieurs sens, rue simple ou boulevard, rue en U ou rue en tissu ouvert. Nous n'avons pas réalisé d'enregistrements sur des rues pavées.

Nous faisons l'acquisition du signal sonore à un taux d'échantillonnage de 51 200 Hz avec un matériel constitué d'un microphone BRUEL & KJAER de classe 1 calibré¹ par le calibre acoustique CAL 01 SL de 01dB et d'un système d'acquisition SYMPHONIE de 01dB. Le logiciel qui a servi à l'acquisition des données numériques est dBTrig de 01dB.

Les mesures ont été réalisées à Paris, dans une rue à sens unique tout d'abord (les passages sont ainsi qualifiés de nets), puis dans une rue à double sens avec une légère montée afin d'avoir le bruit moteur en évidence (en effet nombre d'automobilistes ralentissaient à la vue du dispositif de mesure), et enfin sur un boulevard avec deux fois deux voies. Les mesures ont été réalisées durant différents moments de la journée, la chaussée étant qualifiée de sèche.

3.3.2.2. Catégories

Nous formons quatre catégories ou types en rapport avec nos modèles subjectifs : voitures et utilitaires, cyclomoteurs et petits scooters, moto et enfin bus ou camion.

Ces groupes ont déjà été formés de la sorte dans des études précédentes à partir des caractéristiques physiques et perceptives des sources. La dernière en date est celle de la thèse de David Gaulin [GAULIN-00].

Nos bases de données contiennent au minimum cinquante passages de véhicules par catégorie. Pour un véhicule roulant à une allure de 50 Km/h, nous extrayons une période de 2 à 3 s ce qui

¹ Le mot *calibré* est ici employé dans le sens de la définition donnée par la norme NFS 31-010 de décembre 1996 intitulée *Caractérisation et mesurage des sons de l'environnement. Méthodes particulières de mesurage.*

correspond à une mesure sur un passage d'une trentaine de mètres. Sur cette durée, son allure émerge nettement du bruit de fond de la ville [Figure 63].

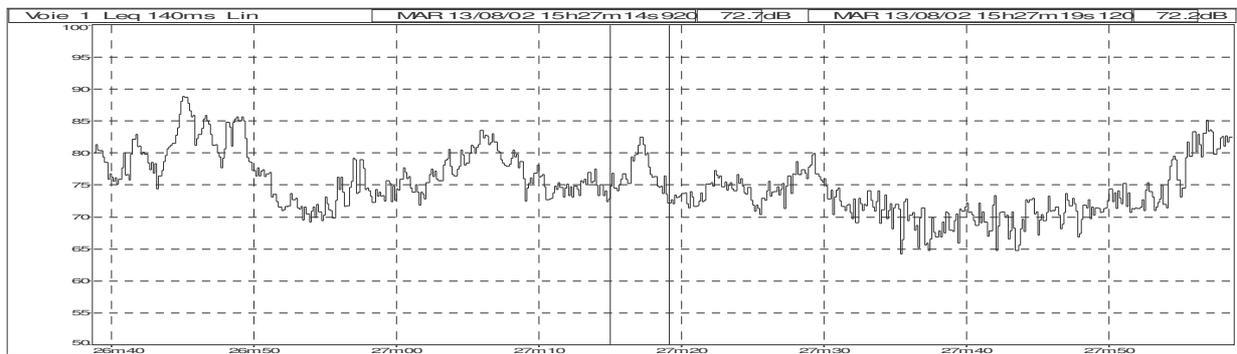


Figure 63 : Evolution du $L_{eq,100ms}$ d'un bruit de trafic urbain. Entre les deux barres, le passage sur 3 s d'un véhicule. Celui-ci émerge nettement du bruit de fond.

3.3.2.3. Utilisations des sonagrammes

Le sonagramme est un outil permettant de visualiser l'évolution temporelle des fréquences d'un signal combiné à leur intensité. A titre d'exemple, la figure 64 illustre les différences entre le passage d'une voiture suivie de celle d'un cyclomoteur (caractérisée par ses bandes horizontales). En fin de stimulus, les taches jaunes (entre trois et six milles hertz) sont caractéristiques d'un chant d'oiseau.

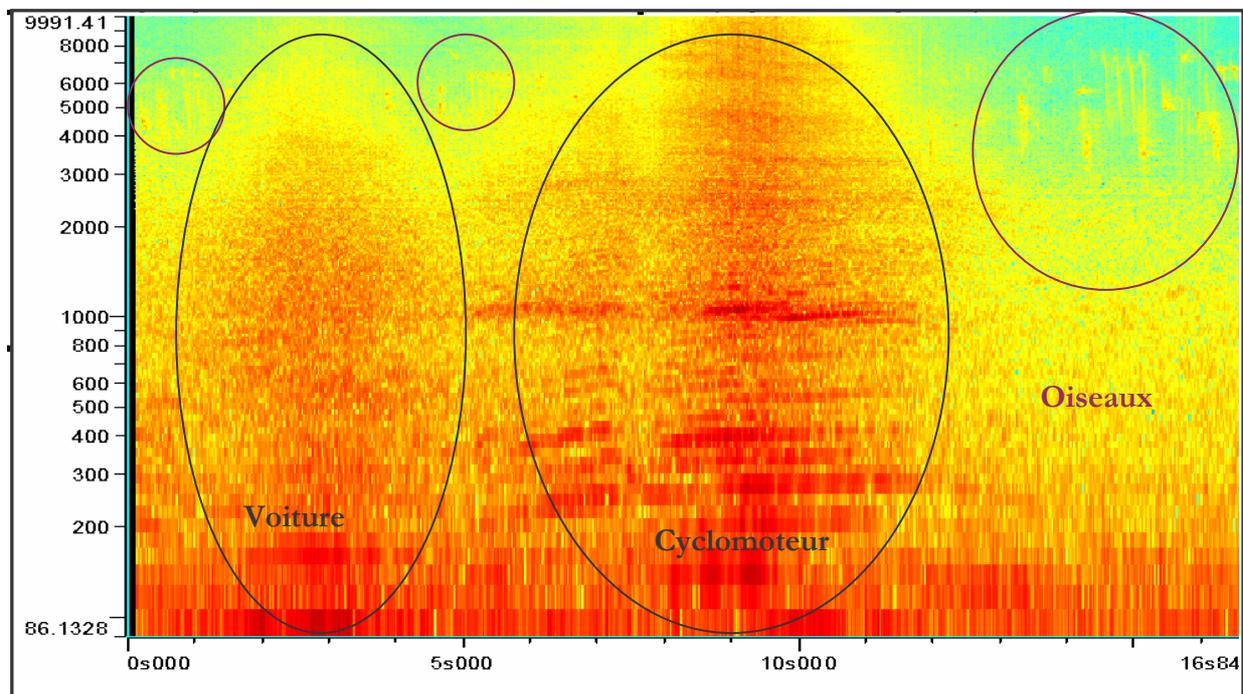


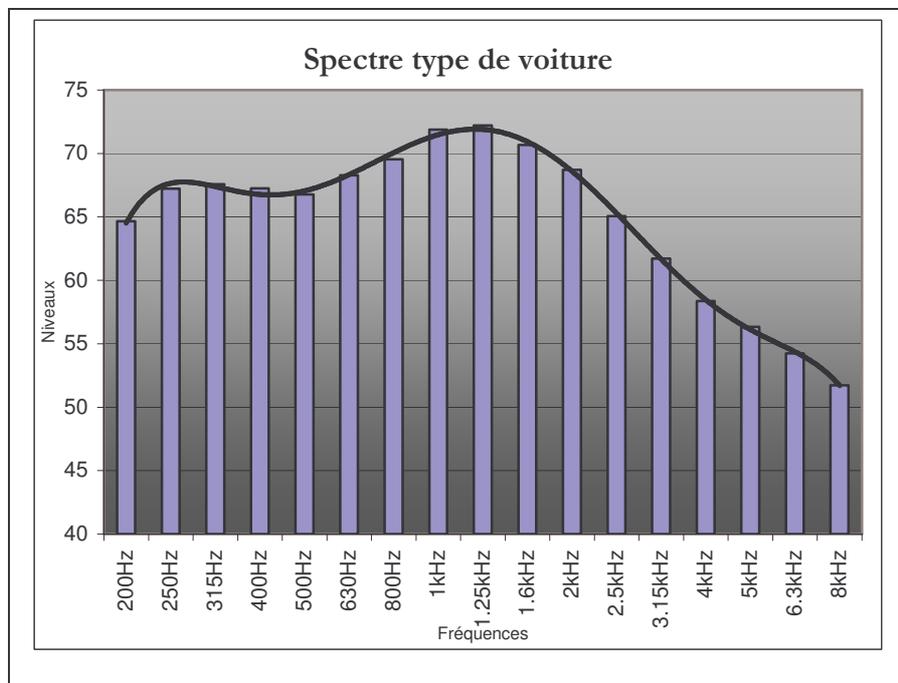
Figure 64 : Sonagramme d'une séquence de 17 s comprenant une voiture, un cyclomoteur et des oiseaux, (fenêtre de hanning, résolution 12,5 Hz, pas temporel = 20 ms).

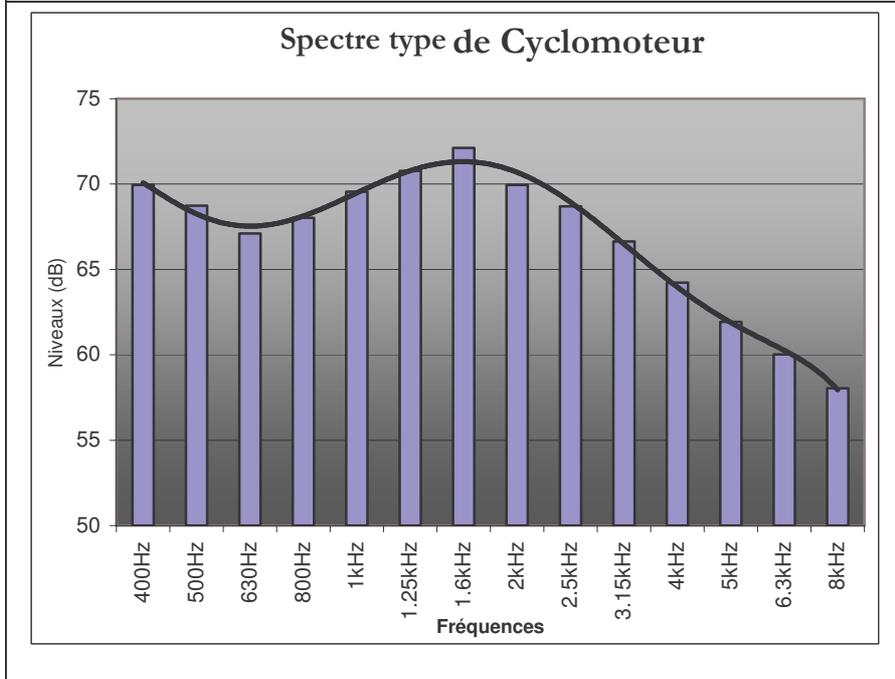
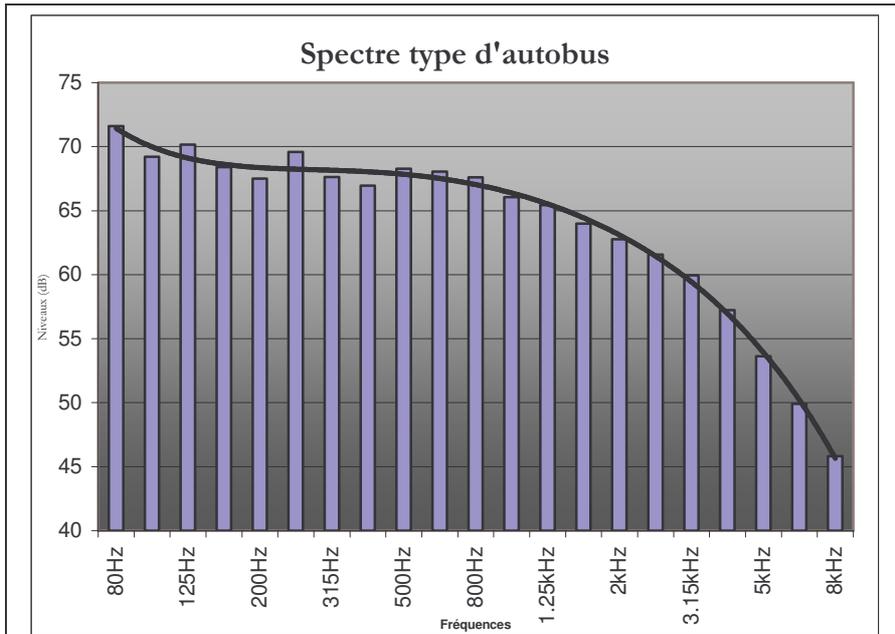
L'utilisation initiale du sonagramme par tâtonnements nous a permis de constater que l'analyse par bandes d'octaves n'est pas assez précise. Nous perdions des informations spectrales telles que les « raies » spectrales caractéristiques du cyclomoteur que nous voyons sur la figure 64. Nous prenons donc une analyse par bandes de tiers d'octaves. De plus, nous retenons une largeur de spectre allant de 80 à 8000 Hz. En effet, les perturbations sont trop importantes en basses fréquences (interférences à 50 Hz avec les lignes électriques) et des différences de niveaux sont mesurables jusqu'à 8000 Hz pour différentes catégories de véhicules.

À l'issue de cette première analyse, les spectres types issus sont des moyennes sur une seconde. Cette durée est extraite à partir d'une demi-seconde de part et d'autre de la valeur maximale en dB(A) lors du passage du véhicule. En effet, nous avons généralement constaté une symétrie des profils énergétiques autour des maximums d'intensité [Figure 63].

3.3.2.4. Gabarits

La figure 65 montre la répartition spectrale des quatre types de véhicules que nous appellerons gabarits. Nous avons fait la moyenne arithmétique d'une cinquantaine de passages par catégorie après avoir égalisé leur niveau global à 80 dB(A).





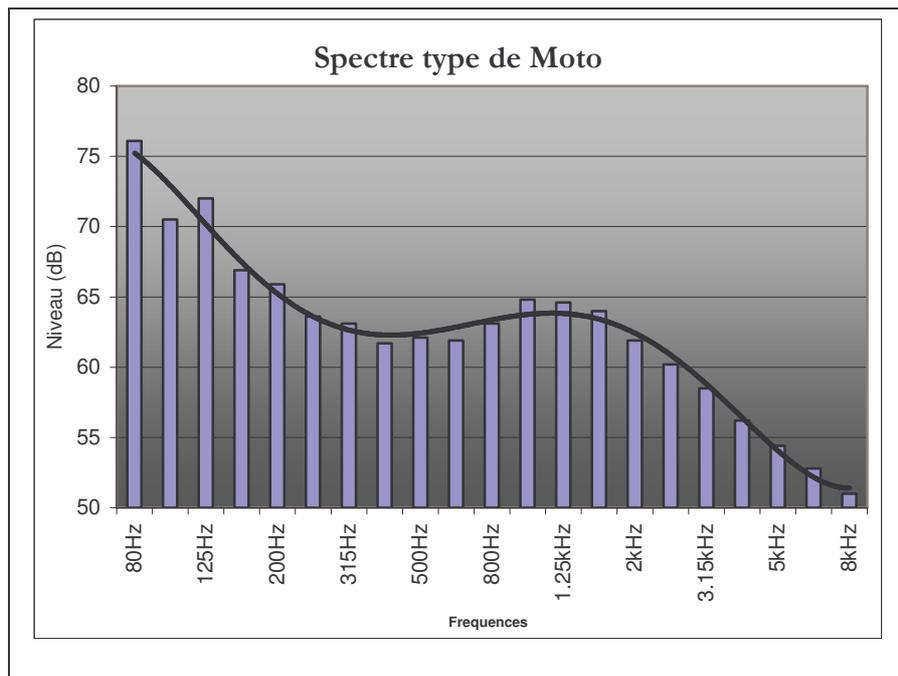


Figure 65 : Gabarits spectraux avec courbe d'évolution pour les quatre catégories de véhicules.

Les mesures en basses fréquences et bas médium sont trop dépendantes du régime moteur, en particulier pour les voitures et les cyclomoteurs. Nous avons donc, par la suite, tronqué les parties des spectres ayant une forte dispersion entre les différents enregistrements. (En dessous de 250 Hz pour les autobus, en dessous de 400 Hz pour les cyclomoteurs et scooters, en dessous de 200 Hz pour les voitures).

Les deux principales sources de bruits d'un véhicule motorisé sont largement mises en évidence. Le bruit émis par le moteur ou bruit rayonné par le groupe moto-propulseur est représenté dans les basses fréquences et les bas médiums. Celui-ci dépend fortement du rapport de la boîte de vitesse engagé, du type de moteur (essence ou diesel), de la charge du véhicule. L'autre source de bruit prédominante provient du contact du pneumatique avec la chaussée. Elle se retrouve au niveau des hautes fréquences avec des pics communs à chaque type de véhicule (1 kHz pour les voitures, 1.25 kHz pour les autobus et les motos et 1.6 kHz pour les cyclomoteurs). Notons qu'aux vitesses urbaines (aux alentours de 50 km/h, conditions in situ de nos mesures), les deux bruits (moteur et contact pneumatique-chaussée) ont des niveaux quasiment égaux, ce qui représente une tendance nette pour la majorité des véhicules rencontrés en conditions urbaines.

Sur la figure 66 nous pouvons évaluer visuellement les écarts spectraux entre véhicules. Plus ces écarts sont importants, plus la routine ressemblance spectrale pourra faire une ségrégation pertinente entre les catégories de véhicules.

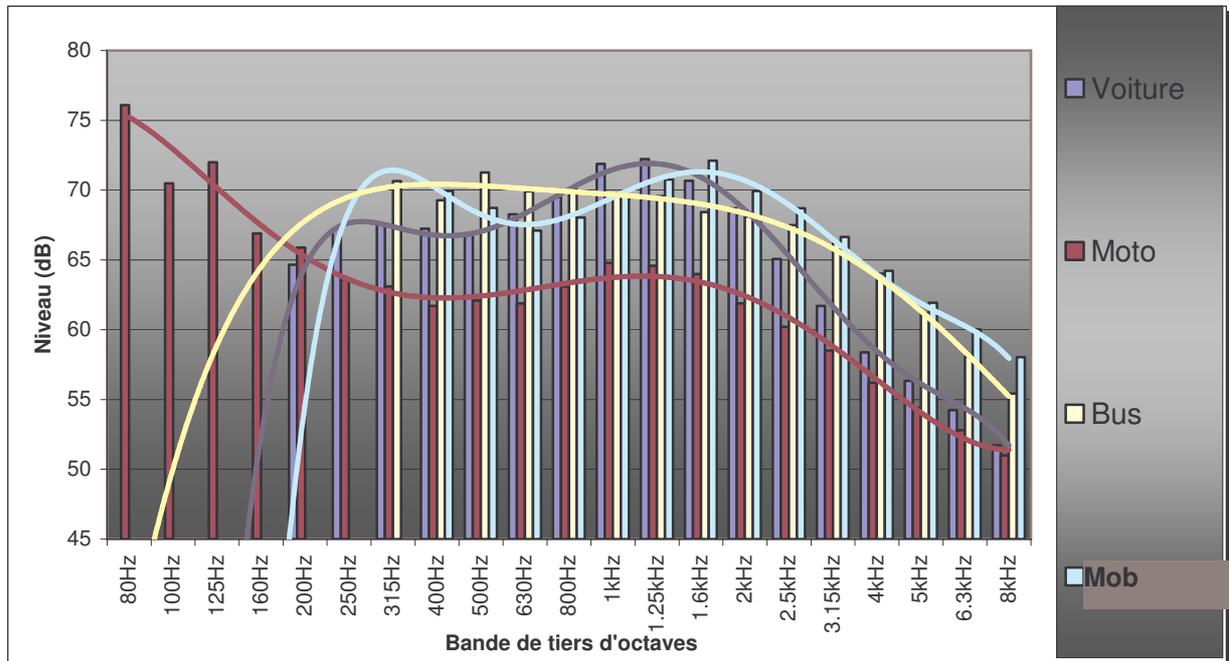


Figure 66 : Gabarits spectraux superposés avec courbe d'évolution pour les quatre catégories de véhicules. Les spectres correspondent à ceux de la figure 65.

3.3.3. Démarche statistique

Bien que la première analyse nous ait permis de nous faire une idée des attributs spectraux importants et ainsi former des groupes de véhicules, cette méthode était assez intuitive dans le choix des patterns et doit être associée à une confirmation plus rigoureuse. C'est l'objet de cette analyse statistique. Nous utilisons ici les mêmes bases de données sonores que pour la première analyse.

L'idée est de générer des séquences d'attributs propres aux sources sonores issus des spectres tiers d'octave (de la bande centrée à 80 Hz à celle centrée à 8000 Hz) calculés toutes les 0.2 s sur une durée de totale de 6 s autour du temps correspondant au niveau maximum en dB(A) [Figure 67]. Nous allons de la période T0 à la période T30.



Figure 67 : Courbe de $L_{Aeq,200ms}$ d'un passage de véhicule avec indication de la période sur laquelle sont calculés les attributs, soit 3 s avant et après le niveau maximum.

En d'autres termes cela revient à former, pour un passage de véhicule, la matrice des valeurs d'un sonagramme avec une résolution spectrale au tiers d'octaves et une résolution temporelle de 200 ms.

Chacun de ces spectres est caractérisé par vingt et une valeurs correspondant au niveau de chacune des bandes de tiers d'octaves. Nous obtenons une matrice (30 X 21) correspondant aux nombres de spectres sur trente instants successifs séparés de 0.2 s. Les six cent trente éléments de la matrice sont appelés attributs [Tableau 25].

Un attribut a ainsi une valeur X correspondant à l'énergie d'une bande de tiers d'octave Y à un temps t.

Temps	0 à 200 ms	200 à 400 ms	...	3 s à 3.2 s	...	5.8 s à 6 s
Instant	T0	T1	...	T15	...	T30
Attribut	1 à 21	22 à 43	...	294 à 315	...	609 à 630
Bande	80 Hz à 8 kHz	80 Hz à 8 kHz	...	80 Hz à 8 kHz	...	80 Hz à 8 kHz

Tableau 25 : Numéros et références des attributs.

Ainsi, pour chaque type de source, nous calculons :

- Les écarts types des attributs.
- Les moyennes des attributs (sur une période de une seconde glissante toutes les 0.2 s).

Ces deux données statistiques (moyennes et écart types), exposées dans les graphiques, nous permettent de mesurer l'importance d'un attribut (plus exactement de la moyenne des attributs sur une seconde glissante).

L'écart type nous renseigne sur la tendance d'un attribut à s'écarter plus ou moins fortement d'une valeur moyenne. Ainsi, un faible écart type est distinctif d'une concentration des valeurs de l'attribut près de la valeur moyenne. Il peut être interprété, à première vue comme étant le signe que l'attribut auquel il fait référence est une caractéristique stable donc représentative du type de véhicule. La valeur de la moyenne, comparée aux moyennes des autres véhicules, permet alors de confirmer ou d'infirmer cette interprétation.

Nous avons pris vingt individus pour chaque catégorie de véhicules. Ces individus sont sélectionnés sur écoute. Les sources sont reconnaissables sur la totalité de la longueur des séquences. De légers bruits de fond de circulation urbaine sont également présents sur la majorité des stimuli car ils sont enregistrés in situ.

Les motos ne sont pas représentées car les résultats obtenus lors de ces analyses sont trop disparates. Cette diversité des bruits de motos rencontrés en site urbain illustre la complexité de leur caractérisation spectrale.

3.3.3.1. Ecart type des attributs

Nous reprenons ci-dessous les graphes de l'écart type en fonction du numéro d'attribut pour les voitures, les cyclomoteurs et les bus [Figure 68].

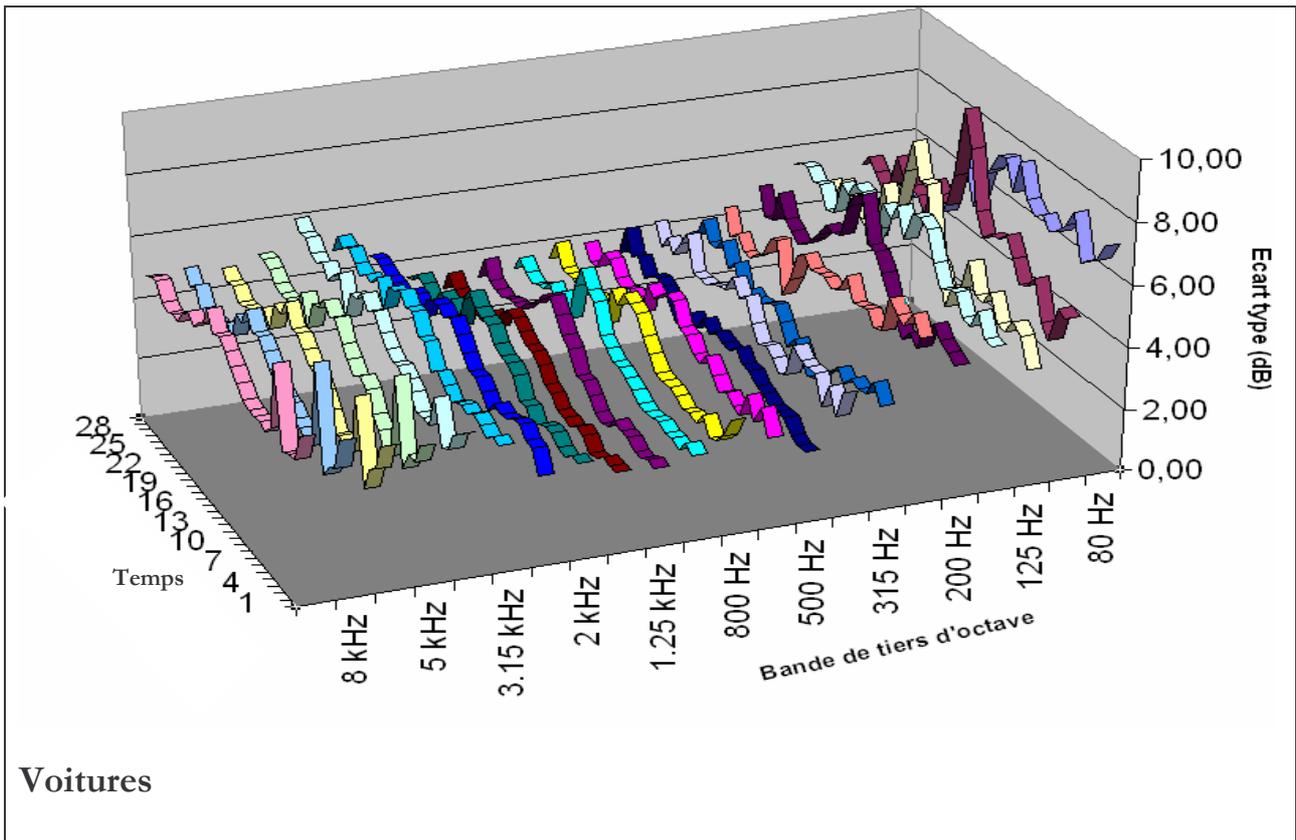


Figure 68 : Ecart types des séquences d'attributs (trente séquences correspondant aux trente instants de 200 ms) sur les vingt et une bandes de tiers d'octave. Ces statistiques ont été faites sur un échantillon de vingt voitures.

D'un point de vue temporel, les écarts types des attributs sont peu sujets aux variations, à l'exception des temps correspondant à la moitié du passage (aux alentours du temps quinze donc 3 s). Ainsi, la tendance des écarts types (qu'ils soient importants ou non) ne subissent que peu de variations au cours des passages.

D'un point de vue spectral, les minima d'écart types ont tendance à se situer sur les bandes centrées entre 200 Hz et 2 kHz. La largeur de cette plage semble néanmoins restreinte pour pouvoir trouver des différences assez robustes avec les autres véhicules.

En dessous de 200 Hz, les écarts types sont importants. Ils sont la conséquence de la variabilité des régimes moteurs.

Il semble donc judicieux de retenir comme plage fréquentielle supplémentaire les bandes situées au dessus de 2 kHz.

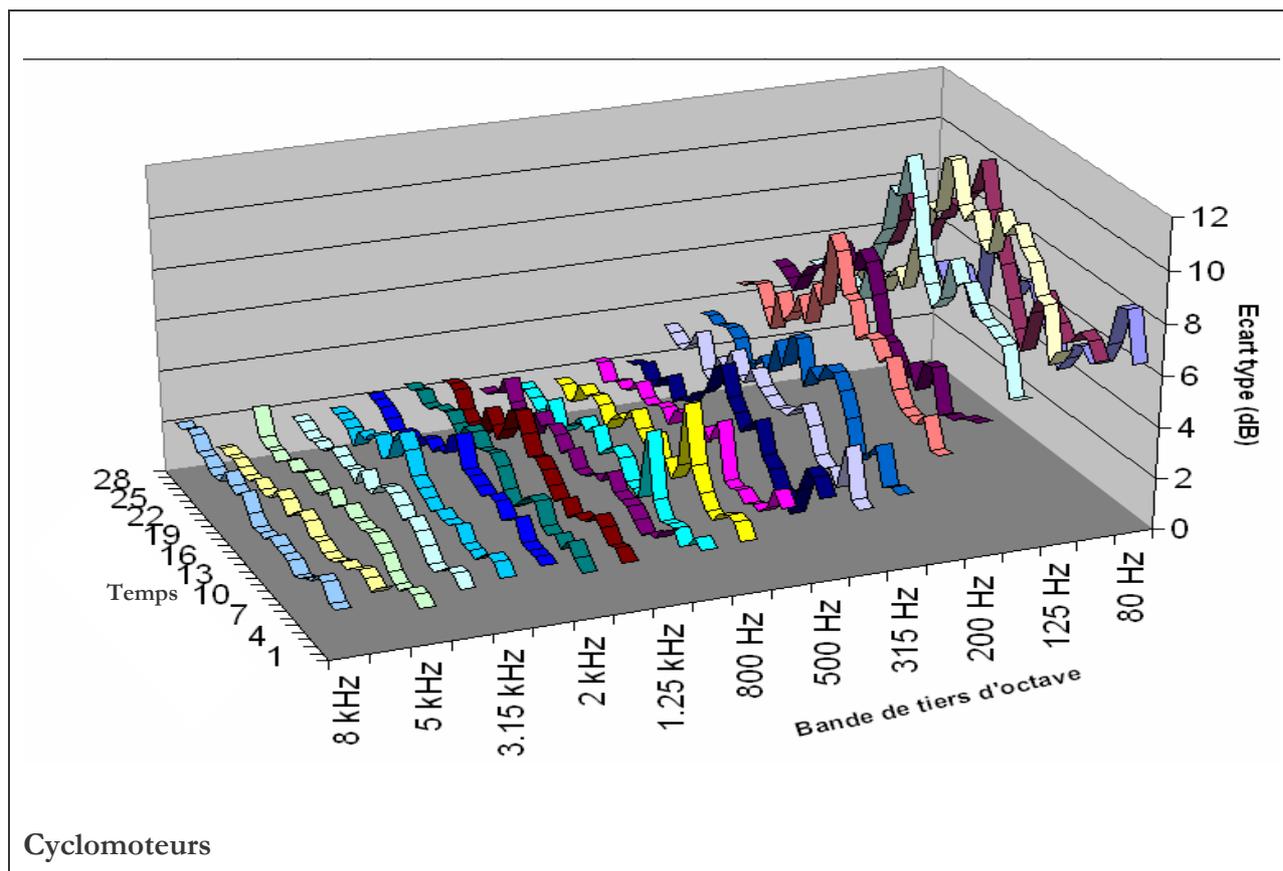
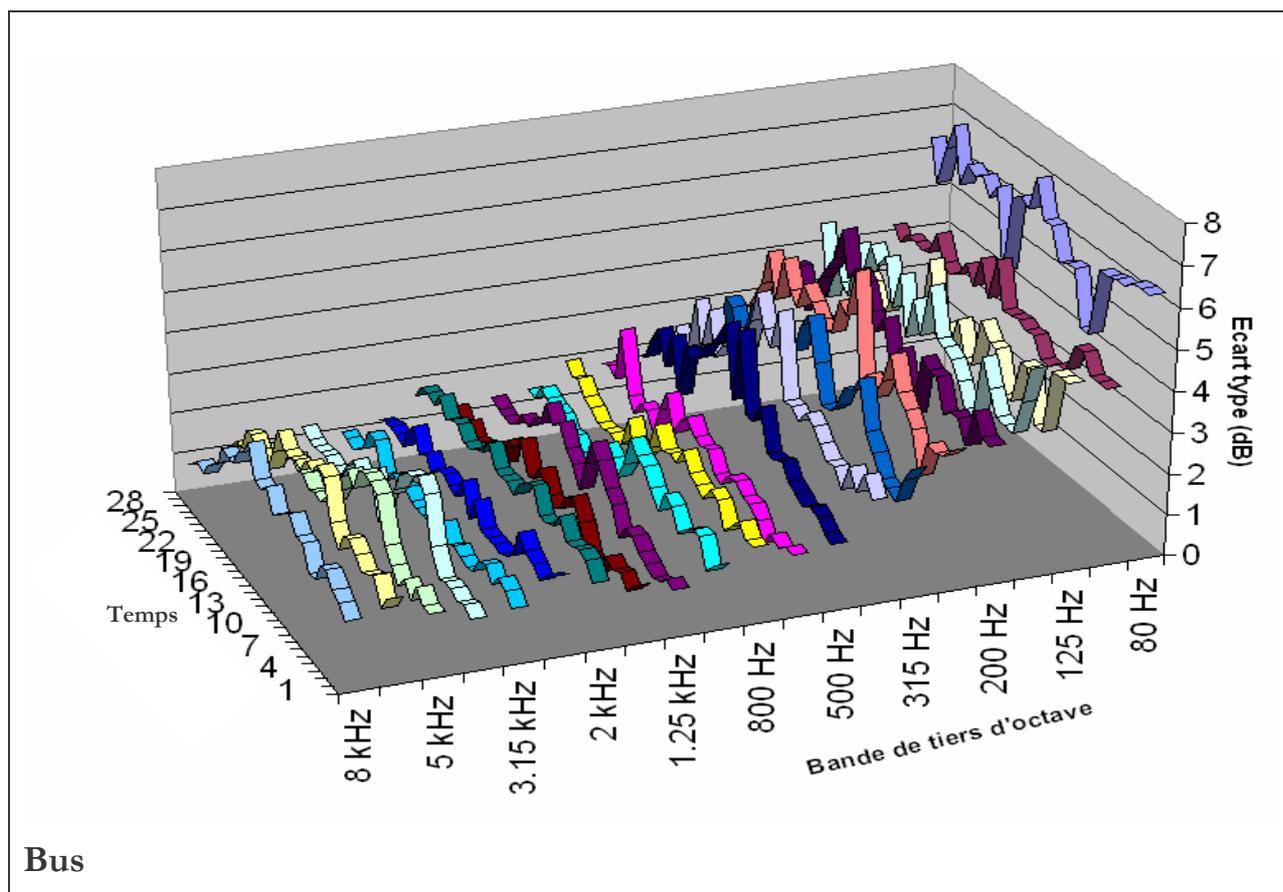


Figure 69 : Ecart types des séquences d'attributs (trente séquences correspondant aux trente instants de 200 ms) sur les vingt et une bandes de tiers d'octave. Ces statistiques ont été faites sur un échantillon de vingt cyclomoteurs.

La figure 69 montre clairement que plus on s'approche des médiums et basses fréquences, plus les écarts types, donc les différences de niveaux entre passages, sont importants. Cela est dû aux fréquences marquées des cyclomoteurs comme nous les voyons sur le sonagramme de cyclomoteur du paragraphe 2.2.1. Cette différence est nettement accentuée aux temps intermédiaires des passages (du temps dix au temps seize) ce qui est peut être dû à l'effet Doppler. En effet, ces temps correspondent à une phase de transition pour laquelle le véhicule passe de l'état d'approche du système d'enregistrement à celui d'éloignement.

Le sonagramme montre également que la stabilité temporelle et fréquentielle des attributs situés dans les aigus et observés sur la figure 69 est intéressante car elle n'est pas une mesure du bruit de fond.



Bus

Figure 70 : Ecart types des séquences d'attributs (trente séquences correspondant aux trente instants de 200 ms) sur les vingt et une bandes de tiers d'octaves. Ces statistiques ont été faites sur un échantillon de vingt bus.

La figure des écarts types des attributs des bus [Figure 70] a la même allure que celle des cyclomoteurs : une stabilité qui décroît lorsque l'on tend vers les basses fréquences ainsi que des périodes temporelles intermédiaires moins favorables. Les explications à ces phénomènes sont les mêmes que pour les cyclomoteurs : effet Doppler et bruits de moteur prédominants.

Nous observons toutefois une bonne stabilité de faibles écarts types de la part d'attributs correspondant à certaines bandes de tiers d'octaves : entre 500 Hz et 3.15 kHz.

De l'observation des trois figures précitées, nous en formulons les conclusions suivantes :

- Nous ne prenons pas en compte les bandes de tiers d'octaves des basses fréquences dans l'établissement de nos gabarits types,
- Les périodes temporelles optimales pour établir les gabarits types (qui sont des moyennes spectrales sur une période de 1 à 2 s environ) se situent avant et après les

passages de véhicules au droit des microphones. Cependant, en effectuant des tests d'identification avec des gabarits extraits avant le maximum et des gabarits extraits autour du maximum, il s'est avéré que les résultats ne variaient guère, si ce n'est sur un décalage temporel. Ainsi, nous avons retenu le gabarit extrait autour de la moyenne car l'identification coïncide mieux avec le moment du passage des véhicules (c'est-à-dire lorsqu'il est à son niveau maximum).

- Les attributs de certaines bandes de tiers d'octave, caractérisés par de faibles variations temporelles et les écarts types peu élevés sont très pertinents mais leur nombre est généralement trop faible pour ne pas tenir compte de bande de tiers d'octaves moins distinctives dans l'établissement des gabarits types.

3.3.3.2. Moyennes des attributs

L'étude des écarts types nous a permis de définir des périodes temporelles et des largeurs spectrales propices à l'intérieur des trois catégories de véhicules. La visualisation de la moyenne des attributs permet de comparer ces zones entre les véhicules [Figure 71]. Pour des raisons de clarté, nous exposons les moyennes des attributs calculées sur une seconde (de la période « 11 » à la période « 15 »).

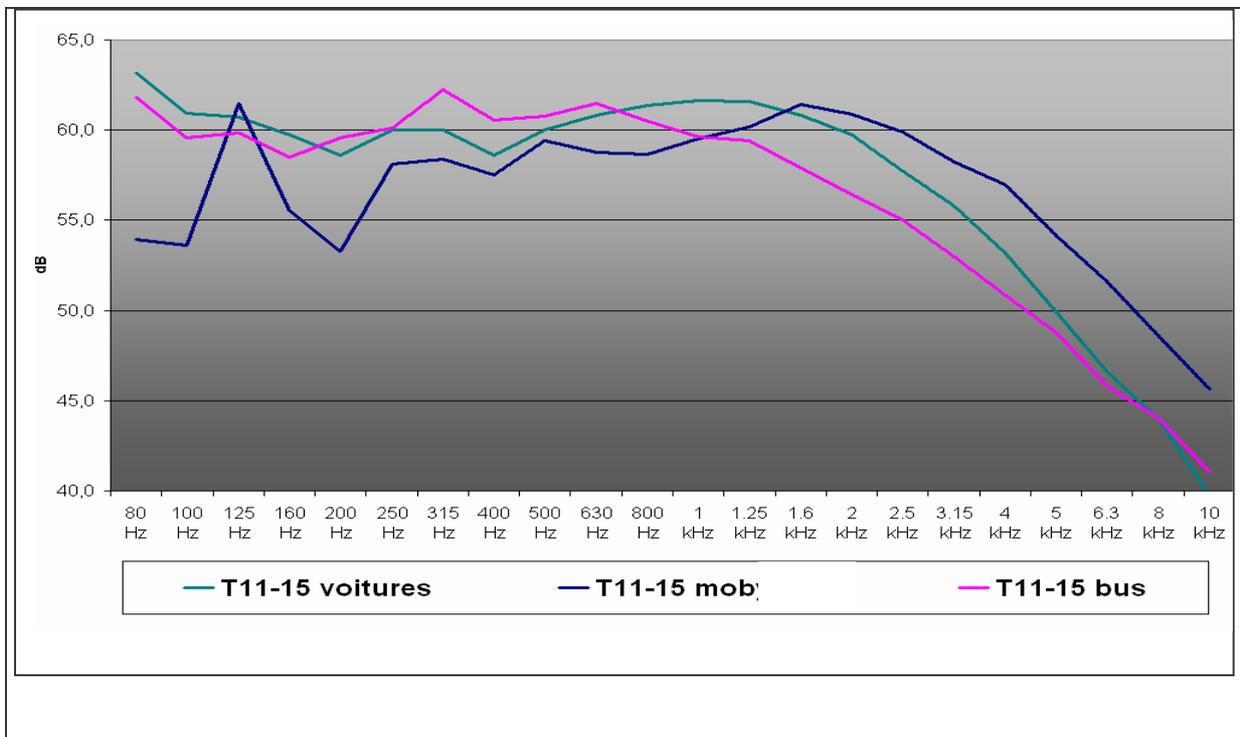


Figure 71 : Moyenne spectrale par bandes de tiers d'octave (attributs « 148 » à « 169 ») sur une durée d'une seconde (T11 à T15) pour les trois types de véhicules.

Nous avons vu que les bandes d'octaves situées dans les basses fréquences ne sont pas prises en compte pour la détermination des gabarits types.

La bande centrée sur 1 kHz est intéressante car les voitures se détachent des bus et des cyclomoteurs.

Les bandes précédentes présentent relativement peu de différences et nous avons vu que les attributs associés étaient sujets à de fortes variations.

Les bandes situées après 1 kHz ont des allures décroissantes à peu près similaires. Les écarts entre attributs peuvent être utiles pour les discriminations.

L'allure de ces graphes révèle que les spectres extraits sur une seconde sont sensiblement identiques aux spectres trouvés lors de l'analyse empirique [Figure 66], qui était réalisée sur des périodes comprises entre 2 et 3 s.

3.4. Résultats des identifications

3.4.1. Oiseaux

Les oiseaux sont identifiés de manière satisfaisante car leur chant est assez bien caractérisé par des émergences spectrales détectables en analyse de tiers d'octaves. Ces émergences ressortent de manière significative du bruit de fond urbain ainsi que lors de l'approche ou de l'éloignement des véhicules. Elles se situent dans une fourchette allant de 3000 à 7000Hz.

Néanmoins, la pertinence des détections d'oiseaux pourrait être améliorée si le système calculait un spectre tiers d'octave sur une période de 50 ms au lieu des 100 ms actuelles.

3.4.2. Voix

Les voix ne sont pas bien identifiées par ce système. La variabilité de leur spectre est trop importante dans le temps (principalement à cause des silences entre les mots) et le système présenté ici n'est pas adapté pour être sensible à ces fluctuations.

Par exemple la figure 72 montre le sonagramme du stimulus « 1 » avec une coupe sur une voix d'enfant (trait bleu vertical). Le spectre correspondant (à droite) montre bien un trait

caractéristique tel les harmoniques (500, 1000, 2000 et 4000 Hz), mais son manque de stabilité temporelle empêche le système de les détecter.

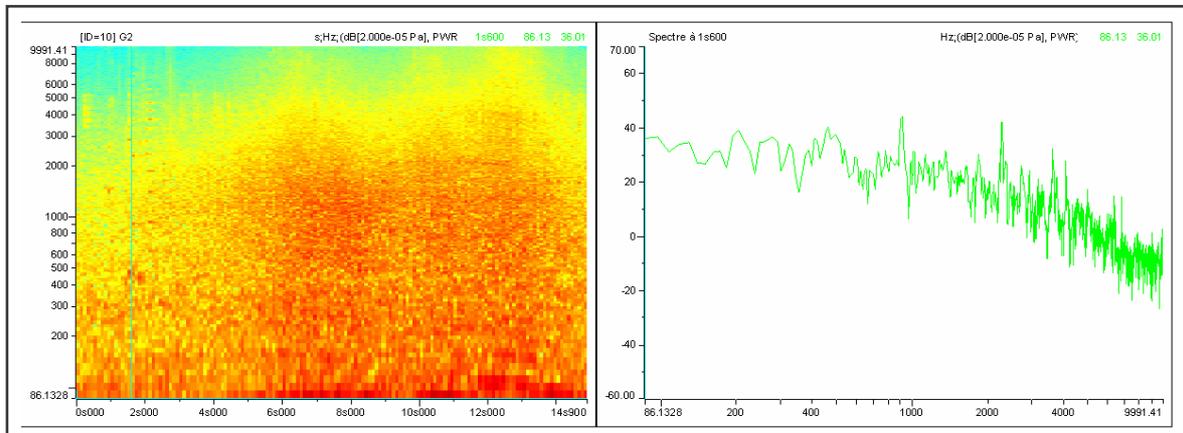


Figure 72 : Sonagramme du stimulus « 1 » (à gauche) avec une coupe sur une voix d'enfant (trait bleu vertical).

Pour les voix d'adultes, les harmoniques sont moins présentes mais nous avons une émergence d'énergie aux alentours de 500 Hz, ainsi que l'illustrent le sonagramme et le spectre instantané du stimulus « 6 » [Figure 73].

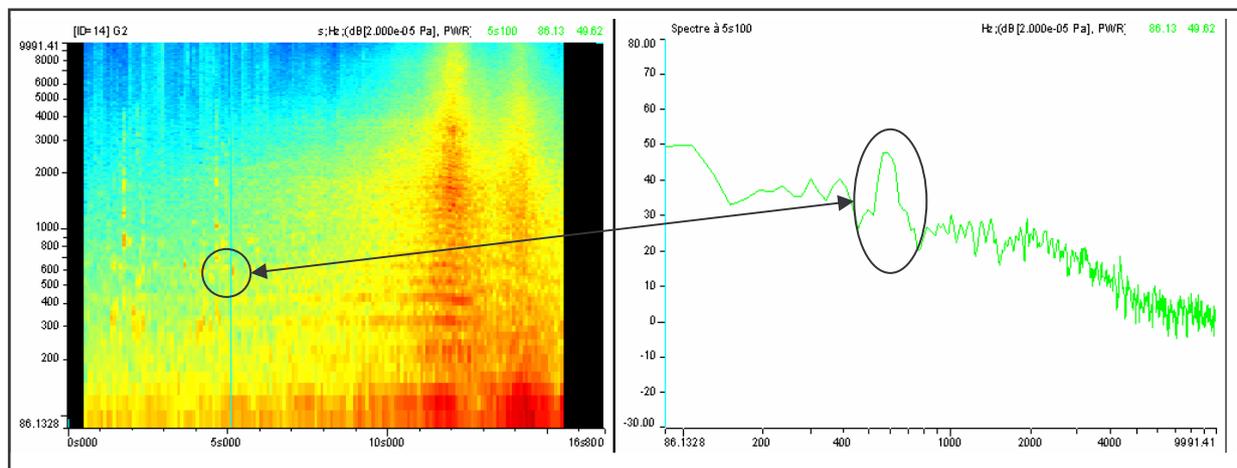


Figure 73 : Sonagramme du stimulus « 6 » (à gauche) avec une coupe sur une voix d'adulte (trait bleu vertical). On remarque une émergence spectrale aux alentours de 500 Hz.

Les deux routines implémentées dans A2S (ressemblance et émergence spectrale) ne sont pas en adéquation avec la complexité des variations spectro-temporelles des voix.

3.4.3. Véhicules

La figure 74 illustre la visualisation du logiciel telle que nous pourrions la voir sur l'interface sur une durée de 12 s. Les valeurs de ressemblance sont alors stockées et récupérées dans un fichier de données.

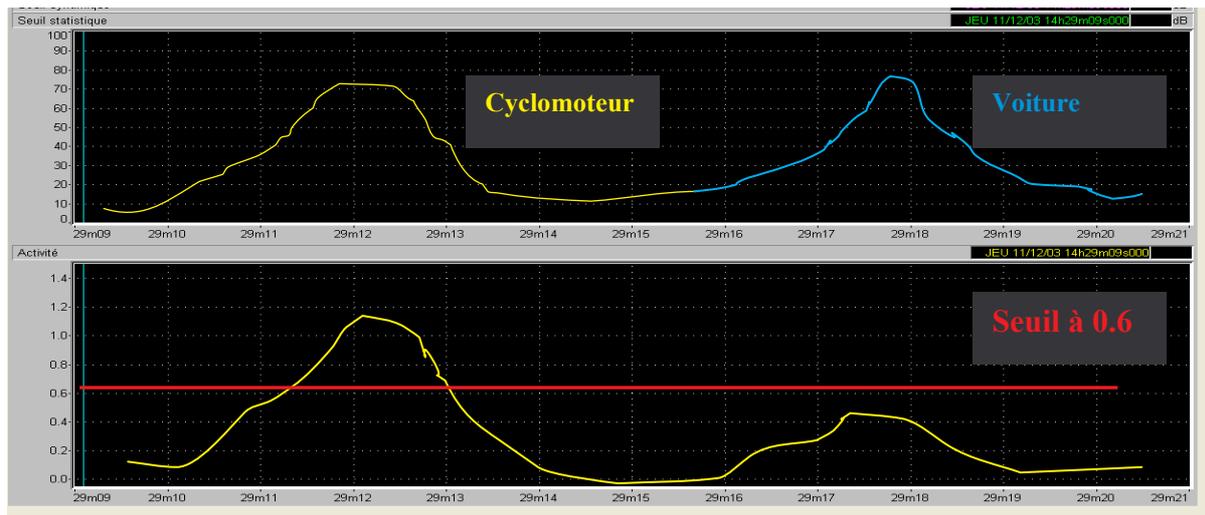


Figure 74 : Exemple de courbe d'identification pour un cyclomoteur telle que nous pourrions la voir sur l'interface du logiciel et sur une durée de 12 s.

Au dessus, nous avons la courbe de L_{Aeq} (codée en jaune pour le cyclomoteur et en bleu pour une voiture). En dessous nous avons la courbe de ressemblance spectrale des cyclomoteurs. Le seuil au-delà duquel la présence de la source est confirmée est fixé à une valeur égale à « 0.6 ».

Pour évaluer le comportement du logiciel, nous comparons dans le tableau 26 les durées d'apparition des véhicules relevées selon un codage expert et selon la reconnaissance du logiciel. Les durées sont exprimées sous forme de pourcentages de temps d'apparition

Stimulus	%T expert Voit	%T A2S Voit	%T expert Mob	%T A2S Mob	%T expert Bus	%T A2S Bus	%T expert Moto	%T A2S Moto
1	63	100	0	0	0	0	0	0
2	16	25	75	36	0	0	0	0
3	3	33	0	0	33	10	33	10
4	27	100	0	0	0	13	0	0
5	3	54	46	10	0	0	0	0
6	0	0	20	7	80	51	0	0
7	0	0	85	10	0	3	0	7
8	18	29	26	32	0	3	44	25
9	25	33	53	14	0	0	0	36
10	0	0	0	0	63	7	0	7
11	23	80	20	10	0	0	0	0
12	8	70	0	0	0	0	0	21
13	63	100	0	0	0	0	0	27
14	0	0	33	30	0	0	14	14
15	54	100	0	0	0	4	0	0
16	15	47	0	0	0	14	0	0
17	16	34	66	36	0	0	0	0
18	9	46	0	0	54	3	0	8
19	0	0	0	0	100	42	0	0
20	0	0	0	0	0	0	40	25

Tableau 26 : Comparaison des pourcentages de temps de présence (%T) des véhicules sur les vingt stimuli de 15 s obtenus par une écoute experte (expert) et par le système A2S (A2S).

Globalement les identifications les plus pertinentes sont celles obtenues pour les voitures et les cyclomoteurs. Même si les durées manquent encore de précision, tous les stimuli contenant ces sources sont détectés par le système.

En revanche, il y a des erreurs d'identification pour les bus et les motos. Le système a tendance à identifier des événements recherchés qui ne sont pas présents dans les stimuli.

3.5 Activité des sources sonores

L'*activité des sources* a été présentée de façon schématique comme étant une évaluation de la contribution énergétique d'une source dans un environnement multi source. Elle permet, lorsque deux sources se superposent, d'évaluer la partie (spectrale) utile à leur reconnaissance. Son calcul est défini comme étant le niveau sonore en dB(A) (L_{Aeq}) obtenu après filtrage par un gabarit type (voir paragraphe 3.2.1.). Ce niveau est calculé (par pas de 100 ms) pendant la période durant laquelle l'identification d'une source est validée par le système : *ressemblance spectrale + identification de pic*.

Ainsi, pour chacun des vingt stimuli, nous avons calculé la courbe d'*activité* de chacun des véhicules. Nous avons ensuite relié les différents critères calculés sur cette courbe d'*activité* avec la prégnance et la proximité des véhicules (Tableau 27). Les critères calculés sont les suivants : *activité* moyenne, *activité* maximale, *activité* dépassée pendant 10% du temps et *activité* dépassée pendant 5% du temps.

Activité	Prégnance Mob	Proximité Mob	Prégnance Voit	Proximité Voit
$L_{Aeqfiltré}$	0,62	0,72*	0,67*	0,04
$L_{Amaxfiltré}$	0,84**	0,90**	0,23	0,40
$L_{A5filtré}$	0,83**	0,89**	0,70*	0,01
$L_{A10filtré}$	0,77*	0,84**	0,70*	0,01

Tableau 27a : Coefficients de corrélation entre la prégnance, et la proximité des cyclomoteurs et des voitures et les critères calculés sur la courbe d'*activités*. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

La mesure de l'*activité* telle que nous l'avons définie permet d'obtenir une bonne correspondance entre la prégnance et la proximité des voitures et des cyclomoteurs.

Toujours pour les vingt stimuli, le tableau 27b indique les corrélations obtenues entre la courbe de L_{Aeq} codée à l'oreille et la prégnance et la proximité des voitures et cyclomoteurs.

Critères codés voit ou mob	Prégnance Mob	Proximité Mob	Prégnance Voit	Proximité Voit
L_{Aeq}	0,76*	0,85*	0,27	0,66*
L_{Amax}	0,82**	0,84**	0,57*	0,80**
L_{A5}	0,74*	0,82**	0,44	0,76**
L_{A10}	0,67*	0,77*	0,31	0,55*

Tableau 27b : Coefficients de corrélation entre la prégnance et la proximité des cyclomoteurs, et des voitures et les critères calculés sur les courbes de L_{Aeq} codées voitures et cyclomoteurs. Les résultats avec ** indiquent une probabilité avec un niveau de confiance $p < 0.01$ et les résultats avec * indiquent une probabilité avec un niveau de confiance $p < 0.05$.

En observant les chiffres des deux tableaux, nous constatons que le codage automatique réalisé selon le système permet d'obtenir une mesure de critères perceptifs. Pour la prégnance et la proximité des cyclomoteurs, cette mesure est aussi pertinente qu'une mesure extraite de la courbe de L_{Aeq} codée de manière experte.

Elle permet également d'obtenir de meilleurs résultats concernant l'évaluation métrologique de la prégnance des voitures. Elle s'avère cependant insuffisante pour la mesure de la proximité des voitures.

3.6. Conclusion

Nous avons créé une méthode simple, qui a l'avantage de fonctionner en temps réel. Cependant, elle n'est pas très robuste à la variabilité spectrale des événements recherchés. Elle fonctionne bien pour les voitures et les cyclomoteurs et est efficace pour une certaine catégorie d'autobus mais est insuffisante pour les motos. Cette méthode peut être utile pour identifier rapidement une source dont les caractéristiques spectrales présentent peu de variabilité au point de mesure. Elle permet également de se rapprocher de la mesure automatique de la prégnance et de la proximité des cyclomoteurs ainsi que de la prégnance des voitures de façon similaire à une méthode nécessitant l'intervention d'un opérateur.

Pour l'identification des oiseaux, le logiciel est limité par sa résolution temporelle qui empêche de réaliser des filtrages à des cadences inférieures à 100ms. Or, si nous arrivons à descendre à un pas de 50 ms, nous pourrions alors avoir une courbe qui permette de superposer l'identification de deux types de sources en simultané.

Remarque : Pour des raisons industrielles, la société 01dB s'est retirée du partenariat établi, sans transmettre le code du logiciel, bloquant ainsi la poursuite des recherches. Comme nous souhaitons aller plus loin, nous nous sommes orientés sur d'autres pistes permettant l'identification.

IV. Identification par utilisation d'opérateurs de traitement du signal

Ce travail a été effectué en collaboration avec une branche de la société SONY dédiée aux recherches sur les sciences informatiques : Sony Computer Science Laboratory (SONY CSL Paris). Nous avons utilisé le logiciel EDS (Extractor Discovery System). Les résultats obtenus (formalisés par la mise en évidence d'opérateurs de traitement du signal) sont importants car ils exposent des descriptions intrinsèques des sources ou événements sonores étudiés.

EDS est un système en cours de développement qui n'a pas pour but d'être commercialisé. Il représente un dispositif puissant car il est capable d'effectuer une grosse quantité de calcul sans nécessité de nombreuses manipulations et qu'un grand nombre d'opérateurs de traitement du signal y est implémenté. Il représente un outil permettant de simplifier certaines tâches (visualisation des données, modèles de classification déjà implémentés, génération automatique d'un exécutable...).

4.1. Description d'EDS

Le logiciel EDS (Extractor Discovery System) permet de rechercher des descripteurs caractéristiques de la constitution d'une base de données sonores. A partir d'un ensemble de signaux étiquetés suivant une propriété donnée (intensité, tempo, type d'instrument, voix d'une chanteuse...), il permet de générer un modèle général de cette propriété, en construisant des fonctions de traitement du signal pertinentes. Toutefois, ce modèle général ne s'appuie uniquement que sur une description du problème qui lui a été posé à travers la base de données sonore disponible. Même si la base de données sur laquelle le modèle s'est construit peut sembler assez large (en termes de nombre et de variété de signaux) pour couvrir un vaste ensemble de cas, la robustesse du modèle ne sera pas forcément établie. En effet, il est toujours possible de trouver un modèle descriptif du problème posé, ne serait-ce qu'en employant une fonction de traitement adaptée à chaque unité de la base de données, mais il est évident que le modèle ne sera pas efficace sur de nouveaux sons. Ainsi, la robustesse d'un modèle se détermine avant tout par un nombre limité d'opérations qu'il doit effectuer et donc par l'utilisation d'opérateurs de traitement du signal pertinents. Ce modèle constitue en définitif le reflet d'un minimum de traits communs imputables aux signaux de la base de données. La base de données constitue donc l'élément central dans la recherche de descripteurs pertinents par une utilisation d'EDS.

4.2. Mode opératoire

4.2.1. Constitution d'une base de données

La base de données de *recherche* doit comporter des séquences sonores étiquetées de durées identiques. Cette durée est choisie en fonction de ce que l'on cherche à détecter. Une détection de sons brefs demandera des séquences de durée brève ; une détection de sons « longs » demandera des séquences de durée longue. Par ailleurs, l'étiquetage de chaque séquence est une valeur permettant de l'identifier au sein d'un groupe ou bien de l'identifier sur une échelle de valeurs. Seule une classification booléenne sera étudiée ici suivant le paradigme oui/non, les séquences prendront alors les valeurs « 0 » ou « 1 ».

Chaque séquence de la base de *recherche* possède ainsi une étiquette correspondant à l'ensemble auquel elle appartient. Par exemple, si l'on veut détecter un cyclomoteur, les séquences cyclomoteurs prendront la valeur « 1 », les autres séquences prendront la valeur « 0 ».

4.2.2 Recherche des fonctions

Le principe d'EDS est de découvrir des traits. Un trait est une fonction qui décrit un événement, une caractéristique ou une spécificité. Dans notre cas, les caractéristiques sont des types de sources sonores. Chaque fonction est une composition d'opérateurs de traitement du signal adaptés à la description de l'événement sonore. L'association de traits dans un modèle de classification forme un descripteur physique propre à l'événement sonore.

Beaucoup d'opérateurs connus dans la littérature sont implémentés dans EDS (une liste est fournie en Annexe 6 et leurs descriptions en Annexe 7). De plus, la recherche par combinaison de ces opérateurs est automatisée.

Ainsi, l'avantage principal de l'utilisation du logiciel est un gain de temps dans la recherche des traits. De plus cette recherche est très complète.

Le principe d'EDS est de chercher des fonctions permettant de caractériser le mieux possible les sons des séquences de la base de données de *recherche*. Une fonction (combinaison d'opérateurs du signal) sera qualifiée d'excellente si elle est capable de discriminer correctement les séquences étiquetées « 1 » des séquences étiquetées « 0 ». Pour cela, EDS crée tout d'abord une première population de fonctions (quarante par exemple) et les teste sur la base de données.

Il garde ensuite les meilleures fonctions discriminantes et en recombine les opérateurs suivant des règles logiques.

Soit EDS génère une première population peuplée d'opérateurs pris au hasard : c'est la recherche simple.

Soit on peut avoir une idée du type d'opérateur qui peut être pertinent pour caractériser une source (on peut par exemple s'aider d'un sonagramme). Dans ce cas EDS crée une population de fonctions de départ qui contient cet opérateur : c'est la recherche orientée.

La recherche s'arrête lorsque qu'un certain nombre de recombinaisons n'a pas amélioré la pertinence des fonctions.

4.2.3. Assemblage des opérateurs

Si l'on considère X comme étant une séquence sonore, une fonction, qui est un assemblage d'opérateurs de traitement du signal, peut avoir une expression de la forme :

$$f(X)=\text{Power}[\text{Log10}[\text{Abs}[\text{Max}[\text{SpectralCentroid}[\text{Split}[\text{Sqrt}[\text{Bartlett}[\text{Normalize}(X)],7421.0]]]],-1]$$

Les fonctions trouvées ont comme opérateur central ce que nous allons appeler des *opérateurs de bases*. Dans la fonction exposée ci-dessus, nous avons *Spectral Centroid* qui représente la fréquence barycentrique du spectre. Un opérateur de base est une description d'une caractéristique de l'agencement des composantes spectrales d'un son (ils sont définis dans l'Annexe 7). L'un d'eux est très souvent présent dans une fonction. Les opérateurs qui s'organisent autour (fenêtrage, filtrages, intensités...) servent à définir sa fenêtre de calcul pour y suivre son évolution.

On peut visualiser une représentation graphique des répartitions des valeurs booléennes attribuées à chaque séquence sonore en fonction des résultats de la fonction. Le graphe de la fonction est proposé ci-dessous [Figure 75] et indique le résultat de chaque séquence ainsi que la classification de chaque séquence (une séquence est un carré bleu). L'abscisse correspond à l'échelle des résultats de la fonction discriminante appliquée à chaque séquence et l'ordonnée représente la classification des séquences dans la catégorie « 1 » (cyclomoteur) ou « 0 » (non cyclomoteur).

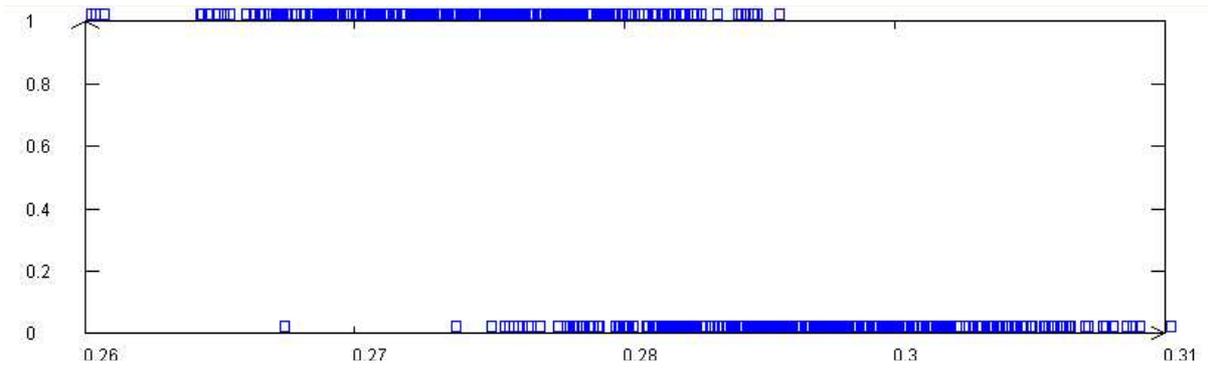


Figure 75 : Graphe de la fonction $f(X)=\text{Power}[\text{Log}10[\text{Abs}[\text{Max}[\text{SpectralCentroid}[\text{Split}[\text{Sqrt}[\text{Bartlett}[\text{Normalize}(X)],7421.0]]]]],-1]$ Les carrés bleus sont les séquences sonores de la base de données. Les séquences étiquetées « 1 » sont des cyclomoteurs, les séquences étiquetées « 0 » sont des sons de véhicules autres que des cyclomoteurs. En abscisse nous avons les valeurs des séquences par application de la fonction $f(x)$.

Ainsi, les points sur la ligne horizontale correspondant à la valeur « 1 » sont les valeurs que prennent les sons *cyclomoteurs* (de la base de données) étiquetés « 1 », par application de la fonction $f(X)$.

Par exemple,

$$f(\text{son_mob001}) = 0.27$$

Idem pour les points correspondant à la valeur « 0 » qui sont les valeurs que prennent les sons différents des cyclomoteurs (comme les *automobiles* par exemple) étiquetés « 0 », par application de la fonction $f(X)$.

Par exemple,

$$f(\text{son_automobile 001}) = 0.3$$

On comprend aisément que plus les groupes de valeur « 0 » et « 1 » sont éloignés, et plus la discrimination est réussie.

4.2.4. Pertinence d'une fonction

Dans le cas d'une discrimination booléenne entre deux classes A et B (A pour bus et B pour autres véhicules par exemple), la pertinence (ou fitness) d'une fonction s'évalue par la formule :

$$Pertinence(fitness) = \frac{2 \times (\mu_A - \mu_B)^2}{(\sigma_A^2 + \sigma_B^2)}$$

où, pour la classe i , μ = Moyenne des valeurs de la fonction sur les séquences,

et σ = Variance des valeurs de la fonction sur les séquences,

La pertinence de la fonction présentée au-dessus est $F = 1.66$. La figure 84 montre que cette fonction permet de discriminer un certain nombre de cyclomoteurs par rapport aux autres sources.

La valeur de pertinence d'une fonction sert à indiquer quelle fonction nous pouvons retenir pour constituer un modèle. Il faut toutefois s'assurer que les fonctions retenues soient faiblement corrélées. En général les fonctions retenues ont un critère de pertinence supérieur ou égal à « 0,6 » et on des *opérateurs de bases* différents.

4.2.5. Création d'un modèle par association de fonctions

La création d'un modèle s'effectue par l'association des meilleures fonctions de chaque population en utilisant différentes lois de construction (ou classificateurs). Nous avons retenu deux modèles : le modèle de type *régression linéaire* et le modèle de type *arbre de décision*.

Le modèle de type *régression linéaire multiple* sera de la forme :

$$F(f_i(séq)) = \sum_i \alpha_i f_i,$$

i : nombre de fonctions du modèle,

α : coefficient du modèle,

f : fonction du modèle

Un seuil discriminatoire du résultat du modèle permettant de classer les sons est alors fixé

Le modèle de type arbre de décision utilise les fonctions de la façon suivante :

$$F(f_i(séq)) = \begin{cases} si.f_1 \leq seuil \rightarrow f_2... \\ si.f_1 \geq seuil \rightarrow f_3... \end{cases}$$

f : fonction du modèle

Un compromis existe entre le nombre de fonctions dans le modèle et le minimum de fonctions utilisées pour une performance correcte (plus le nombre de fonctions dans le modèle est élevé, moins notre modèle pourra être exécuté en temps réel).

REMARQUE : La performance d'un modèle est calculée à partir de la base d'apprentissage. Il ne donne qu'une indication sur une détection correcte, mais ne donne absolument pas un taux de détection du modèle sur des séquences « *tests* ». Pour s'assurer du pouvoir discriminant de nos *fonctions*, on étudie la performance du modèle sur une *base de données tests* étiquetée dont les séquences sont différentes de celles utilisées dans la base d'apprentissage. Ceci permet d'obtenir une nouvelle évaluation sur des séquences non connues par le système d'apprentissage.

La méthode générale établie pour trouver des descripteurs est présentée sur la figure 76 ci-après.

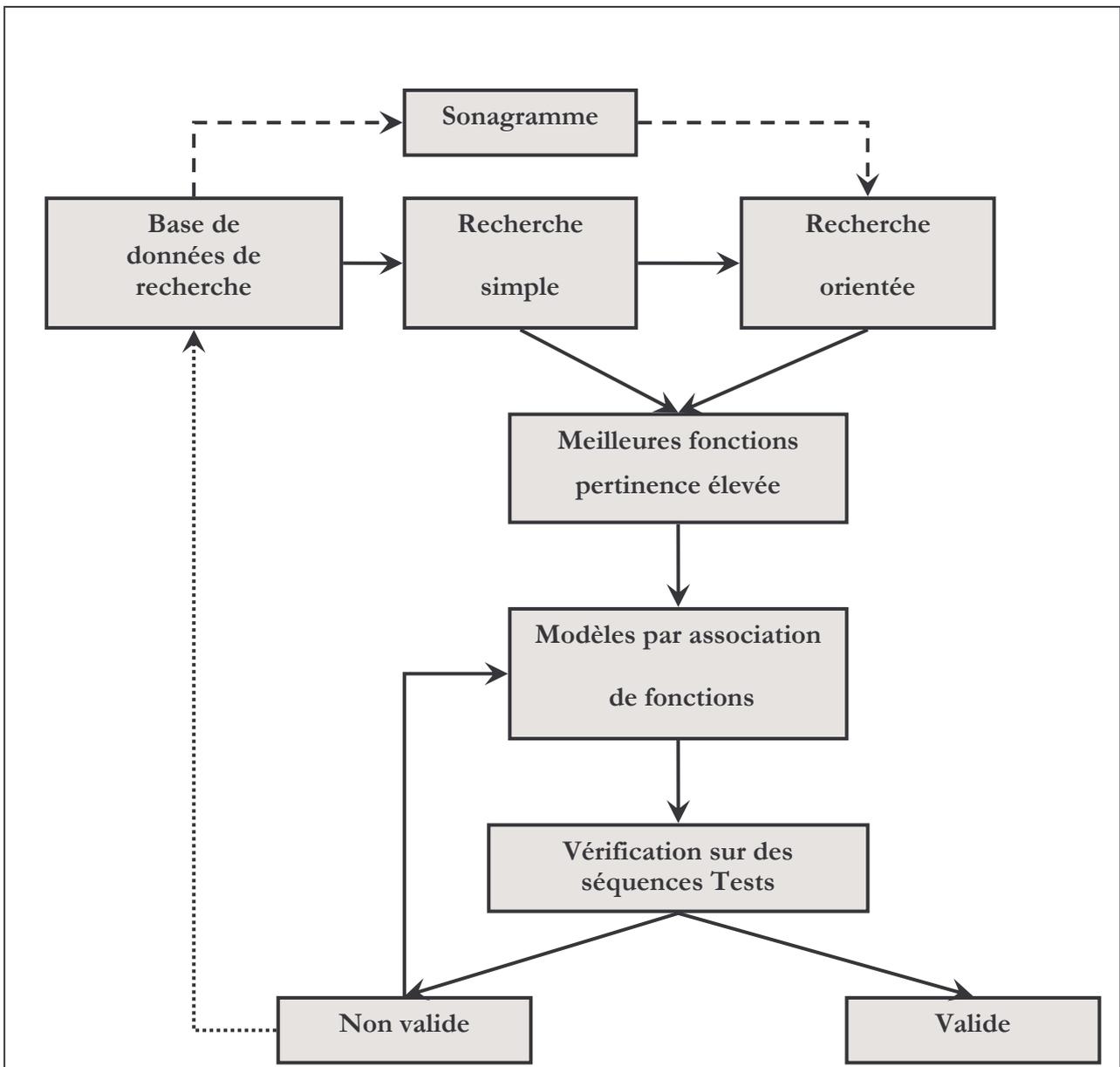


Figure 76 : Méthode générale pour la recherche de descripteurs.

4.3. Constitution d'une base de données

4.3.1. Métrologie et mesure du signal

Nous reprenons la même méthode que celle présentée au paragraphe 3.3.2.1 (chapitre 3) pour la constitution de la base de données ayant servie pour A2S. D'autres mesures ont été également effectuées.

Au total, plus de cinq cents séquences de véhicules et plus de cinq cents séquences différentes des véhicules composent la base de données sonore.

4.3.2. Détermination de la durée des séquences

La durée des séquences de *recherche* est choisie en fonction de la durée de l'émergence d'un son. Par exemple, un cyclomoteur est identifiable en moyenne pendant deux secondes. On peut considérer que la répartition temporelle de l'énergie s'effectue d'une manière gaussienne avec un maximum d'énergie pendant un quart de la durée du passage du cyclomoteur [Figure 77]. Cette durée d'énergie maximum sert donc de base pour les durées des séquences de *recherche* de la base de données pour les cyclomoteurs.

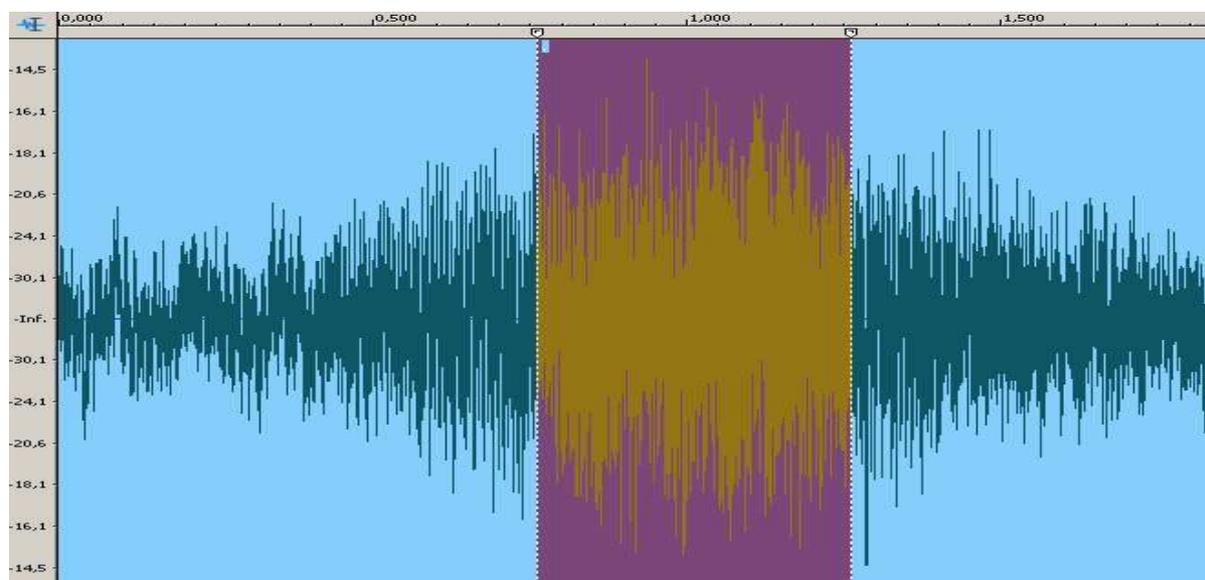


Figure 77: Amplitude d'un passage d'un cyclomoteur en fonction du temps.

Pour notre étude concernant la détection des cyclomoteurs, la durée pendant laquelle l'énergie est maximale est de 0.5 s. L'ensemble des séquences de la base données doit être par conséquent d'une durée maximale de 0.5 s.

Au moment de l'automatisation des mesures d'identification, le calcul pourra s'effectuer par pas de 0.1 s.

4.3.3. Normalisation des séquences

Dans notre méthode de détection, l'objectif est d'obtenir des fonctions s'appuyant sur des similarités fréquentielles de catégories sonores. Il est donc nécessaire de s'affranchir de l'aspect d'intensité des séquences. Les opérateurs utilisés dans les fonctions permettent de réunir des séquences sonores d'une même catégorie suivant un aspect temporel, spectral ou d'intensité, ou encore par des combinaisons entre ces trois points. Ainsi, l'intensité ne doit pas intervenir dans la discrimination. C'est pourquoi chaque séquence est égalisée sur son niveau maximum en dBLin qui est ainsi ramené au niveau numérique maximum avant saturation.

4.4. Classification véhicules/non véhicules

4.4.1. Pourquoi cette distinction ?

Après quelques premiers essais sur des bases de données comprenant un ensemble des sons de l'environnement urbain qui soit le plus exhaustif possible (véhicules, voix, oiseaux, klaxons, sirènes, bruits de pas, crissements de freins...), il est apparu que les meilleures fonctions trouvées ne pouvaient être assez bonnes pour discriminer une source de toutes les autres. La palette de sons était donc trop large et cela posait problème en particulier lorsqu'un événement n'avait pas de caractéristiques acoustiques différentes et assez robustes par rapport à toutes les autres sources sonores. Ainsi il fallait recomposer les bases de *recherche* pour « simplifier » le problème.

La répartition des individus de certaines fonctions trouvées et leur réécoute laissent à penser qu'une première distinction peut être effectuée. En effet, après l'écoute des stimuli ordonnés par certaines fonctions, il est apparu que les véhicules pouvaient être « assez facilement » discriminés des autres sources. De plus, instinctivement, il semble que les véhicules ont au moins un trait distinctif assez robuste par rapport aux autres sources : leur mouvement qui crée l'effet Doppler.

4.4.2. Méthode

Ainsi, en première approche, on peut considérer que la principale distinction entre les véhicules et les autres sons qui composent l'espace sonore urbain s'effectue sur le plan d'une variation spectrale dans le temps. Cette variation provient d'abord de la modification du régime

moteur des véhicule pendant leurs passages devant le point de mesure et ensuite de la variation naturelle de la fréquence suivant le principe de l'effet Doppler.

Le choix de la durée des séquences de la base de données doit donc être déterminé par la durée totale d'un passage de véhicule et par un temps suffisant pour obtenir une variation des fréquences.

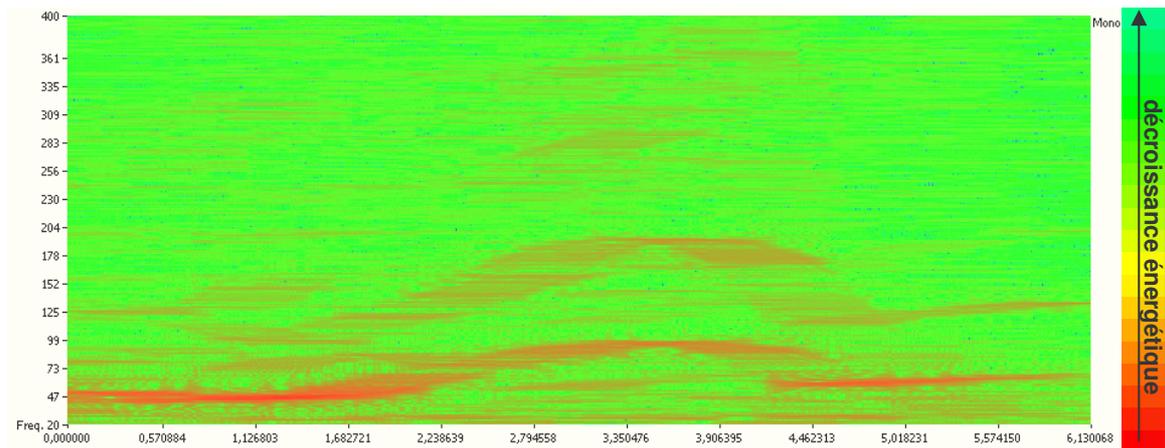


Figure 78 : Variations spectrales (de 20 à 400 Hz) d'une accélération du régime moteur dans le temps (s).

La figure 78 est un sonagramme qui indique les variations de l'énergie spectrale d'un passage de véhicule (automobile) dans la bande passante 20-400 Hz, par rapport au temps. Le choix d'une durée de 2.5 s semble convenable.

4.4.3. Résultats

Une base de données constituée de séquences véhicules et de non-véhicules de 2.5 s est créée. Au total, deux cent quatre-vingt-six séquences composent la base de données de *recherche*. Pour que les fonctions trouvées soient les plus robustes possibles, nous avons inclus dans la catégorie non-véhicules des sons tels que des voix, des bruits de pas, des bruits de poussettes...

Trois fonctions permettent effectivement une bonne séparation de séquences véhicules [Tableau 28].

N°	Fonctions	Pertinence
1	Square (Log10 (Rms (Variance (Sqrt (Mfcc (Bartlett (Split (Normalize (Testwav), 968.0)), 11.0))))))	3.65284876
2	Square (Log10 (Rms (SpectralDecrease (Split (Normalize (Testwav), 2813.0))))))	2.42002477
3	Rms (Sum (Blackman (Split (Normalize (Testwav), 3215.0))))	0.83843983

Tableau 28 : Meilleures fonctions descriptives des variations spectrales des véhicules sur le problème véhicule/non-véhicules, ainsi que leur pertinence.

Les fonctions trouvées (f1, f2) font ressortir deux opérateurs de base : *Mfcc* et *Spectral Decrease* qui sont des descriptions de la forme du spectre. La fonction construite autour de *Mfcc* indique une forme globale du spectre commune aux véhicules tandis que celle construite autour de *Spectral Decrease* décrit la variation temporelle de cette forme commune. Ces deux opérateurs semblent manifestement traduire la présence de l'effet Doppler dans le signal. La troisième (f3) ne contient pas d'opérateur de base. L'opérateur *Split* semble décrire une *cassure* de niveau à la fréquence 3215 Hz mais la fonction est difficilement interprétable.

On peut visualiser la répartition de chaque type de sons pour chacune des trois meilleures fonctions [Figures 79, 80 et 81]. En abscisse, on lit le résultat de la fonction appliquée à chaque séquence et en ordonnée, le classement des différentes sous-catégories : autos, oiseaux, bus, camions, chien, enfants, cyclomoteurs, motos, pas de piétons, poussettes, rien...

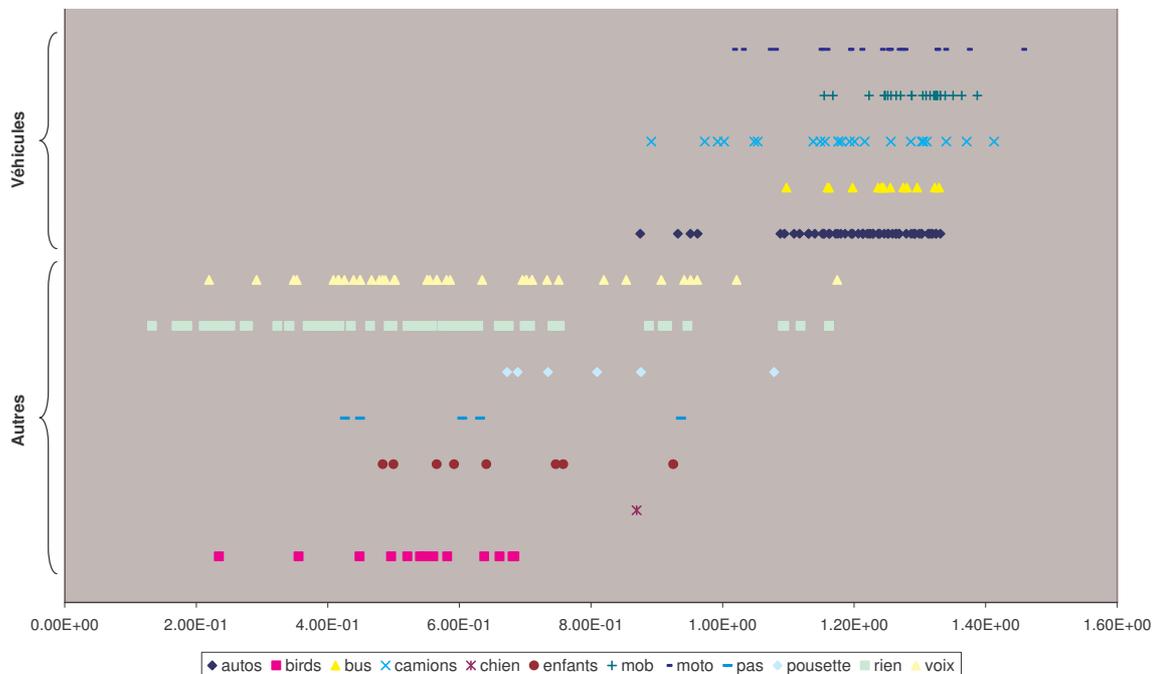


Figure 79 : Résultats de la fonction n°1 pour chaque sous-catégorie sonore. En abscisse, on lit le résultat de la fonction appliquée à chaque séquence et en ordonnée, le classement des différentes sous-catégories.

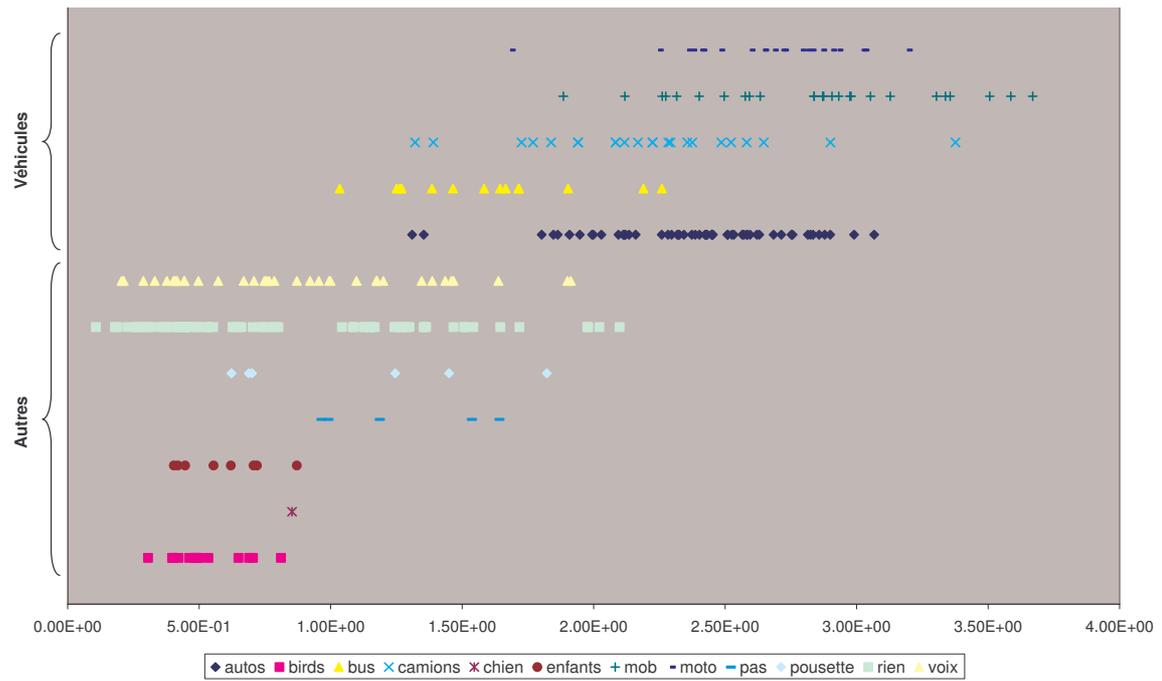


Figure 80 : Résultats de la fonction n°2 pour chaque sous-catégorie sonore. En abscisse, on lit le résultat de la fonction appliquée à chaque séquence et en ordonnée, le classement des différentes sous-catégories.



Figure 81 : Résultats de la fonction n°3 pour chaque sous-catégorie sonore. En abscisse, on lit le résultat de la fonction appliquée à chaque séquence et en ordonnée, le classement des différentes sous-catégories.

On constate que pour ces trois fonctions, les catégories *voix* et *rien* ont une répartition large sur l'échelle des valeurs en abscisse. Cela signifie que les séquences sont hétérogènes au sein d'un même groupe, ce qui montre la difficulté de trouver de bonnes fonctions discriminantes à la fois pour une catégorie et pour l'ensemble des catégories.

A partir de ces trois fonctions, le meilleur modèle que l'on peut construire est un modèle de type *arbre de décision*. Sa performance est élevée (97.55% de bonnes détections). La détection sur la base d'apprentissage a donc une erreur inférieure à 5%.

La figure 82 montre comment est organisé l'*arbre de décision* autour des trois fonctions avec indication des valeurs seuils.

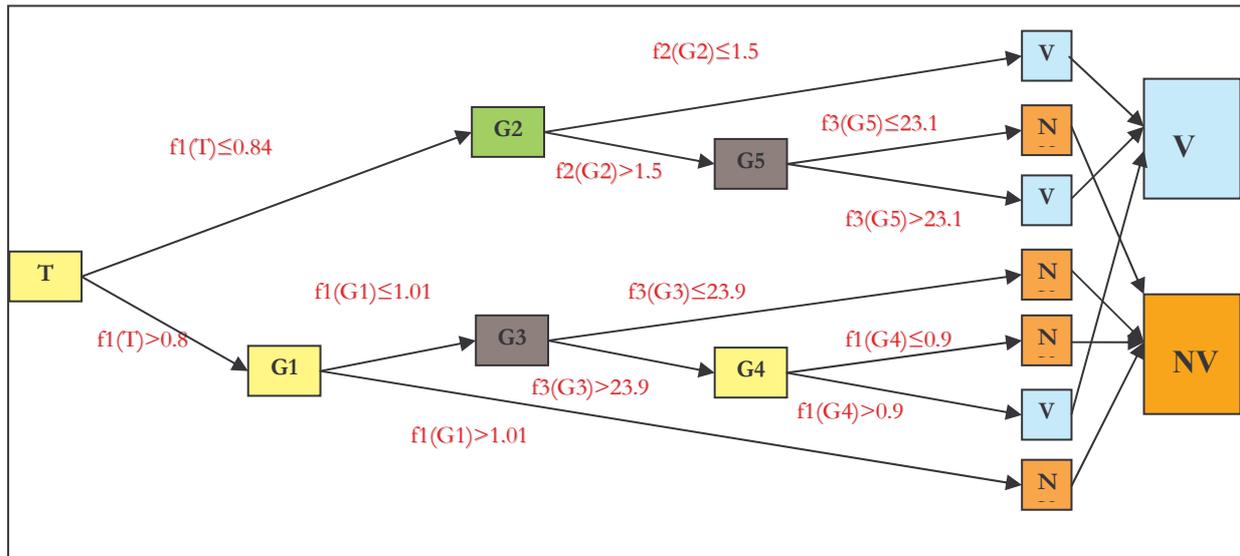


Figure 82: Arbre de décision du modèle de séparation des véhicules/non véhicules. Les trois fonctions utilisées sont numérotées de f_1 à f_3 , les groupes de séquences sont numérotés de G1 à G5 (T étant le nombre total de séquences). Les nombres indiquent la valeur seuil qui oriente les branches, V pour véhicules, NV pour non véhicules.

Ce modèle ou descripteur permet de réaliser la séparation suivante (Figure 83):

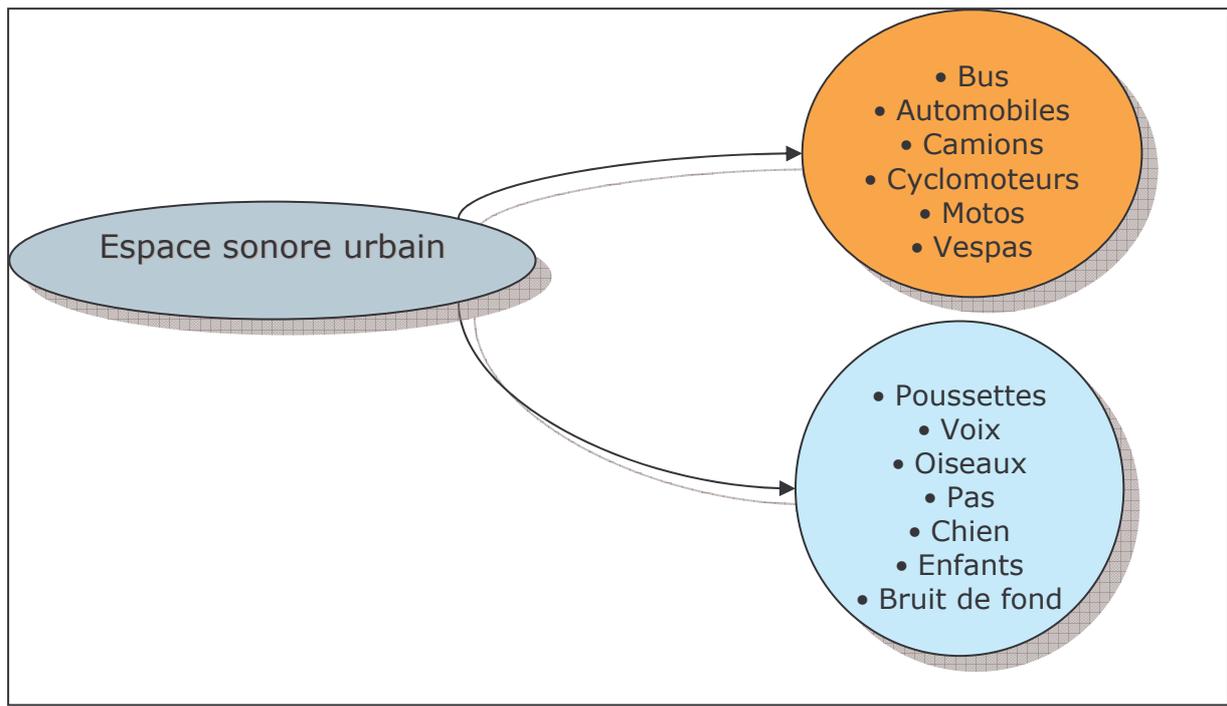


Figure 83 : Séparation effectuée par le modèle utilisant trois fonctions.

L'application de ce modèle sur des séquences tests donne des résultats satisfaisants aussi bien pour une détection par pas de 2.5s que pour une détection par pas de 0.5s.

4.4.4. Conséquences pour la suite des recherches

Par la suite, nous allons diriger nos recherches de fonctions en partant de cette première division.

Nous allons par exemple construire un descripteur permettant de discriminer les cyclomoteurs des autres véhicules. Cependant, ce descripteur ne sera pas forcément pertinent pour discriminer les cyclomoteurs des oiseaux.

Finalement, un cyclomoteur sera détecté comme telle s'il est déjà identifié comme véhicule.

4.5. Classification cyclomoteurs – autres véhicules

Avant d'effectuer les recherches en tenant compte de la première séparation véhicule/non véhicule, une recherche a été effectuée en tentant de séparer les cyclomoteurs de tous les autres types d'événement sonores. Les résultats sur plusieurs bases de données préliminaires ont permis de déterminer que certains sons classés dans une même catégorie sont identifiés séparément. Par exemple les vespas sont toutes identifiées comme étant non cyclomoteurs. Une écoute poussée de chaque séquence mal identifiée permet de confirmer les résultats d'EDS, c'est-à-dire que les vespas ne sont ni des motos, ni des cyclomoteurs. D'ailleurs, notre système auditif effectue très facilement la distinction entre les vespas et les autres deux temps.

Ainsi, les catégories *cyclomoteurs, voitures, motos, bus...* créées au départ sont amenées à évoluer. Des regroupements et des séparations de classes vont être effectués tout au long de ce travail, au gré des résultats d'EDS.

4.5.1. Base de données cyclomoteurs/véhicules

Une base de données composée uniquement de véhicules est créée. Cette base de données est constituée à 50% de sons de cyclomoteurs et à 50% d'autres sons. Au total, six cent soixante-dix-neuf séquences composent la base. Les voix, les oiseaux, les avions, les piétons, les klaxons sont supprimés, seuls sont gardés les ensembles bus, motos, autos, camions et Vespas, en plus des séquences de cyclomoteurs : la base de données est donc constituée uniquement de véhicules.

L'étiquetage des séquences est la valeur « 1 » pour les sons cyclomoteurs et la valeur « 0 » pour tous les autres véhicules.

4.5.2. Fonctions

Pour obtenir les meilleures fonctions, nous avons en premier lieu effectué des recherches simples. Celles-ci ont permis de dégager quelques opérateurs pertinents qui sont des opérateurs relatifs aux caractéristiques spectrales des sources : *Spectral Centroid*, *Spectral Flatness* et *Spectral Spread*.

Puis, en second lieu nous avons effectué des recherches orientées à partir des ces opérateurs, c'est-à-dire que le système a cherché les meilleurs combinaisons autour de ces opérateurs de base en y intégrant la possibilité de suivre leur évolution temporelle.

Nous exposons ci-dessous quatre fonctions trouvées parmi les plus pertinentes [Tableau 29] :

N°	Fonctions	Pertinence
1	Power (Log10 (Abs (Max (SpectralCentroid (Split (Sqrt (Bartlett (Normalize (Testwav))), 7421.0))))), -1.0)	1,6692
2	Power (Log10 (Abs (Max (SpectralCentroid (Split (LpFilter (Sqrt (Normalize (Testwav)), 21946.0), 7938.0))))), -4.0)	1,6133
3	Log10 (Abs (SpectralFlatness (Normalize (Testwav))))	1,5873
4	Abs (Log10 (SpectralSpread (Testwav)))	1,2939

Tableau 29: Quelques fonctions pertinentes de la description des variations spectrales des cyclomoteurs sur le problème cyclomoteur/véhicule ainsi que leur pertinence.

Ces quatre fonctions utilisent chacune un opérateur qui agit sur le spectre. Le suivi de leur comportement temporel correspond en effet aux variations de fréquences du bruit de moteur caractéristique des cyclomoteurs. Ces opérateurs semblent s'appuyer sur une caractéristique pertinente du son de cyclomoteur, à savoir son centre de gravité spectral.

Par exemple *Spectral Centroid* correspond au centre de gravité spectral, c'est-à-dire que cette fonction calcule la fréquence barycentre de l'énergie du spectre.

Voici le graphique indiquant la classification booléenne des séquences de la base de données en fonction du résultat obtenu sur chaque séquence par une simple application de la fonction pertinente (Figure 84) :

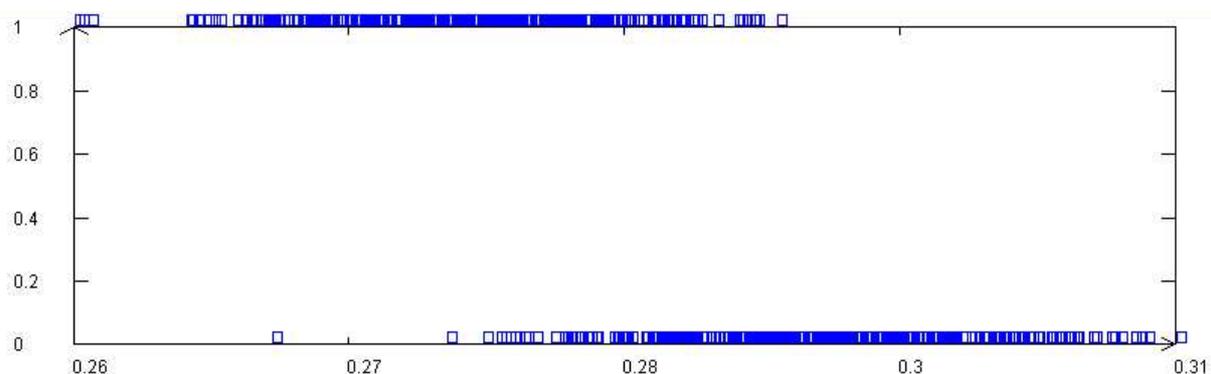


Figure 84: Graphe de la fonction n°1 ; les cyclomoteurs sont étiquetés « 1 », les véhicules non cyclomoteur sont étiquetés « 0 ».

Une autre fonction de traitement du signal ressort plusieurs fois dans les fonctions n°3, 9 et 10 : il s'agit de la fonction *Spectral Flatness*. Elle mesure la proximité du son avec un bruit blanc et

donne donc une information sur la *platitude* du spectre. Elle doit renseigner de la présence d'émergences spectrales mentionnées au paragraphe 3.2.2.

Voici le graphe généré par EDS pour la classification à partir de la fonction n°3 [Figure 85] :

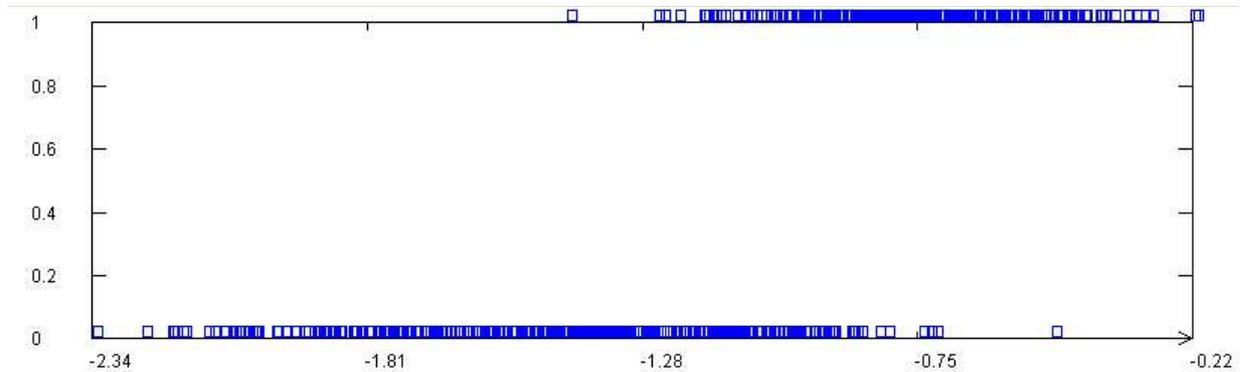


Figure 85: Graphe de la fonction n°3 ; les cyclomoteurs sont étiquetés « 1 », les véhicules non cyclomoteur sont étiquetés « 0 ».

Une troisième fonction de traitement du signal permet de caractériser les sons de cyclomoteurs : *Spectral Spread* ou la mesure de l'étendue de l'énergie spectrale autour du centre de gravité spectral. Elle intervient dans la quatrième fonction générée par EDS pour décrire les sons de cyclomoteurs. Voici le graphe montrant la répartition booléenne en fonction des résultats obtenus par l'application de la fonction sur les séquences de la base [Figure 86] :

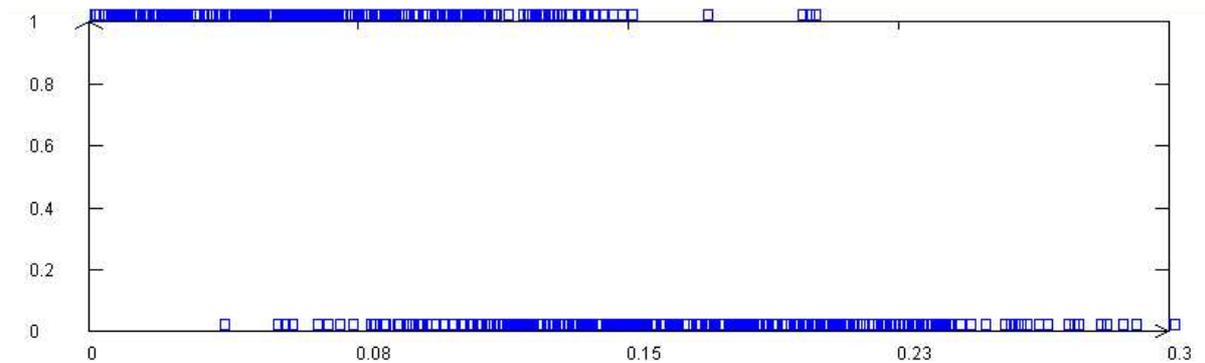


Figure 86: Graphe de la fonction n°4 ; les cyclomoteurs sont étiquetés « 1 », les véhicules non cyclomoteur sont étiquetés « 0 ».

Deux des trois opérateurs de base employés ici (*Spectrale Spread*, *Spectral Centroid*) traduisent une caractéristique forte du son de cyclomoteur, à savoir l'influence de son centre de gravité spectral.

4.5.3. Modèles

Si l'on fabrique un modèle intégrant l'ensemble des meilleures fonctions trouvées, on peut déterminer sa performance par application de celui-ci sur l'ensemble des *séquences d'apprentissage*. Les modèles dont les résultats sont exposés ici sont de type *arbre de décision*. Pour un calcul sur l'ensemble des échantillons, la performance est de 93.3% de bonnes détections. Voici la répartition des détections pour chaque ensemble de véhicules composant la base [Tableau 30] :

Véhicules	Cyclomoteurs		Motos		Vespas		Autos		Bus		Camions	
	DC	MD	DC	MD	DC	MD	DC	MD	DC	MD	DC	MD
Nombre	318	17	89	3	16	3	131	2	39	4	55	2
Taux	94,9%	5,1%	96,7%	3,3%	84,2%	15,8%	98,5%	1,5%	90,7%	9,3%	96,5%	3,5%

DC : détection correcte ; **MD** : mauvaise détection.

Tableau 30: Résultats des détections avec modèle associant trente-quatre fonctions.

Les résultats obtenus sont satisfaisants, mais le modèle utilise trente-quatre fonctions compliquées, ce qui entraîne un temps de calcul élevé et limite ainsi les possibilités de calcul en temps réel. L'idée est donc de simplifier le modèle afin de trouver un compromis entre la diminution du nombre de fonctions (et donc le temps de calcul) et l'augmentation de la marge d'erreur que cela entraîne dans la détection.

Si l'on crée un modèle composé de la meilleure fonction, c'est-à-dire la fonction tracée sur la figure 92, on obtient des résultats par catégories de véhicules légèrement moins bons par rapport au modèle précédent, mais avec une erreur globale inférieure à 10% [Tableau 31].

Véhicules	Cyclomoteurs		Motos		Vespas		Autos		Bus		Camions	
	DC	MD	DC	MD	DC	MD	DC	MD	DC	MD	DC	MD
Nombre	306	29	77	15	16	3	127	6	39	4	51	6
Taux	91,3%	8,7%	83,7%	16,3%	84,2%	15,8%	95,5%	4,5%	90,7%	9,3%	89,5%	10,5%

DC : détection correcte ; **MD** : mauvaise détection.

Tableau 31: Résultats des détections avec modèle utilisant uniquement la meilleure fonction.

Notons que seuls les chiffres concernant les cyclomoteurs sont intéressants, les autres chiffres sont à étudier d'un point de vue global car les fonctions sont sélectionnées pour être capables de séparer uniquement les cyclomoteurs des autres véhicules.

4.5.4. Optimisation d'un modèle

En sélectionnant les fonctions qui sont faiblement corrélées entre elles (par exemple celle qui utilisent des opérateurs centraux différents), nous augmentons nettement la performance du modèle à 94.7% de bonnes détections en utilisant simplement dix fonctions.

Les nouveaux résultats des détections par ce modèle sur l'ensemble des séquences témoignent d'une nette amélioration du taux de détection des véhicules qui composent la base de données [Tableau 32].

Véhicules	Cyclomoteurs		Motos		Vespas		Autos		Bus		Camions	
	DC	MD	DC	MD	DC	MD	DC	MD	DC	MD	DC	MD
Nombre	317	18	88	4	16	3	131	2	39	4	55	2
Taux	94,6%	5,4%	95,7%	4,3%	84,2%	15,8%	98,5%	1,5%	90,7%	9,3%	96,5%	3,5%

Tableau 32: Résultats des détections avec modèle utilisant les meilleures fonctions faiblement corrélées.

On remarque que tous les ensembles de véhicules ont une mauvaise détection inférieure à 10%, sauf la catégorie Vespa. Ceci s'explique par le petit nombre de séquences qui composent l'ensemble Vespa. Par contre, les cyclomoteurs ont un taux de bonnes détections de 94.6% et les automobiles de 98.5%.

4.5.5. Vérification du modèle sur des séquences tests

Pour l'instant, les résultats que nous avons obtenus sont les applications directes des modèles de détection sur des séquences qui ont elles-mêmes servi à créer les modèles. Il est donc nécessaire de vérifier l'efficacité de notre meilleur modèle sur des *séquences tests*, d'une durée très supérieure à 0.5 s pour valider notre méthode de détection.

EDS permet de générer un fichier exécutable capable d'appliquer le modèle à un fichier son. Le fichier de résultat comporte des valeurs booléennes établies toutes les 0.5 s.

Les tests sur deux séquences d'une durée de 100 s chacune, permet de visualiser la présence ou non des cyclomoteurs. Une présentation graphique permet de comparer les résultats de détection avec la présence réelle des cyclomoteurs [Figure 87 pour la séquence « 1 » et Figure 88 pour la séquence « 2 »].

La partie supérieure est la représentation du signal de la séquence test en fonction du temps, et la partie inférieure indique s'il y a une détection par la couleur orange (en foncé), ou pas de détection par la couleur bleu (en clair). Pour la séquence « 1 », deux cyclomoteurs sont détectés alors qu'un seul cyclomoteur est présent. En revanche, les deux cyclomoteurs de la séquence « 2 » sont bien détectés. En outre, la mauvaise détection des cyclomoteurs ne s'effectue que pour deux tranches de 0.5 s consécutives alors que les bonnes détections des cyclomoteurs le sont pour trois ou quatre tranches de 0.5 s consécutives.

Pour palier à des erreurs de ce type, nous pouvons toujours imaginer une amélioration de la détection en imposant qu'une détection soit valide uniquement pour un certain nombre de tranches consécutives, soit par exemple pour trois détections consécutives.

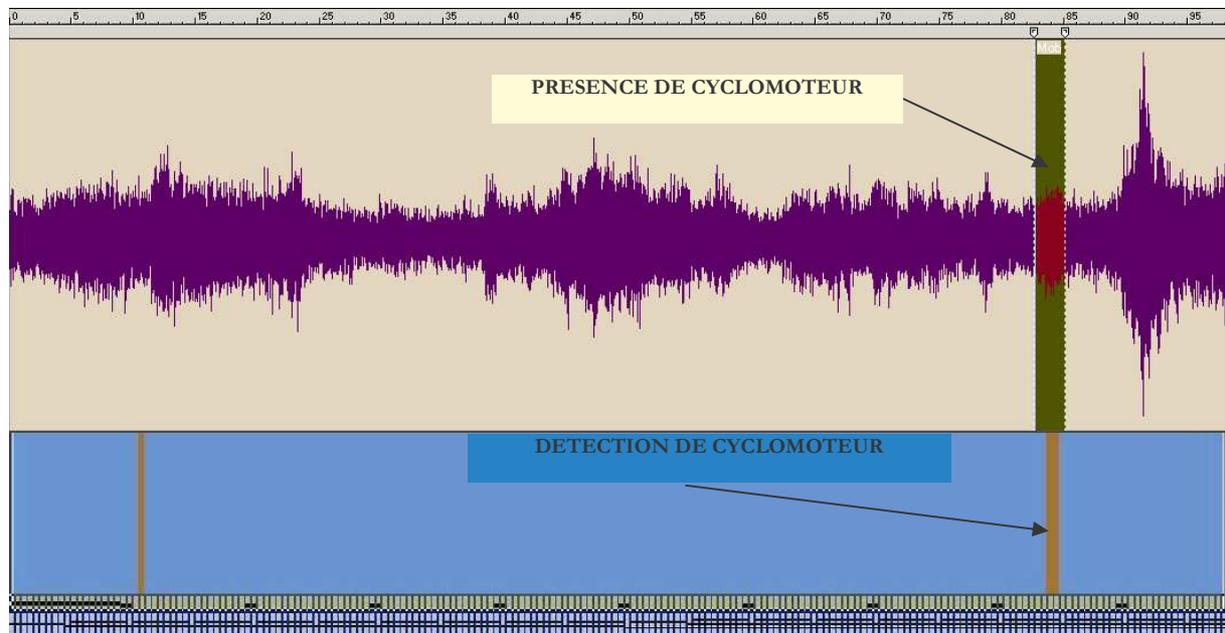


Figure 87: Vérification du modèle cyclomoteur sur la séquence test « 1 ». En haut nous avons le signal de 100s avec la présence de cyclomoteur détectée par une écoute experte, en bas nous avons la présence détectée par notre modèle.

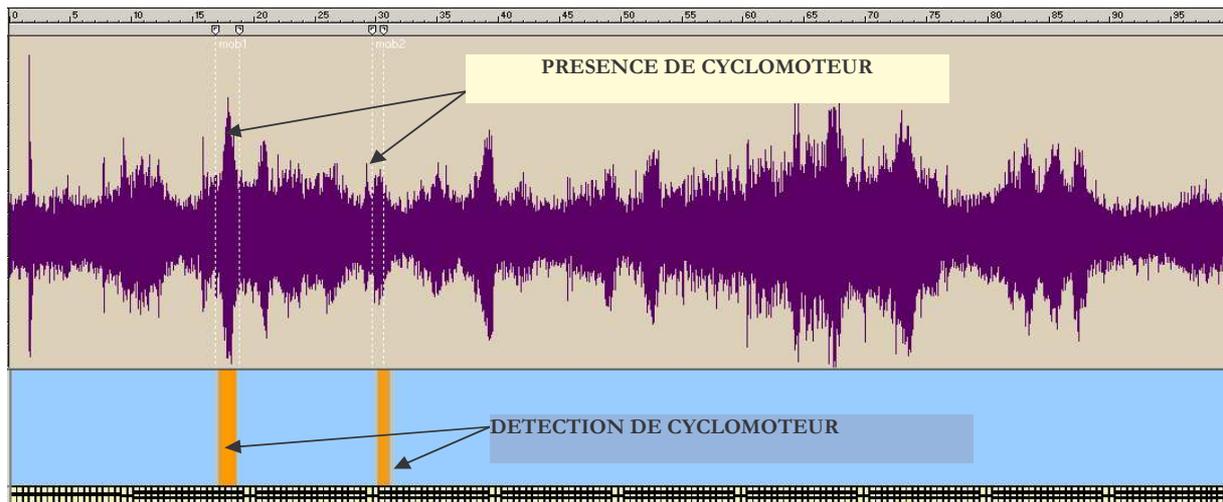


Figure 88: Vérification du modèle cyclomoteur sur la séquence test « 2 ». En haut nous avons le signal de 100s avec la présence de cyclomoteur détectée par une écoute experte, en bas nous avons la présence détectée par notre modèle.

4.6. Bus

Nous allons voir que la grande variété de signaux de bus présent dans l'espace urbain va nous obliger ici à recomposer les bases de données de départ. Nous allons alors trouver des fonctions pour chaque base de données construite. Au final, nous créerons un modèle général qui intégrera les meilleures fonctions trouvées pour les différentes bases de données.

4.6.1. Bases de données

Comme pour les cyclomoteurs, la durée choisie pour les séquences de bus de la base de données est de 0.5 s. Au départ, sept cent soixante-dix-sept séquences composent la base de données, soit 50% de séquences bus et 50% de séquences autres que bus. Ce nombre élevé doit permettre d'obtenir une robustesse de notre modèle malgré des valeurs de pertinence assez faibles pour les fonctions.

4.6.2. Fonctions

a) Base de départ B1

Une première recherche simple doit permettre de dégager certains opérateurs pertinents pour la détection des bus. Cette première approche a pour finalité des recherches orientées autour de fonctions utilisant ces opérateurs.

Cependant, aucune fonction pertinente ne permet d'effectuer des recherches orientées car elles ne sont pas caractéristiques du spectre des sons. La première recherche sur la base de données cyclomoteur a fait ressortir des opérateurs pertinents comme *SpectralKurtosis* ou *SpectralSpread*. Il était alors plus facile d'orienter par la suite les recherches autour de ces opérateurs, ce qui n'est pas le cas pour la détection des bus.

b) Base B2

Une nouvelle base de données est alors créée en sélectionnant des séquences de bus très caractéristiques et en ajoutant de nouveaux sons de bus. Certaines séquences sont supprimées. Au total, cinq cent quatre-vingt-dix-neuf séquences composent la base de données, dont deux cent soixante-dix sont des séquences de bus.

Nous obtenons des fonctions appelées $f(B2)$. Le meilleur modèle de type *régression linéaire* est composé de huit fonctions et obtient une performance de 90.15%.

Cette performance est encourageante puisqu'elle indique une incertitude inférieure à 10%.

Cependant, l'application de ce modèle sur une séquence de test donne des résultats totalement décevants : les bus ne sont pas détectés lorsqu'ils sont présents et les sons différents des bus (mais qui ne sont pas des véhicules) sont souvent détectés comme étant des bus.

La figure 89 montre que lors de l'exécution d'une fonction sur la base de données, certains sons, désignés dans le rond bleu, ne sont pas bien discriminés (il en est de même pour un modèle qui est une composition de fonctions). Ainsi, il y a lieu d'effectuer une nouvelle recherche de fonctions en créant une nouvelle base de données qui garde uniquement les sons mal détectés (base B3).

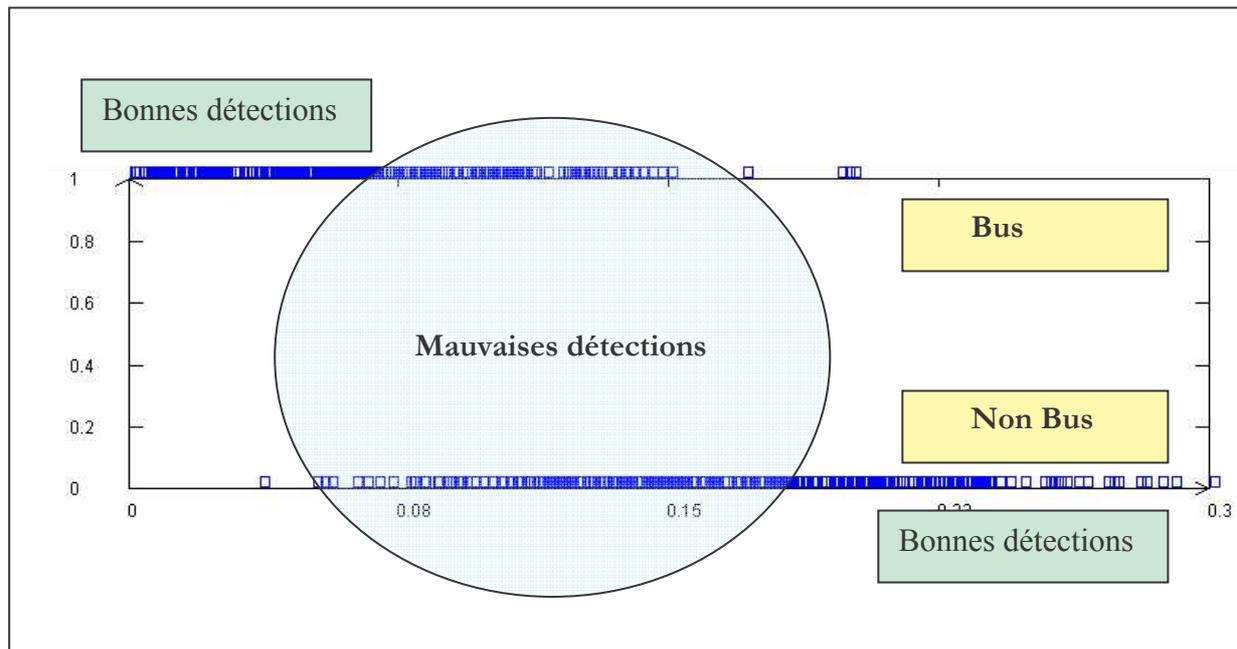


Figure 89 : Résultat d'une fonction sur la base de données B2. Au centre nous avons les sons mal détectés. Ces séquences mal ordonnées serviront ainsi à constituer une nouvelle base de données (B3).

En outre, lors des passages des cyclomoteurs, des motos ou des autos, la détection ne se fait pas. C'est un fait intéressant qui montre que l'on a besoin d'augmenter la pertinence des fonctions en leur demandant de discriminer les bus des sons qui ne sont pas des véhicules. Il faut alors envisager la constitution d'une nouvelle base de données plus complète cette fois-ci qui incorporerait l'ensemble des sons présents dans l'espace urbain (base B4).

c) Base B3

La base de données B3 est constituée de l'ensemble des séquences (bus comme non-bus) les moins bien détectées par le modèle fabriqué à l'aide de la base précédente (B2). Il s'agit ici de trouver des fonctions permettant d'établir une séparation des séquences entourées en bleu dans la figure 97. En d'autres termes, des séquences de bus mal détectées une première fois pourront être bien détectées par la suite, et inversement, des séquences autres que bus mais reconnues comme bus pourront être détectées en tant que non bus par la suite.

Nous obtenons ainsi des fonctions appelées $f(B3)$.

d) Base B4

La base de données B4 est constituée de l'ensemble des séquences composant l'espace sonore urbain. Voici la répartition des différentes catégories de la base [Tableau 33]:

Catégorie	Séquences	Quantité	Répartition
	bus	289	50.0%
Véhicules	autos	48	16.6%
	camions	48	
Différent de véhicules	oiseaux	20	16.6%
	enfants	22	
	pluie	3	
	poussettes	8	
	rien	22	
	voix	21	
Deux roues	cyclo	40	16.8%
	motos	40	
	vespas	17	
Total		578	

Tableau 33: Composition de la base de données B4

Nous obtenons ainsi des fonctions appelées $f(B4)$.

4.6.3. Modèle

A présent, il faut construire un modèle permettant de détecter le mieux possible les sons de bus à partir des différentes fonctions trouvées par EDS à l'aide des différentes bases de données de calcul.

Le premier modèle [Figure 90] construit est *un arbre de décision* qui combine certaines des fonctions $f(B2)$ et $f(B4)$.

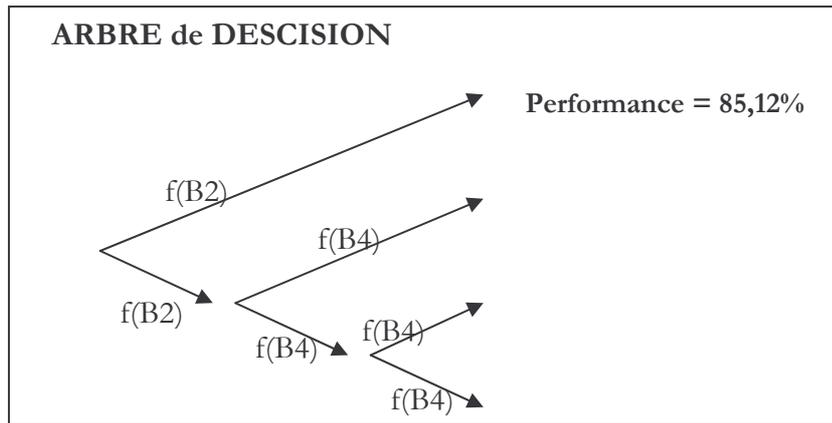


Figure 90 : Meilleur arbre de décision obtenu sur la base B4. Ce modèle utilise des fonctions (différentes ou identiques pour chaque branche) extraites parmi $f(B2)$ et $f(B4)$.

Un autre modèle créé automatiquement par EDS, de type *arbre de décision* est créé à partir du calcul de la performance de celui-ci sur la base de donnée B2. Il utilise certaines des fonctions parmi $f(B2)$, $f(B3)$ et $f(B4)$. Ce modèle est schématisé sur la Figure 91 ci-dessous :

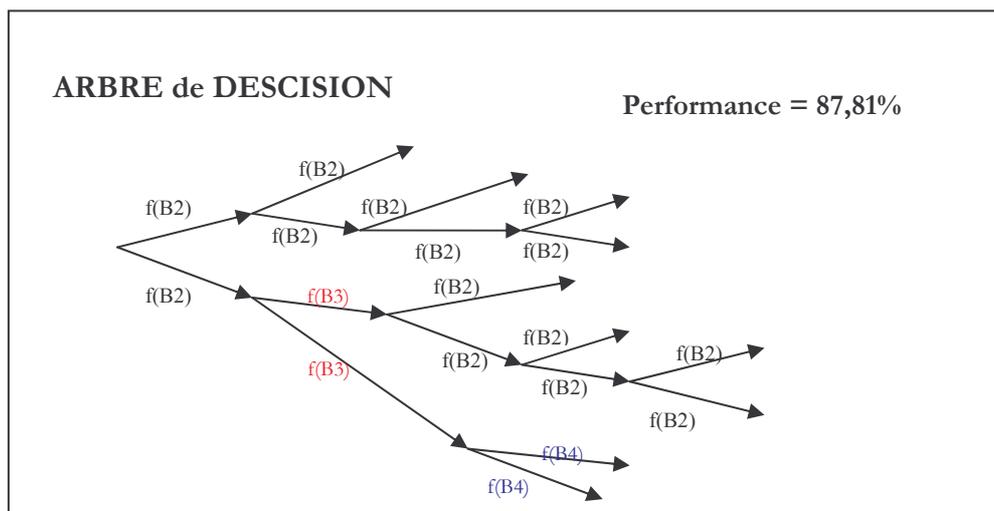


Figure 91: Meilleur arbre de décision obtenu sur la base B2. Ce modèle utilise des fonctions (différentes ou identiques pour chaque branche) extraites parmi $f(B2)$, $f(B3)$ et $f(B4)$.

On constate que les arbres de décisions générés par le logiciel sont difficilement interprétables par rapport à ce que l'on pourrait attendre comme hiérarchie logique. De plus, EDS ne permet pas de paramétrer l'arbre suivant notre logique, mais plutôt suivant une logique de recherche automatique du meilleur modèle en utilisant les fonctions qui lui sont présentées. Ainsi, il est intéressant d'imaginer quel type de modèle pourrait correspondre à nos attentes.

4.6.4. Modèle d'arbre décisionnel logique

D'après les différents résultats obtenus sur les trois bases de données, si l'on considère un modèle de combinaisons de fonctions de type arbre de décision, le choix que doit effectuer le modèle est d'abord une première sélection des séquences non-bus, puis ensuite une sélection des bus dans un panel comprenant des bus et d'autres véhicules et enfin une vérification de chaque sous-sélection pour déterminer si les séquences bus sont bien détectées ou non. C'est ce qui est transcrit dans la figure 92.

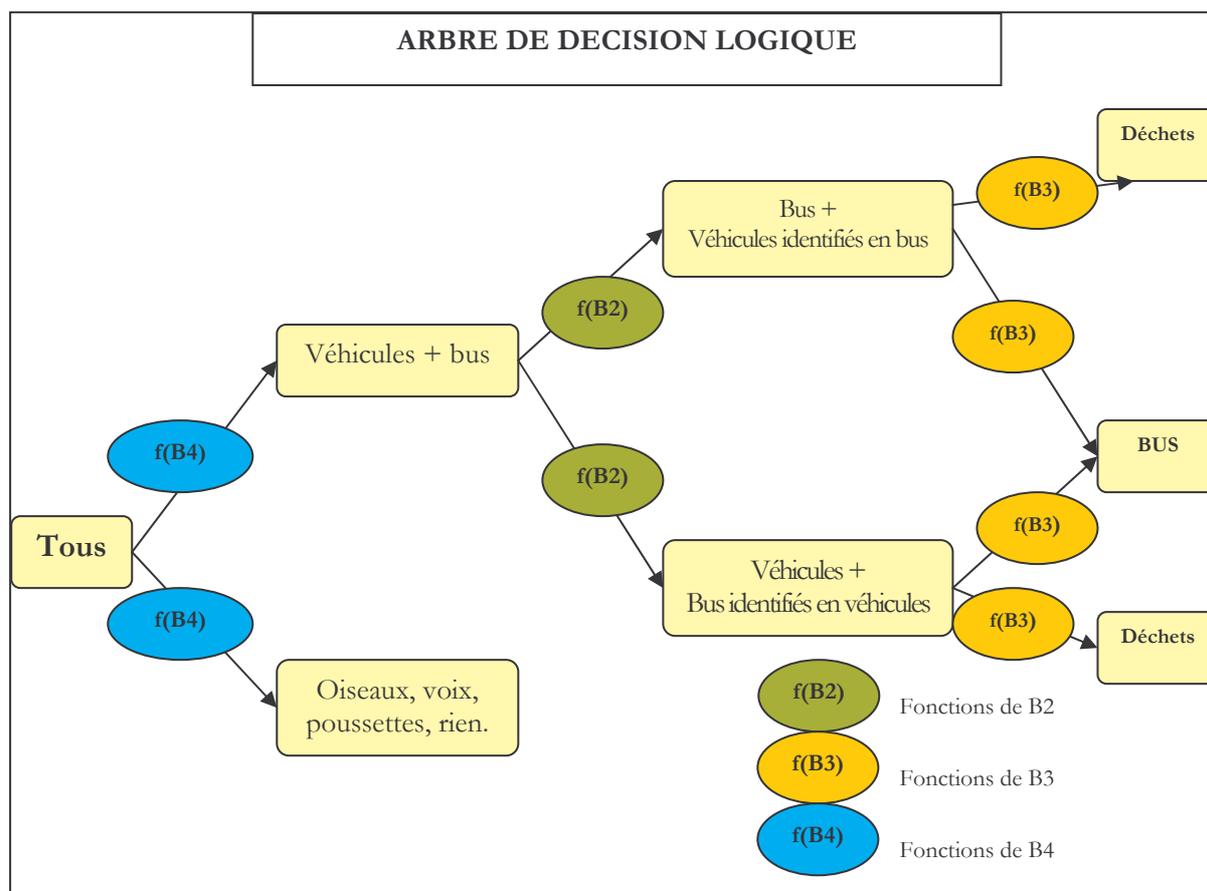


Figure 92 : Modèle à envisager permettant une détection optimale des bus dans l'espace sonore urbain.

Ce schéma illustre le fait que la détection doit se faire par étapes, c'est-à-dire qu'une première sélection des véhicules et des non-véhicules doit être opérée pour détecter ensuite plus facilement les sous catégories comme les bus, les autos, les cyclomoteurs et les motos. Il en est de même pour détecter les oiseaux ou les voix où il faut dans un premier temps déterminer une catégorie non-véhicule.

La recherche sur l'identification des autobus a permis de construire un nouveau type de modèle. Classiquement un modèle était construit à partir de fonctions issues de recherches sur une unique base de données. Désormais, pour obtenir des meilleurs résultats, il est parfois

nécessaire de créer des modèles intégrant des fonctions issues de recherches sur des bases de données différentes.

4.7. Autres sources

En reprenant la méthode que nous avons établie et indiquée à la figure 76 nous obtenons des résultats satisfaisants pour l'identification des voitures et des oiseaux. Des pistes de recherches sont indiquées pour l'identification des voix et des motos qui demandent à être étudiées d'avantage.

4.7.1. Automobiles

La *base d'apprentissage* est constituée uniquement de séquences de véhicules. 50% des séquences sont des sons d'automobiles et les 50% restants sont divers véhicules. Au total, deux cent soixante-seize séquences composent la *base de données d'apprentissage*. La durée des séquences est de 0.5s.

Plusieurs recherches préliminaires sont effectuées et permettent de connaître les opérateurs les plus adéquats pour décrire les véhicules comme *Spectral Flatness*, *Spectral Centroid*, *Spectral Kurtosis* qui sont des descriptions fréquentielles des signaux.

Puis ces opérateurs sont rentrés dans les recherches orientées qui, grâce à des combinaisons avec des opérateurs temporels ou énergétiques, permettent de suivre l'évolution temporelle des caractéristiques spectrales.

Les meilleures fonctions obtenues par opérateur de bases sont les suivantes (Tableau 34a) :

Fonctions	Pertinence
Square (Arcsin (Power (SpectralCentroid (Bartlett (HpFilter (Sqrt (Testwav), 1299.0))), -1.0)))	0,56
Square (Log10 (Abs (SpectralFlatness (Sqrt (Normalize (BpFilter (Normalize (Testwav), 1008.0, 3751.0)))))))	0,82
Log10 (SpectralFlatness (Sqrt (BpFilter (Normalize (Testwav), 928.0, 2841.0)))	0,75
Log10 (Rms (Percentile (FilterBank (Hamming (PeakPos (Blackman (Normalize (Testwav))), 4.0), 60.0)))	0,53
Square (Log10 (Multiplication (Variance (MelBands (Derivation (Normalize (Testwav)), 3.0), 35415.0)))	0,43
Power (Abs (SpectralKurtosis (Testwav)), -0.5)	0,30

Tableau 34a : Meilleures fonctions obtenues pour les voitures.

La pertinence des fonctions obtenues est moins bonne que celle des fonctions obtenues pour la classification véhicule/non véhicule ou pour les cyclomoteurs. En effet, le meilleur score est ici

de 0.82 contre 3.65 et 1.66 précédemment. Cela traduit le fait que, dans les problèmes discriminatoires présentés, les voitures ont des caractéristiques spectrales moins prononcées que les cyclomoteurs ou les véhicules.

Finalement, à partir de ces six fonctions, deux modèles sont fabriqués. Leurs performances sont les résultats obtenus par l'application des modèles sur la *base de données d'apprentissage* [Tableau 34b].

Modèle	Performance
Arbre de décision	84.06%
Régression linéaire	86.59%

Tableau 34b : Meilleurs modèles composés à partir des meilleures fonctions.

Plusieurs vérifications de ces modèles sur des *séquences tests* indiquent que les fonctions obtenues sont optimum pour créer les meilleurs modèles permettant d'identifier l'ensemble des séquences de la catégorie automobile.

Notons également que tous les véhicules de type *utilitaires* sont détectés comme étant des voitures. Ce regroupement a déjà été effectué dans d'autres études qui traitaient des similarités spectrales entre véhicules (voir [GAULIN-00] pour une bibliographie). Les *utilitaires* ne sont différents des voitures que d'un point de vue visuel.

4.7.2. Motos

Les premiers résultats concernant la détection des motos nous entraîne à diviser la catégorie *motos* en sous catégories comme nous avons déjà pu le faire en séparant les sons de *vespas* et les sons de *cyclomoteurs*. De ce fait, la difficulté pour trouver de bons descripteurs réside dans la constitution de nouvelles bases de données qui doivent avoir un nombre suffisant d'individus.

Néanmoins, une écoute approfondie des séquences de motos incite à séparer naturellement à l'oreille l'ensemble des sons de motos en quatre classes de sons. Ces classes sont les suivantes :

- Motos au ralenti,
- Motos à vitesse normale,
- Explosion continue du pot d'échappement (Type Harley Davidson...),
- Motos de grosse cylindrée,

Ainsi, on peut prédire que l'ensemble des *fonctions* que l'on obtiendra permettra de détecter chaque sous-catégorie de motos. Une somme des détections de chaque sous-ensemble indiquera alors une détection de toutes les motos.

Il apparaît que nos bases de données sonores des quatre catégories de motos ne sont pas assez fournies en nombre pour nous permettre d'obtenir des descripteurs pertinents.

4.7.3. Oiseaux

La base de données est composée de deux cent soixante-dix séquences dont la répartition en pourcentage par source est la suivante [Tableau 35] :

oiseaux	50.0%
animaux	0.7%
piétons	7.8%
voix	22.2%
avertisseur	13.3%
autres	5.9%

Tableau 35: Composition de la base de données « oiseaux »

Pour identifier les oiseaux, nous avons guidé les recherches à partir d'indices trouvés sur la lecture d'un sonagramme. Sur la figure 93, on visualise le signal ainsi que le spectrogramme d'un son d'oiseau. La bande passante de cet oiseau s'étend de 5000 à 9500 Hz. Dès lors, un filtrage passe-haut semble être un bon départ pour caractériser les sons d'oiseaux dans l'espace urbain

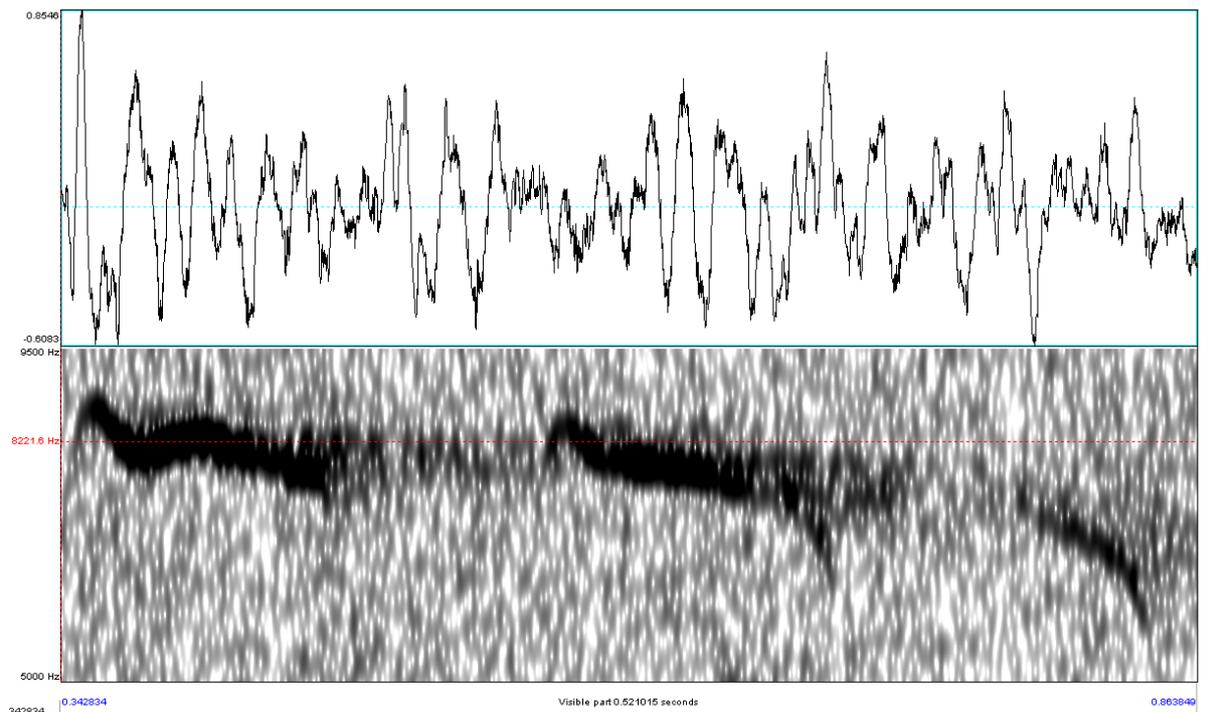


Figure 93: Signal et sonagramme d'un son d'oiseau.

Plusieurs visualisations de spectrogrammes de différents oiseaux montrent que la fréquence de coupure du filtre passe-haut est optimale pour 1500 Hz.

Nous effectuons ainsi un premier filtrage du signal avec l'emploi dans les fonctions de l'opérateur *HpFilter*.

Au final, nous pouvons obtenir un modèle de type « arbre de décision » qui permet d'obtenir une performance de 90.37% sur *la base d'apprentissage*. Ce modèle classe correctement cent vingt séquences d'oiseaux et cent vingt-quatre séquences de non-oiseaux. Les erreurs portent sur quinze séquences d'oiseaux mal détectées et onze séquences de non-oiseaux mal détectées.

Ce modèle fonctionne suivant le schéma suivant (Figure 94) :

```

Log10Log10SpectralRolloffPowerNormalizeCorrelationPowerNormalizeTestwav4.0NormalizeTestwav7.0 <= 0.411583
| SpectralRolloffDerivationNormalizeDerivationDerivationNormalizeDerivationNormalizeTestwav <= 18834: c0 (13.0)
| SpectralRolloffDerivationNormalizeDerivationDerivationNormalizeDerivationNormalizeTestwav > 18834
| | Log10VarianceVariancepitchBandsLpFilterSplitOverlapBpFilterNormalizeHpFilterNormalizeTestwav31.0579.01083.03593.00.890328145749266212891.05.0 <= -0.432409: c1 (87.0/3.0)
| | Log10VarianceVariancepitchBandsLpFilterSplitOverlapBpFilterNormalizeHpFilterNormalizeTestwav31.0579.01083.03593.00.890328145749266212891.05.0 > -0.432409
| | | Log10Log10SpectralRolloffPowerNormalizeCorrelationPowerNormalizeTestwav4.0NormalizeTestwav7.0 <= 0.372532: c0 (10.0)
| | | Log10Log10SpectralRolloffPowerNormalizeCorrelationPowerNormalizeTestwav4.0NormalizeTestwav7.0 > 0.372532: c1 (13.0/5.0)
Log10Log10SpectralRolloffPowerNormalizeCorrelationPowerNormalizeTestwav4.0NormalizeTestwav7.0 > 0.411583: c0 (60.0)

```

Figure 94: Description du modèle de type "arbre de décision" utilisant trois fonctions pour détecter les sons d'oiseaux.

Parmi les trois fonctions utilisées dans ce modèle, deux utilisent l'opérateur *SpectralRolloff* et une l'opérateur *PitchBands*.

SpectralRolloff est une approximation de la fréquence de coupure entre les parties sinusoïdales et bruitées du signal. Il caractérise l'aspect harmonique du chant d'oiseau.

L'opérateur *PitchBands* est quant à lui, utilisé après le filtrage passe-haut. Il représente une mesure de la *hauteur* perçue. Cet opérateur caractérise la *hauteur tonale* du chant d'oiseau.

4.7.4. Voix

Les premiers essais sur les voix n'ont pas été satisfaisants. C'est étonnant car le logiciel a été développé pour des applications musicales. Bien que des recherches plus poussées puissent être effectuées, ce résultat est un peu décevant car la détection des voix n'était également pas satisfaisante avec le logiciel A2S. Aucun *opérateur de base* disponible n'a pu être extrait comme étant pertinent. Des essais peuvent par exemple être effectués en orientant la recherche autour d'un opérateur caractéristique de la voix humaine reflétant une modulation d'énergie comprise entre 3 et 20 Hz. Cette modulation traduit la fréquence du mouvement des organes producteurs de phonèmes (le palais, la langue, la mâchoire et les lèvres) qui fonctionnent par cycles complets d'une durée comprise entre 50 et 300 ms [STEVENS-98]. Cet opérateur n'est pas disponible au moment de notre recherche.

4.8. Ebauche d'un organigramme des détections

Confrontés tout d'abord à des résultats satisfaisants pour classifier les bases de recherche mais très décevants pour identifier les sources sur des séquences tests, nous avons dû changer notre approche en redéfinissant le problème à résoudre. En examinant la composition des modèles obtenus, nous avons constaté qu'ils employaient un nombre trop élevé de fonctions différentes. Les modèles s'avérant ainsi uniquement descriptifs des sons de la base de données sur laquelle le système avait effectué leur recherche. Ainsi, l'absence de fonction pertinente (dû à la non présence d'un opérateur de base) est significative d'une base de données trop éclectique en termes de diversité. Nous avons alors décomposé le problème présenté en plusieurs sous-problèmes pour ensuite intégrer l'ensemble des résultats dans un modèle général.

De ce fait, pour identifier correctement ce que nous classons habituellement dans la catégorie cyclomoteur, nous devons identifier tout d'abord les véhicules ; puis, parmi ces véhicules, nous recherchons indépendamment plusieurs catégories (Vespa, scooter ou cyclomoteur de 50 cm³ de cylindrées) que nous regroupons ensuite sous la dénomination de cyclomoteurs. Pour cela il nous faut effectuer une écoute attentive des séquences pour établir de nouvelles catégories et s'assurer que ces différences perceptibles peuvent être détectées par des opérateurs de traitement du signal, et non pas à l'aide de facteurs cognitifs.

Les opérateurs les plus utiles aux détections sont ceux qui décrivent des particularités spectrales des sources sonores. Ce sont les opérateurs du type *Spectral Centroid*, *Spectral Decrease*, *Mfcc*, *Spectral Flatness*, *Spectral Spread*... Les autres opérateurs (fenêtrages, calculs d'énergie, filtres...) permettent essentiellement de définir le cadre (temporel, spectral et énergétique) de recherche des traits pertinents ainsi que de suivre leur évolution.

Les résultats trouvés avec EDS permettent ainsi d'établir un organigramme pour les détections des sons de l'environnement urbain (Figure 95).

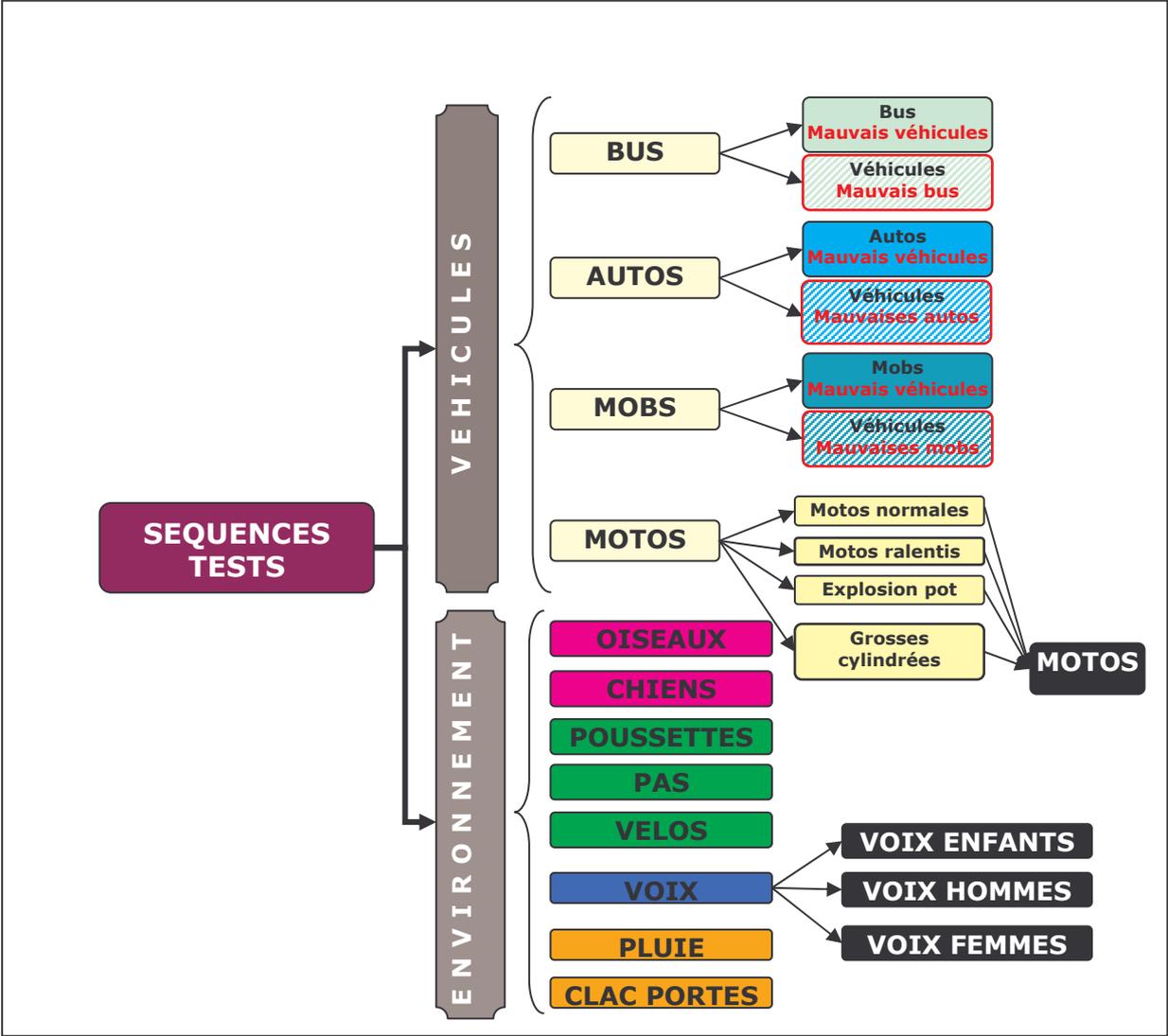


Figure 95: Organigramme des détections tel que nous pouvons l'entrevoir et qui pourra se dessiner dans la suite des recherches.

4.9. Validation des descripteurs sur les vingt séquences tests de 15 s.

A titre d'exemple, nous montrons les détections effectuées sur le stimulus n°2 par pas de 0.5 s (Figure 96). Chaque ligne correspond au résultat du calcul d'un descripteur sur les 15 s de la séquence. La présence d'une couleur est le signe d'une détection positive.



Figure 96 : Détections effectuées sur le stimulus n°2 par pas de 0.5 s. Chaque ligne correspond au résultat du calcul d'un descripteur sur les 15 s de la séquence. La présence d'une couleur est le signe d'une détection positive.

Le cyclomoteur en vert est effectivement détecté comme cyclomoteur et comme véhicule.

Les voitures (auto-500ms) en jaune sont également détectées comme voitures et comme véhicules.

Il y a deux erreurs concernant la détection de bus. Cependant, un bus n'apparaît que rarement sur une durée de 500 ms seulement. Il est donc aisé de conclure que c'est une erreur.

Quant aux oiseaux (*birds*), ils sont bien présents au début du stimulus. Ils sont détectés en même temps que les véhicules. Les véhicules font partie du bruit de fond.

Ainsi, sur des stimuli complexes, des règles d'interprétation sont parfois nécessaires pour conclure à la présence ou non d'une source (voir paragraphe 4.5.5). Ces stratégies employées à

travers la mise en œuvre de règles d'interprétation sont à rapprocher des traitements de « haut niveau » décrits au paragraphe 2.1.3. A titre de comparaison, la présence d'un bus, détectée sur une période unique de 0.5 s, mais finalement rejeté est l'équivalent *computationnel* de l'indice appelé *information contextuelle*.

Ces remarques renforcent l'idée qu'un système d'identification a parfois besoin, au même titre que le système perceptif humain, d'indications complémentaires aux seules informations acoustiques disponibles.

CONCLUSION GENERALE & PERSPECTIVES

« Vous allez entendre vingt séquences d'une durée de 15 s chacune. Imaginez-vous lors d'une promenade dans Paris telle que vous l'avez déjà vécue. Par exemple vous allez à un rendez vous mais vous n'êtes pas pressés ».

Telle est la consigne donnée aux auditeurs au début des tests perceptifs qui ont eu lieu tout au long de ce travail.

C'est donc à travers des expériences perceptives que nous avons entrepris de mieux comprendre comment les citoyens évaluent leur environnement sonore quotidien. Tout au long de cette démarche expérimentale, nous sommes partis du principe que d'une part l'évaluation de la qualité sonore à travers l'expression du désagrément, est fonction du lieu et de l'activité du sujet et que d'autre part, l'unité perceptive qui lui permet d'évaluer cet environnement sonore est la source productrice de bruit car elle lui donne un sens.

En premier lieu, les résultats de ces expériences, à travers les jugements des auditeurs, mettent en avant des observations qui nous permettent de formuler des hypothèses explicatives de nature à mieux appréhender la perception de séquences sonores. Elles apportent en outre des éléments qui à termes peuvent constituer des paramètres utiles au management de l'environnement sonore placé sous le signe de la qualité de vie des citoyens.

I. Evaluation d'une ambiance urbaine

Ces hypothèses expérimentales ont permis d'observer et de mettre en évidence les caractéristiques suivantes, quant à l'évaluation d'une ambiance sonore :

Dans une ambiance de rue avec passages de différents types de véhicules (bus, cyclomoteurs, motos...), les auditeurs sont sensibles à une représentation spatiale des voitures plutôt que temporelle ou liée à l'intensité. Ces observations tendent à montrer que, contrairement aux autres véhicules, considérés comme des « objets bruits » car émergeant clairement du bruit de fond et évalués selon leur caractère prégnant, les voitures sont perçues comme faisant partie intégrante de la route.

Le désagrément sonore est fortement corrélé à la force sonore. Cependant, nous pouvons améliorer cette corrélation par l'adjonction de critères temporels relatifs à des sources sonores. Ces sources ont comme caractéristique de ressortir de l'ambiance sonore en attirant l'attention. Elles peuvent avoir un effet positif comme les voix d'enfants, ou un effet négatif comme les cyclomoteurs et les bus. L'influence positive des oiseaux n'est en revanche pas observée.

Bien que pertinente pour une ambiance de type rue circulée, la force sonore n'est plus adaptée lorsque l'on veut évaluer le caractère hédonique d'une ambiance de type parc ou marché.

Ce sont alors les sources sonores occurrentes et leur caractère temporel qui rendent compte du côté agréable ou désagréable de l'ambiance. De plus, dans ces ambiances qualifiées de calmes ou agréables, les voitures sont alors perçues comme « objet bruit ». Il est alors possible d'expliquer le désagrément sonore uniquement en tenant compte de la composition de l'ambiance et donc sans avoir recours à des considérations énergétiques. Ainsi, le fait de compter les véhicules en présence ou bien de tenir compte de leur temps d'apparition permet d'être très proche du désagrément.

Ces observations illustrent le fait que la présence d'une source suffit à l'auditeur pour établir un jugement qualitatif d'une ambiance.

Ainsi, sur la base d'une hypothèse qui soutient que « c'est la signification imputée aux signaux à partir de l'identification de la source qui contribue au caractère plaisant, désagréable ou pénible de la stimulation sensorielle » [GUASTAVINO-03], nous avons établi un nouvel indicateur qui ne se base plus uniquement sur le niveau sonore mais sur la composition de l'environnement. Cet indicateur est calculé à partir de plusieurs modèles de type régression linéaire multiple qui ont été déterminés pour des séquences de relatives courtes durées (15 s). Les variables des modèles sont relatifs à la durée d'apparition des sources sonores. Les modèles s'adaptent au type d'ambiance (marché, rue de quartier et parc) selon la valeur des coefficients.

Une étude en laboratoire et une étude in situ ont permis d'obtenir plusieurs observations sur des séquences plus longues (5 min en laboratoire et 45 min in situ).

L'intégration perceptive de critères liés aux sources sonores a tendance à s'effectuer sur les événements marquants des séquences.

Par contre, l'intégration perceptive du niveau sonore et du désagrément sonore a tendance à être évalué à partir de valeurs moyennes de la séquence.

Ainsi, en tenant compte de ces observations, il est possible d'extrapoler la mesure du désagrément à partir de séquences courtes sur des séquences plus longues.

II. Détection automatique des sources

Notre second but étant également de qualifier la qualité d'un environnement sonore urbain par la mesure et de manière simple, nous avons exposé deux méthodes d'identification de sources.

La première méthode a permis d'obtenir un outil simple qui est capable de détecter des sources ayant des caractéristiques spectrales prononcées telles que les voitures et les cyclomoteurs.

Cette méthode permet d'extraire, à travers l'évaluation de l'*activité* des sources, une mesure de la prégnance et de la proximité des cyclomoteurs ainsi que de la prégnance des voitures dans une ambiance urbaine, sans nécessiter de réécoute des stimuli par un opérateur.

Cependant, il y a manque de robustesse dès qu'il existe une trop grande variabilité de cas, par exemple pour les motos, les bus ou les voix.

Pour la seconde méthode, l'utilisation du logiciel EDS sur des bases de sonorités urbaines a été réalisée pour la première fois. Cela a permis de mettre en évidence des descripteurs physiques solides capables de détecter les principaux événements sonores de l'environnement urbain parisien. Les descripteurs trouvés ne sont bien sûr pas universels mais permettent de résoudre un problème ad hoc. Ils s'appuient sur la pertinence d'opérateurs de traitement de signal qui sont descriptifs du comportement spectral des sources sonores.

En partant de classes de sources pré-déterminées, il a été rapidement constaté que les similitudes spectro-temporelles extraites des signaux ne sont pas forcément en accord avec les classes correspondant à nos images perceptives. Ainsi, un son de cyclomoteur de type Vespa, s'il rentre a priori dans une catégorie perceptive dite « cyclomoteur », ne l'est plus a posteriori. Une écoute attentive des sons de Vespa nous permet toutefois de les distinguer des sons de

cyclomoteurs classiques. Ainsi, la taxinomie effectuée ici se construit à partir des similarités spectro-temporelles des événements sonores, contrairement aux propriétés visuelles du départ.

Dans notre recherche de descripteurs pertinents, nous avons également vu qu'une base de données trop large (en termes de diversité de sources sonores) pouvait nuire à l'extraction de fonctions intéressantes. La démarche nous a permis de constater que nous obtenons de meilleurs résultats en orientant les recherches selon l'arborescence d'un arbre de décision : en formant des classes en partant du plus général et en allant au plus particulier. Ainsi, la détection d'un bus sera meilleure si l'on effectue en premier lieu une différenciation entre véhicules et non-véhicules, puis en appliquant un second descripteur faisant la différence entre bus et véhicules plutôt que de chercher directement identifier des bus parmi l'ensemble des sources de l'environnement urbain.

Les premiers descripteurs physiques établis ont montré leur efficacité sur les nombreux tests de vérification effectués. En employant une méthodologie à peu près similaire à celle présentée ici, il semble que des descripteurs relatifs aux autres sources (voix et motos) employées dans nos modèles de désagrément pourront être trouvés.

Il faudra également, pour renforcer la robustesse des descripteurs, étoffer les bases de données de *recherche* de sons plus divers, que nous pouvons rencontrer en milieu urbain : passages d'avions, d'hélicoptère, musique...

Enfin, le but de ces recherches étant d'obtenir un outil permettant l'identification des sources en temps réel, les modèles définitifs retenus devront être optimisés en fonction du nombre de fonctions et opérateurs auxquels ils font appel. De plus, les compositions d'opérateurs devront certainement être « épurées » (par exemple si nous avons le calcul de la valeur absolue de la racine carrée) afin d'économiser du temps de calcul. L'identification constitue un élément essentiel pour pouvoir automatiser le calcul de la qualité sonore d'une ambiance telle que nous l'avons caractérisée au chapitre 2 de ce mémoire.

Ces perspectives sont en cours de réalisation dans le cadre du projet de recherche « Bruit et Nuisances Sonores » parrainé par le ministère de l'Environnement et du Développement Durable. Ce projet, intitulé « Evaluation de la qualité des ambiances sonores urbaines à l'échelle du quartier : représentations cartographiques adaptées à la communication avec les riverains, à partir d'une approche perceptive et au travers d'indicateurs pertinents » court jusqu'à fin 2006.

BIBLIOGRAPHIE

[AFSSE-94]

Agence française de la Sécurité Sanitaire Environnementale. (2004).

Impact sanitaire du bruit, état des lieux, indicateurs bruit-santé.

Mai 2004

<http://www.grenoble.fr/download/bruit04.pdf>

[AMPHOUX-91]

AMPHOUX, P. (1991).

Aux écoutes de la ville. La qualité sonore des espaces publics européens – Méthode d'analyse comparative – Enquête sur trois villes suisses.

Rapport de recherche, IREC. Ecole Polytechniques Fédérales de Lausanne, n°94. Lausanne.

[ANDERSON-83]

ANDERSON, L.M., MULLIGAN, B.E., GOODMAN, L.S., REGEN, H.Z. (1983).

Effects of sounds on preferences of outdoor settings.

Environment and Behavior, 15(5), 539-566.

[ASA-60]

AMERICAN ACOUSTICAL SOCIETY. (1960).

Acoustical Terminology, S1-1

American Standard Association, New York, 1960

[ATTIAS-97]

ATTIAS, H., SCHREINER, C.E. (1997).

Temporal low-order statistics of natural sounds.

Advances in Neural Info Processing Systems, 9, edited by M. Mozer (MIT Press, Cambridge, MA), 27-31.

[AUDITORY-05]

Furum auditory@lists.mcgill.ca

www.auditory.org, 2005

[BAGOT-96]

BAGOT, J.D. (1996).

Information, sensation et perception

Armand Colin.

[BALLAS-82]

BALLAS, J.A., HOWARD, J.H., KOLM, C. (1982).

Listener descriptions of isolated and patterned acoustic transients.

Psychology documents 12,3

[BALLAS-86]

BALLAS, A. J., SLIWINSKY, M.J. (1986).

Uncertainty and response time in identifying non-speech sounds.

The Journal of the Acoustical Society of America, 86, S 47.

[BALLAS-87]

BALLAS, J.A., HOWARD, J.H. (1987).

Interpreting the language of environmental sound.

Environment and behaviour, 19, 91-114, 1987.

[BALLAS-93]

BALLAS, A.J. (1993).

Common Factors in the Identification of an Assortment of Brief Everyday Sounds.

Journal of Experimental Psychology, Human Perception and Performance, 19, N° 2, 250-267, 1993.

[BARTHET-03]

BARTHET, M. (2003).

Détection automatique d'insectes xylophages

Rapport de stage mars-mai 2003. LAM, LRMH, 01dB-Stell.

[BÉKÉSY-60]

BEKESY, G. (1960).

Experiments in Hearing.

E. G. Wever Ed., Mc Graw Hill, New York, 1960.

[BERGLUND-04]

BERGLUND, B., NILSSON, M.E. (2004).

Soundscape perceived in built environment.

Proc. ICA 2004, Kyoto, 2004.

[BERGSON-00]

BERGSON, H. (1900).

Le rire. Essais sur la signification du comique.

Presses Universitaires de France.

[BERNARD-94]

BERNARD, Y. (1994).

Connaître et se représenter un espace.

Le courrier du CNRS n° 81, La Ville.

[BISMARCK-74]

BISMARCK, G. (1974).

Timbre of steady sounds: a factorial investigation of its verbal attributes.

Acustica, 30, 146-158.

[BJÖRK-84]

BJÖRK, E.A. (1984).

The perceived quality of natural sounds.

Acustica, Volume 57, 1985.

[BLAUERT-97]

BLAUERT, J., JEKOSCH, U. (1997).

Sound quality evaluation – a multi-layered problem.

Acustica . Acta Acustica., 83, 747-753.

[BLAUERT-94]

BLAUERT, J. (1994).

Product-sound assessment: an enigmatic issue from the point of view of engineering.

Proc. Internoise 94, Yokoama, 2, 857-862.

[BONEBRIGHT-01]

BONEBRIGHT, T.L. (2001).

Perceptual scaling of everyday sounds: a multidimensional scaling approach.

Proceeding of the 2001 International conference on auditory display, Espoo, Finlande, 2001.

[BORG-02]

BORG, E., BORG, G. (2002).

A comparison of AME and CR100 for scaling perceived exertion.

Acta Psychologica, 109, 157-175.

[BOTTE-88]

BOTTE, M.C., CANEVET, G., DEMANY, L., SORIN, C. (1988).

Perception de l'intensité sonore.

Psychoacoustique et perception auditive, Editions INSERM/SFA/CENT, 13 – 41, Paris.

[BOULLET-05]

BOULLET, I. (2005).

Sonie des sons impulsionnels : Perception, mesures et modèles.

Thèse de doctorat de l'université d'AIX-MARSEILLE II, 2005.

[BREGMAN-90]

BREGMAN, A.S. (1990).

Auditory scene analysis: The perceptual organisation of sound.

Cambridge, MA. MIT Press.

[BREGMAN-94]

BREGMAN, A.S. (1994).

L'analyse des scènes auditives.

Penser les sons, PUF, 11-39.

[BROADBENT-58],

BROADBENT, D.E. (1958).

Perception and communication.

Oxford : Pergamon Press.

[BRUIT-04]

Bruit des infrastructures de transports terrestres, plan national d'actions contre le bruit du 6 octobre 2003.

Le Ministère de l'Ecologie et du Développement Durable, Le Ministère de l'Equipeement des Transports, de l'Aménagement du Territoire, du Tourisme et de la Mer, Le Ministère de l'Emploi, du Travail et de la Cohésion sociale, 25 mai 2004.

[CAN-05]

CAN, A. (2005).

Modèle de caractérisation du désagrément sonore en milieu urbain, basé sur l'identification des sources, élaboré à partir d'un test sur site.

Mémoire de Master Recherche « mécanique, physique et Modélisation » spécialité acoustique. Université Aix Marseille. Collaboration LASA, Université de Cergy Pontoise. Mars – août 2005.

[CANEVET-00]

CANEVET, G. (2000).

Eléments de psychoacoustique.

Cours du DEA Acoustique et dynamique des vibrations de l'Université d'Aix Marseille II. 2000.

[CANEVET-99]

CANEVET G., SCHARF B., SCHLAUCH R., TEGHTSOONIAN R., TEGHTSOONIAN M. (1999).

Perception of changes in loudness.

Nature, Letter to the Editor, 398, 673-674.

[CARELLO-98]

CARELLO, M., CANDERSON, K.L., KUNKLER-PECK, A.J. (1998).

Perception of object length by sound.

Psychological Science, 9, 211-214.

[CASEY-98]

CASEY, M. (1998).

Auditory Group Theory: with Applications to Statistical Basis Methods for Structured Audio, Ph.D. Thesis, MIT Media Lab, February 1998.

[CONDAMINES-78]

CONDAMINES, R.. (1978).

Stéréophonie. Cours de relief sonore théorique et appliqué.

MASSON, 1978.

[COWLING-03]

COWLING, M., SITTE, R. (2003).

Comparison of techniques for environmental sound recognition.

Pattern Recognition Letter 24, 2895-2907.

[DAMASIO-03]

DAMASION, A.R. (2003).

Spinoza avait raison, joie et tristesse, le cerveau des émotions.

Odile Jacob, Paris.

[DARWIN-59]

DARWIN, C. (1859).

On the origin of species.

Londres: John Murray.

[DELCAMPE-02]

DELCAMPE, D. (2002).

Les perspectives issues de la directive du 25 juin 2002 relative à l'évaluation et à la gestion du bruit dans l'environnement.

Ministère de l'écologie et du développement durable, Nov. 2002.

http://www.infobruit.org/ministeres/publications/sources/medd-trans_04.pdf

[DEFREVILLE-03]

DEFREVILLE, B., LAVANDIER, C., DUFOURNET, D. (2003).

Activity of urban sound sources.

Congrès Euronoise, 2003, Naples, Italie.

[DESS-01]

Illustration sans référence précise issue du Cours du DESS « Bases physiques pour l'acoustique architecturale et urbaine ». (2001).

Université PARIS VI.

[DIRECTIVE-02]

PARLEMENT EUROPEEN ET CONSEIL DE L'UNION EUROPEENNE. (2002).
Directive 2002/49/CE du 25 juin 2002 relative à l'évaluation et la gestion du bruit dans l'environnement.
Journal officiel des communautés européennes. 18.7.2002.
http://www.ecologie.gouv.fr/IMG/pdf/DirectiveUE_BruitAmbiant_25-06-2002.pdf

[DODGE-04]

DODGE, Y., ROUSSON, V. (2004).
Analyse de régression appliquée.
DUNOD, 2^{ème} édition, 2004.

[DUBOIS-04]

DUBOIS, D., GUASATAVINO, C., RAIMBAULT, M. (2004).
Les catégories cognitives du bruit urbain : des discours aux indicateurs physiques.
Acoustique & Techniques, 39, 4ème trimestre 2004.

[DUFAX-00]

DUFAX, A., BESACIER, L., ANSORGE, M., PELLANDINI, F. (2000).
Automatic sound detection and recognition for noisy environment.
Proceedings European Signal Processing Conference. Tampere, Finland, 2000.

[DURBIN-51]

DURBIN, J., WATSON, G.S. (1951),
Testing for serial correlation in least squares regression, II .
Biometrika, 38, 159-178, 1951.

[ELLERMEIER-04]

ELLERMEIER, W., MADER, M., DANIEL, P. (2004).
Scaling the unpleasantness of sounds according the BTL model: Ration-scale representation and psychoacoustical analysis.
Acta Acustica United with Acustica. 90, 101 – 107. (2004).

[ESCOFIER-98]

ESCOFIER, B., PAGES J. (1998).
Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation.
DUNOD. 3^{ème} édition, 1998.

[FASLT-91]

FASTL, H. (1991).

Evaluation and measurement of perceived average loudness.

Fifth Oldenburg Symposium on Psychological Acoustics, Oldenburg, BIS 205-216.

[FERRINGTON-03]

FERRINGTON, G. (2003).

Soundscape studies: listening with attentive ears.

The journal of media literacy. Vol.49/50, Numéro 1, printemps 2003.

[FREED-90]

FREED, D.J. (1990).

Auditory correlates of perceived mallet hardness for a set of recorded percussive events.

The Journal of the Acoustical Society of America, 87, 311-322.

[GABRIELSSON-79]

GABRIELSON, A., SJOGREN, H.. (1979)

Perceived sound quality of sound reproducing systems

The Journal of the Acoustical Society of America, 65(4), 1019-1033.

[GAULIN-00]

GAULIN, D. (2000).

Caractérisation et identification des sources sonores.

Thèse de Doctorat de l'université du Maine, 2000.

[GAVER-88]

GAVER, W. (1988).

Everyday listening and auditory icons.

Université de Californie, San Diego.

[GAVER-93]

GAVER, W. (1993).

What in the World Do We Hear? An Ecological Approach to Auditory Event Perception.

Ecological Psychology, 5, 1-29, 1993.

[GAVER-93]

GAVER, W. (1993).

How do we hear in the world? Explorations in ecological psychoacoustics.

Ecological Psychology, 5, 285-313, 1993.

[GIBSON-66]

GIBSON, J.J. (1966).

The senses considered as perceptual systems.

Boston: Houghton Mifflin.

[GYGI-01]

GYGI, B. (2001).

Factors in the identification of environmental sounds.

Thèse de doctorat de l'Université d'Indiana (USA), département de psychologie, Juillet 2001

[GOLDHOR-93]

GOLDHOR, R..S. (1993).

Recognition of environmental sounds .

Proceedings of ICASSP, New York, USA, 1, 149-152. (1993).

[GOTTSCHLING-99]

GOTTSCHLING, G. (1999).

On the relations of instantaneous and overall loudness.

Acustica – acta acustica, 85, 427-429.

[GRIFFITHS-04]

GRIFFITHS, T.D., WARREN, J.D. (2004).

What is an auditory object?

Nature Reviews, Neuroscience. Vol. 5, novembre 2004.

http://www.staff.ncl.ac.uk/t.d.griffiths/Griffiths_NRN_2004.pdf

[GUASTAVINO-03]

GUASTAVINO, C. (2003).

Etude sémantique et acoustique de la perception des basses fréquences dans l'environnement sonore urbain.

Thèse de doctorat, Université Paris 6, LAM 2003.

[GUSKY-97]

GUSKY, R.. (1997)

Psychological Methods for evaluating sound quality and assessing acoustic information.

Acustica . Acta Acustica, 83, 765-774.

[GUSKY-97]

GUSKY, R.. (1997).

Conceptual, methodological and dose response problems related to annoyance and disturbance.

Inter-Noise 1997, 2, 1077-1082.

[GUSKY-99]

GUSKY, R.. (1999).

Personal and social variables as co-determinants of noise annoyance.

Noise and Health, 1, 45-46.

[GYGI-01]

GYGI, B. (2001).

Factors in the identification of environmental sounds.

PhD thesis, Indiana University, Juillet 2001.

http://matrix.aklab.psych.ubc.ca/uploads/Michael_Sounds_JASA.pdf

<http://www.indiana.edu/~k300bg/dissall.pdf>

[HANDEL-89]

HANDEL, S. (1989).

Listening : An introduction to the perception of auditory events.

Cambridge, Mass., MIT Press (1989)

[HELMHOLTZ-85]

HELMHOLTZ, H. (1885).

On the sensation of tone.

Dover.

[JAMES-90]

JAMES, W. (1890).

The principles of psychology Vol. 2.

Dover.

[KALLMANN-04]

KALLMANN, H., WESTERKAMP, H., WOOG, A.P. (2004) .

World Soundscape Project.

L'Encyclopédie canadienne. Fondation Historica du Canada. 2004.

[KLATT-89]

KLATT, D.H. (1989).

Review of selected models of speech perception.

Lexical representation and process (ed. W. D. Marslen-Wilson), Cambridge, MA.

MIT Press, Chapitre 6, pages 169 à 226.

[KRAUSE-93]

KRAUSE, B.L. (1993).

The niche hypothesis: a hidden symphony of animal sound, the origins of musical expression and the health of habitats.

The Explorer Journal, Hiver 1993, 156-160.

[KUWANO-02]

KUWANO, S. (1992).

Memory of the loudness of sounds and its relation to overall impression.

Forum Acusticum, Seville, septembre 2002 .

[LAKATOS-97]

LAKATOS, S., MCADAMS, S., CAUSSE, R. (1997).

The representation of auditory source characteristics: simple geometric form.

Perception and psychophysics, 59, 1180-1190.

[LASS-82]

LASS, N. J., EASTHAM, S.K, PARRISH, W.C., SCHERBICK, K.A., RALPH, D.M. (1982).

Listener's identification of environmental sounds.

Perceptual and Motor Skills, 57, 995-998.

[LAVANDIER-02]

LAVANDIER, C., BARBOT, B. (2002).

Influence of the temporal scale on the relevance of acoustic parameters selected to characterize urban sound environment.

5^{ème} congrès Euro Noise, Naples, Italie, mai 2003.

[LAVANDIER et al – 03]

LAVANDIER, C. (2003).

Qualité des ambiances sonores liées aux usages des établissements d'enseignement.

Rapport de recherche financé par le PUCA, Ministère de l'Équipement, des Transports et du Logement : Lettre F01107. 2003.

[LEOBON-95]

LEOBON, A. (1995).

La qualification des ambiances sonores urbaines.

Natures et société, 3(1), 26-41.

[LOI BRUIT-92]

Loi n° 92-144 du 31 décembre 1992 relative à la lutte contre le bruit.

Journal officiel du 1^{er} janvier 1993.

<http://www.legifrance.gouv.fr/WAspad/Ajour?nor=ENVX9200186L&num=92-1444&ind=6&laPage=1&demande=ajour>

[MACPHERSON-95]

MACPHERSON, E.A. (1995).

A review of auditory perceptual theories and the prospects for an ecological account.

Preliminary exam paper. Department of Psychology, University of Wisconsin-Madison.

http://www.khri.med.umich.edu/800x600/research/middlebrooks_lab/ewan/research.html

[MAFFIOLO-99]

MAFFIOLO, V. (1999) .

Qualité sonore de l'environnement urbain.

Thèse de doctorat, Université du Maine, Le Mans 1999.

[MCADAMS-94]

MCADAMS, S., BIGAND, E. (1994).

Introduction à la cognition auditive.

Penser les sons, PUF, pp 1-9.

[MCADAMS-97]

MCADAMS, S. (1997)

L'organisation perceptive de l'environnement sonore.

Rencontres IPSEN en ORL, 1997, Serveur© IRCAM.

<http://mediatheque.ircam.fr/articles/textes/McAdams97b/>

[MERLEAU-PONTY-45]

MERLEAU-PONTY, M. (1945).

Phénoménologie de la Perception.

Gallimard 1945.

[MEUNIER-02]

MEUNIER, S., MARCHIONI, A. (2002).

Loudness of sounds with temporal variable intensity.

Proceeding Forum Acusticum, Seville, 16-20.

[MEUNIER-04]

MEUNIER, S. (2004).

Psychoacoustique musicale et psychoacoustique appliquée.

Cours du Master d'acoustique de l'Université d'Aix Marseille II. Année 2004-2005.

[MOLES-95]

MOLES, A. (1995).

Les sciences de l'imprécis.

Editions du Seuil, mars 1995.

[MOLES-81]

MOLES, A. (1981).

L'image sonore du monde environnant. Phonographies et Paysages sonores.

Proc. Colloque : paysage sonore urbain, 265-291, Paris, Plan construction.

[MOORE-97]

MOORE, B.C.J. (1997).

An Introduction to the Psychology of Hearing, 4th Ed.

Academic, San Diego (1997).

[MZALI-00]

MZALI, M. (2000).

Perception de l'ambiance sonore et évaluation du confort acoustique dans les trains

Thèse de Doctorat de l'Université de Paris VI.

[NAMBA-94]

NAMBA, S. (1994).

Noise – quantity and quality.

Proceeding Inter-Noise 94, 171-218.

[NAMBA-04]

NAMBA, S., KUWANO, S. (2004).

Environmental Acoustics: psychological assessment of noise.

Ecological Psychoacoustics, J. Neuhoff (Ed), pp. 175-190.

[NELKEN-99]

NELKEN, I., ROTMAN, Y., YOSEF, O.B. (1999).

Responses of auditory-cortex neurons to structural features of natural sounds.

Nature, 397, 154-157, 14 janvier 1999.

[NEUHOFF-98]

NEUHOFF, J.G. (1998).

A perceptual bias for rising tones.

Nature (Londres) 395, p 123.

[NEUHOFF-04]

NEUHOFF, J.G. (2004).

Localization and motion.

Ecological psychoacoustics, J. Neuhoff (Ed), pp. 87-111

[NORME-96]

NFS 30-010 Décembre 1996 (1996).

Caractérisation et mesurage des bruits de l'environnement.

AFNOR

[NOTBOHM-04]

NOTBOHM, G., VASTFJALL, D., GARTNER, C., SCHWARZE, S. (2004).

Subjective evaluation of the sound quality of urban traffic noise situation.

Proc. Colloque ICA 2004, Kyoto, avril 2004.

[ORDONNANCE-04]

Ministère de l'écologie et du développement durable.

Ordonnance n° 2004-1199 du 12 novembre 2004 prise pour la transposition de la directive 2002/49/CE du Parlement européen et du Conseil du 25 juin 2002 relative à l'évaluation et à la gestion du bruit dans l'environnement.

Journal officiel de la république française, 14 novembre 2004.

<http://www.legifrance.gouv.fr/WAspad/UnTexteDeJorf?numjo=DEVX0400235R>

[PLOMB-02]

PLOMB, R.. (2002).

The Intelligent Ear: on the nature of sound perception.

Lawrence Erlbaum Associates, Inc. pp. 104, 2002.

[POPPER-97]

POPPER, A.N., FAY, R.R.. (1997).

Evolution of the Ear and Hearing: Issues and Questions.

Brain, Behaviour and Evolution, 1997, 50, 213-221.

[POPPER-00]

POPPER., A.N., FAY, R.R.. (2000).

Evolution of hearing in vertebrates: the inner ears and processing.

Hearing Research 149 (2000), 1-10.

[QUEHL-01]

QUEHL, J. (2001).

Comfort studies on aircraft interior sound and vibration.

Thèse de doctorat de l'université de Oldenburg, 2001.

<http://docserver.bis.uni->

[oldenburg.de/publikationen/dissertation/2001/quecom01/quecom01.html](http://docserver.bis.uni-oldenburg.de/publikationen/dissertation/2001/quecom01/quecom01.html)

[RADAU-67]

RADAU, R.. (1867).

L'acoustique et les phénomènes du son.

Hachette, 1867 – 1870. p. 6.

[RAIMBAULT-03]

RAIMBAULT, M., LAVANDIER, C., BERENGIER, M. (2003).

Ambient sound assessment of urban environments: field studies in two French cities.

Applied Acoustics, 64 (12),1241-1256.

[RAIMBAULT-01]

RAIMBAULT, M., DUBOIS, D., BERENGIER, M. (2003).

Common factors in the identification of urban soundscapes (pilot studies in two French cities: Lyon and Nantes).

Proc. ICA CD-ROM, 2001.

[REPP-87]

REPP, B. H. (1987).

The sound of two hand clapping: An exploratory study.

The Journal of the Acoustical Society of America, 81, 1100-1110.

[RISSET-95]

RISSET, J.C. (1995).

Support de cours de psychoacoustique pour le DEA ATIAM.

[ROUSSARIE-99]

ROUSSARIE, V. (1999).

Analyse perceptive de structures vibrantes simulées par model physique.

Thèse de l'université du Maine, décembre 1999.

[ROZEC-99]

ROZEC, V. (1999).

La perception et l'évaluation du confort sonore dans l'ex quartier tranquille (désormais « quartier vert ») de la butte aux Cailles.

Thèse de doctorat, 1999.

[RUSSELL-80]

RUSSELL, J.A. (1980).

The circumplex model of affect.

Journal of Personality and Social Psychology, 39, 1161-1178.

[SCHAEFFER-66]

SCHAEFFER, P. (1966).

Traité des objets musicaux.

Seuil, 2002.

[SCHAFER-67]

SCHAFER, M. (1967).

Ear Cleaning.

BMI. Canada.

[SCHAFER-77]

SCHAFER, M. (1977 puis 1994).

The Soundscape. Our Sonic Environments and the Tuning of the World".

Rochester VT, p 29, (1977).

[SHEPARD-81]

SHEPARD, R.N. (1981).

Psychological relations and psychological scales: on the status of « direct » psychological measurement.

Journal of mathematical Psychology, 24, 21-57.

[SMOLDERS-79]

SMOLDERS, J.W., AERTSEN, A.M., JOHANNESMA, P.I. (1979).

Neural presentation of the acoustic biotope. A comparison of the response of auditory neurons to tonal and natural stimuli in the cat.

Biol. Cybern. Nov. 1979, 35(1), 11-20.

[SOLOMON-58]

SOLOMON, L.N. (1958)

Semantic approach to the perception of complex sounds

The Journal of the Acoustical Society of America, 30, 421-425, Mai 1958.

[SOUTHWORK-69]

SOUTHWORK, M. (1969).

The sonic environment of cities.

Environment and Behavior, 1(1), 49-70.

[STEVENS-98]

STEVENS, K.,N. (1998)

Acoustic phonetics

The MIT Press

[SUSINI-98]

SUSINI, P., MISDARIIS, N., WINSBERG, S., McADAMS, S. (1998).

Caractérisation perceptive de bruits.

Acoustique et Techniques, 13, 11-15.

[SUSINI-00]

SUSINI, P., McADAMS, S. (2000).

Effet de récence dans une tâche de jugement de la sonie.

5^{ème} Congrès français d'acoustique, Lausanne, Suisse.

[SUSINI-02]

SUSINI, P., McADAMS, S. (2002).

Influence of sound-directed attentional focus on average loudness rating.

Proceeding Internoise 2002, Seville 2002.

[SUSINI-05]

SUSINI, P. (2005)

Approches de la qualité sonore.

Cours de DESS acoustique architecturale et urbaine, Université Paris VI, 2005.

http://www.lmp.jussieu.fr/lmp/pages/actens/AcouArch/Info_etudiants2004_05/Qualite_sonore.pdf

[TEGHTSOONIAN-00]

TEGHTSOONIAN, R., TEGHTSOONIAN, M., CANEVET, G. (2000).

The perception of waning signals: Decruitment in loudness and perceived size.

Perception and Psychophysics, 62, 637-646.

[TODD-01]

TODD, N. (2001).

Evidence for a behavioral significance of saccular acoustic sensitivity in humans.

The Journal of the Acoustical Society of America, 110, 380-390, Juillet 2001.

[TOYODA-04]

TOYODA, Y., HUANG, J., DING, S., LIU, YONG. (2004).

Environmental Sound recognition by Multilayered Networks.

Proceedings of the Fourth International Conference on Computer and Information Technics, 00, 123-127, 2004.

[TRUAX-84]

TRUAX, B. (1984).

Acoustic communication.

New Jersey : Ablex publishing, 1984.

[TUKEY-77]

TUKEY, J.W. (1977).

Exploratory data analysis.

EDA, Reading, Massachusetts, (Addison-Wesley).

[VANDERVEER-79]

VANDERVEER, N.J. (1979).

Ecological acoustics : human perception of environmental sounds.

Diss. Abstr. Internat. , 40, 4543 (University Microfilms N° 8004002).

[VÄSTFJÄLL-02]

VÄSTFJÄLL, D. (2002).

Emotion in product sound design.

Proceedings des Journées du design sonore, Paris, 20-21 mars 2002.

[VÄSTFJÄLL-02]

VÄSTFJÄLL, D. (2002).

Influences of current mood and noise sensitivity on judgments of noise annoyance.

The journal of Psychology, 2002, 136(4), 357-370.

[VIOLLON-00]

VIOLLON, S. (2000).

Influence des informations visuelles sur la caractérisation de la qualité acoustique de l'environnement sonore urbain.

Thèse de doctorat de l'université de Cergy-Pontoise, 2000.

[VOGEL-99]

VOGEL, C. (1999)

Etude sémiotique et acoustique de l'identification des signaux sonores d'avertissement en contexte urbain.

Thèse de doctorat de l'université de Paris, 1999.

[VOSS-77]

VOSS, R.F., CLARKE, J. (1977).

1/f, noise in music and speech.

Nature, 258, 317-318 (1977).

[WINSBERG-89]

WINSBERG, S., CARROLL, D. (1989).

A quasi non-metric method for multimodal scaling via an extended Euclidian model. Psychometrika, 54, 217-229.

[WOLOSZYN-98]

WOLOSZYN, P., SIRET, D. (1998).

Du complexe au simplexe, le modèle des objets ambiants.

Cahiers de la Recherche Architecturale, Paris, Octobre 1998.

[WRIGHTSON-99]

WRIGHTSON, K. (1999).

An introduction to acoustic ecology.

Journal of Electroacoustic Music, 12, Mars 1999.

[ZHANG-98]

ZHANG, T., JAY KUO, C.C. (1998).

Hierarchical System for Content-based Audio Classification and Retrieval.

International Symposium on Voice, Video and data Communication, Boston, MA, Novembre 1-6, 1998.

[ZILS-04]

ZILS, A. (2004).

Extraction de descripteurs musicaux: une approche évolutionniste.

Thèse de doctorat de l'Université de Paris 6, 2004.

<http://www.csl.sony.fr/General/People/StaffPage.php?username=aymeric>

[ZILS-04 bis]

ZILS, A., PACHET, F. (2004).

Automatic Extraction of Music Descriptors from Acoustic Signals using EDS.

Proceedings of the 116th AES Convention, May 2004.

<http://www.csl.sony.fr/General/People/StaffPage.php?username=aymeric>

[ZWICKER-81]

ZWICKER, E., FELDTKELLER, R. (1981).

Psychoacoustique.

Masson, Paris, 1981.

[ZWICKER-90]

ZWICKER, E., FASTL, H. (1990).

Psychoacoustics, facts and models.

Springer Verlag, 1990.

[ZWICKER-99]

ZWICKER, E., FASTL, H. (1999).

Psychoacoustics, facts and models, Updated Edition.

Springer Verlag, 1999.

ANNEXES

ANNEXE 1 : consigne des tests de présence, prégnance et proximité sur les séquences de 15 s.

1. Présentation :

On va vous demander de juger trois qualificatifs concernant des sons urbains à travers des séquences sonores d'une quinzaine de secondes chacune.

- **La prégnance** : c'est l'importance, dans les sens où une source sonore s'est démarquée des autres, non pas à cause de sa durée, mais sur le fait qu'elle a attiré votre attention quand elle est apparue.
- **La présence** : c'est l'occurrence temporelle de la source. Cela revient à se demander si la source apparaît souvent durant la séquence.
- **La proximité** : c'est votre distance par rapport à la source (attention une petite note indique une source éloignée).

2. Déroulement du test :

Ainsi, pour chaque séquence, vous préciserez quelles sont la ou les sources qui marquent cette séquence, et pour chacune d'elle, quelle est sa prégnance, sa présence et sa proximité dans son environnement sur une échelle de « 1 » à « 10 ».

Vous pouvez réécouter les séquences sonores autant de fois que vous le voulez.

Durant le test, imaginez-vous un déplacement pédestre urbain tel que vous l'avez déjà vécu : vous avez un rendez-vous mais vous n'êtes pas pressé.

3. Entraînement :

Afin de vous faire une échelle et de vous familiariser au test, trois exemples vous sont présentés en entraînement.

Une source sonore ayant une forte prégnance,
Une source ayant une forte présence
Une source peu prégnante et peu présente.

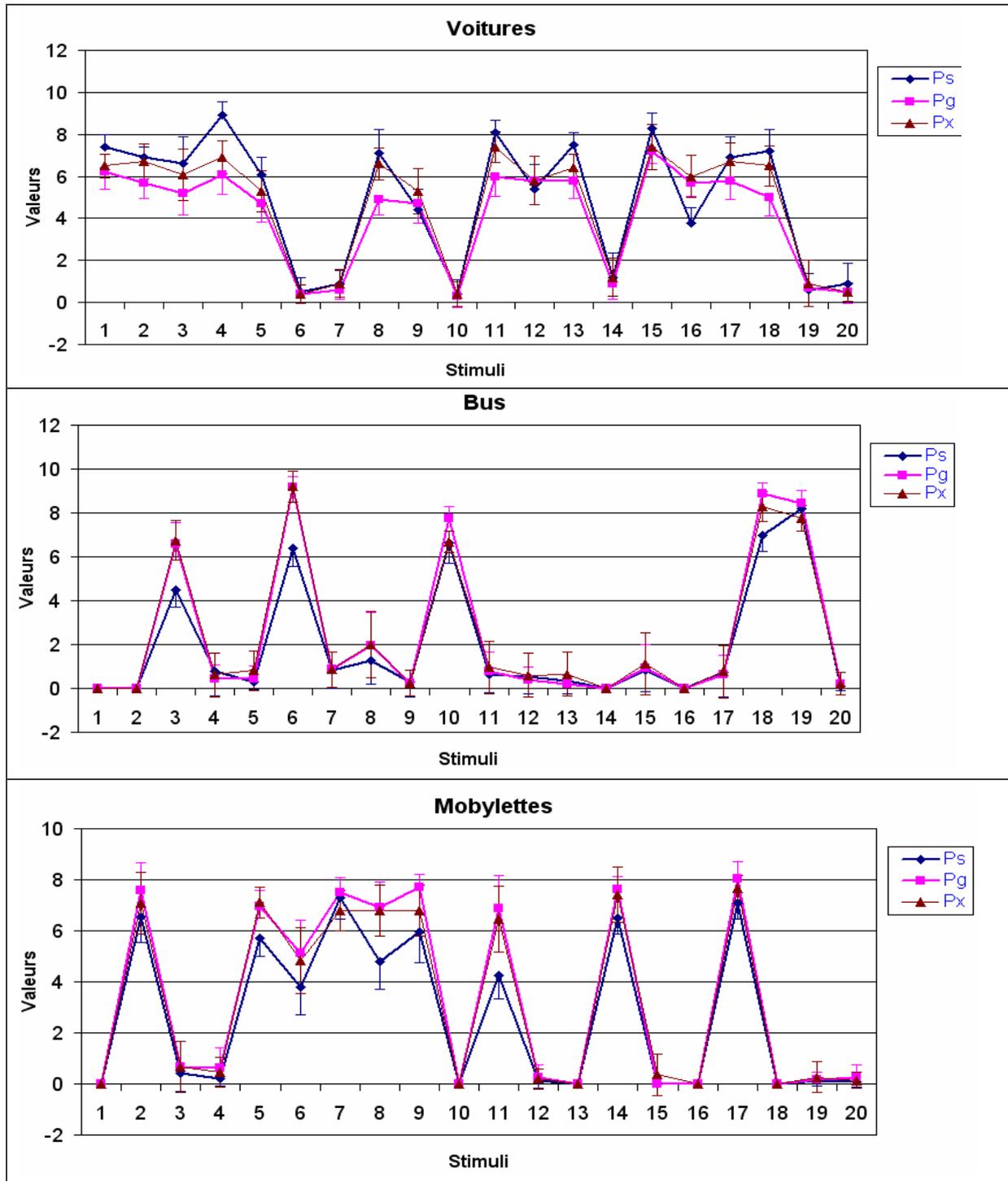
Attention, l'échelle va de « 1 » à « 10 ».

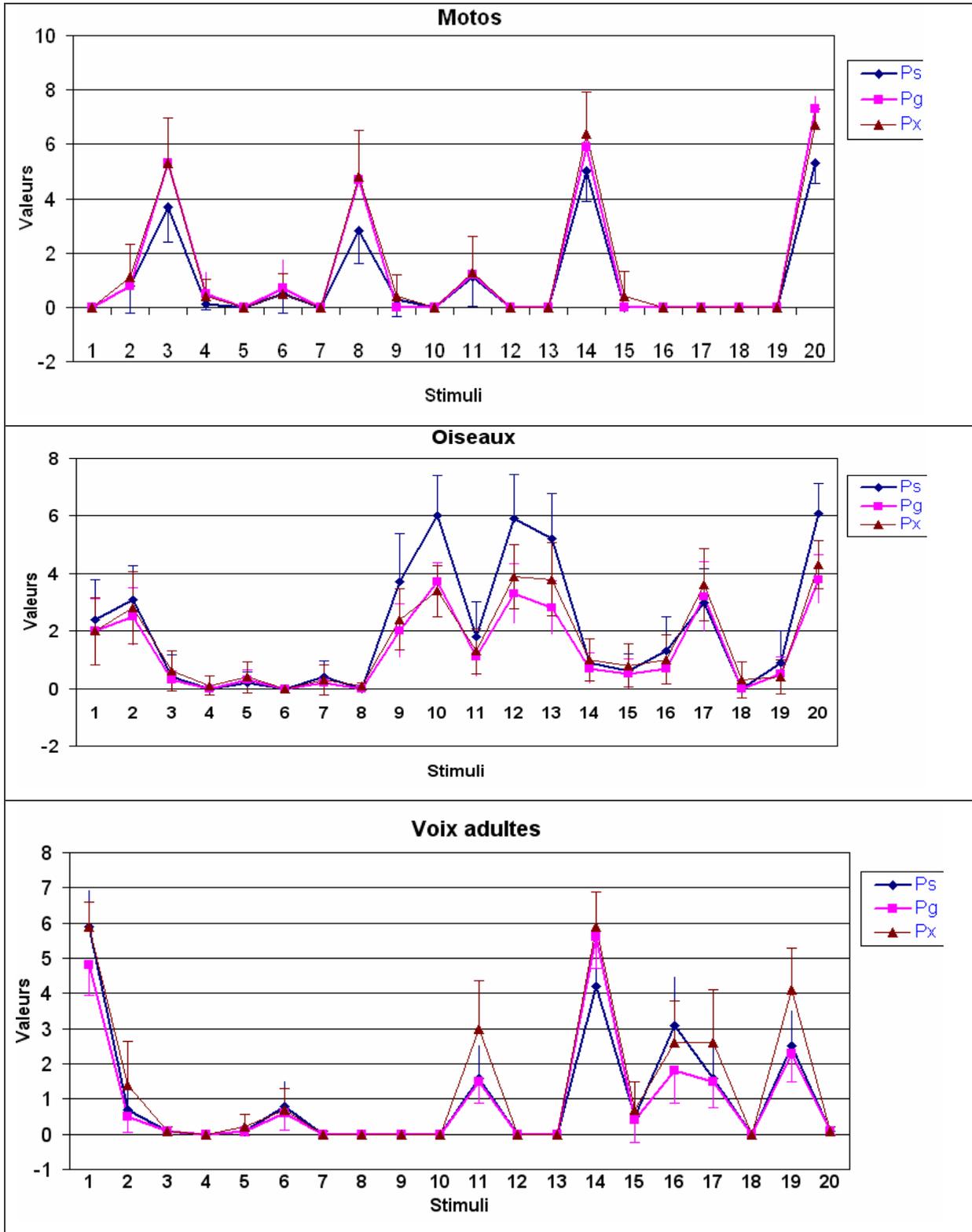
Surtout, limitez vous aux sources qui vous ont marqués.

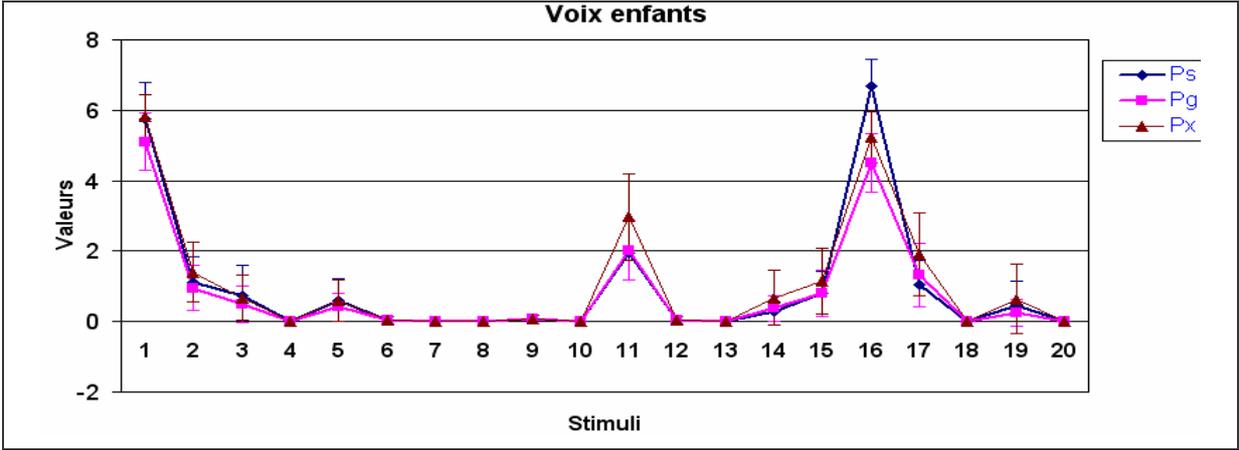
PS : dès qu'une source est suffisamment détectée, elle doit automatiquement obtenir une note pour les trois qualificatifs.

ANNEXE 2 : représentation graphique des résultats du test 1.

Présence (Ps), Prégnance (Pg) et Proximité (Px) des vingt stimuli.



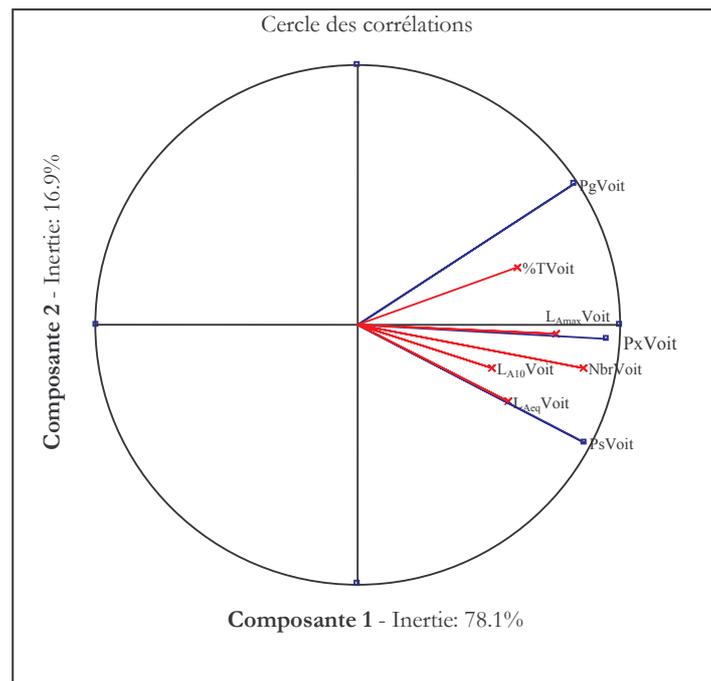




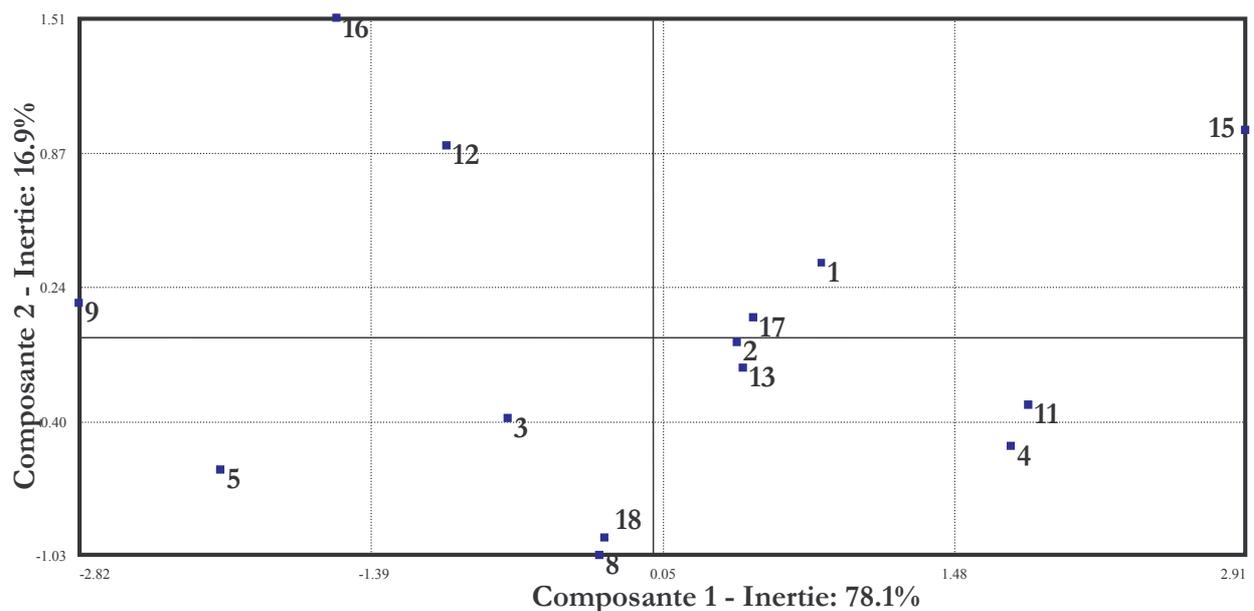
Pour des raisons de visibilité, les écarts types ne sont parfois représentés que sur un seul coté.

ANNEXE 3a : Représentations graphiques de l'ACP faite sur les stimuli contenant des voitures.

Numéro de stimulus	1	2	3	4	5	8	9	11	12	13	15	16	17	18
Nombre de voitures	4	2	1	5	1	3	1	4	1	3	4	1	2	2

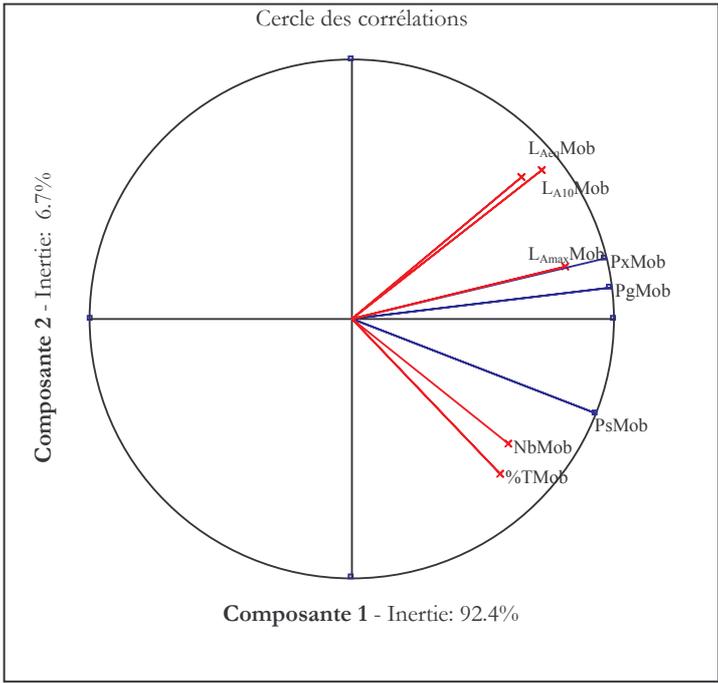


Plan factoriel 1-2

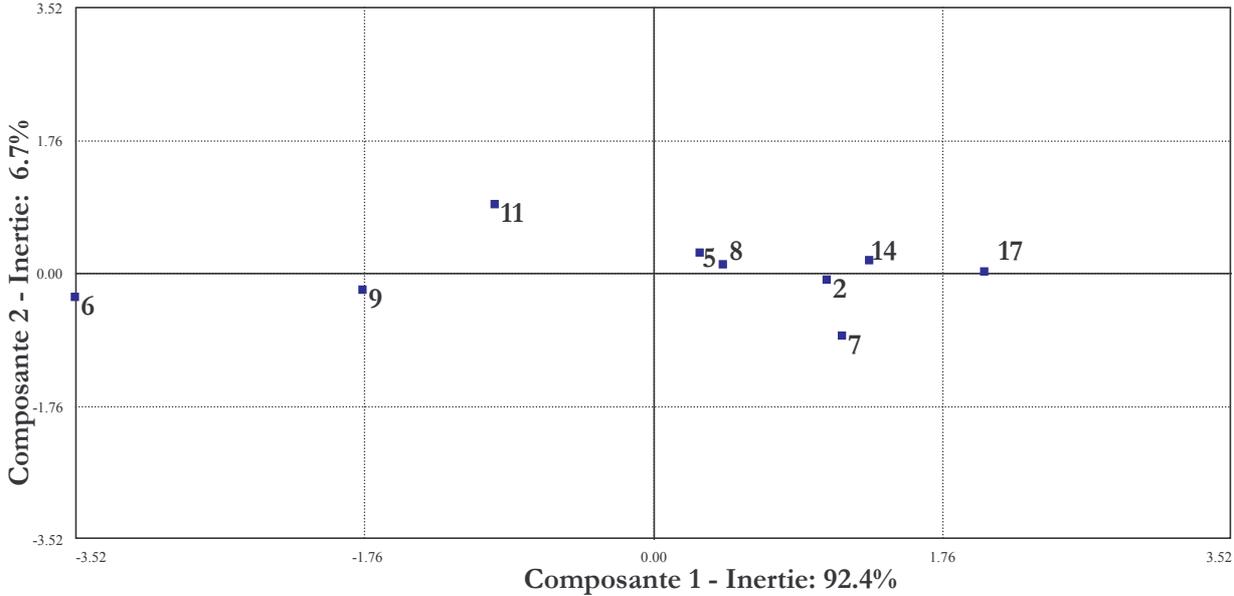


ANNEXE 3b : Représentations graphiques de l'ACP faite sur les stimuli contenant des cyclomoteurs.

Numéro de stimulus	2	5	6	7	8	9	11	14	17
Nombre de cyclomoteurs	2	1	1	2	1	1	1	1	2

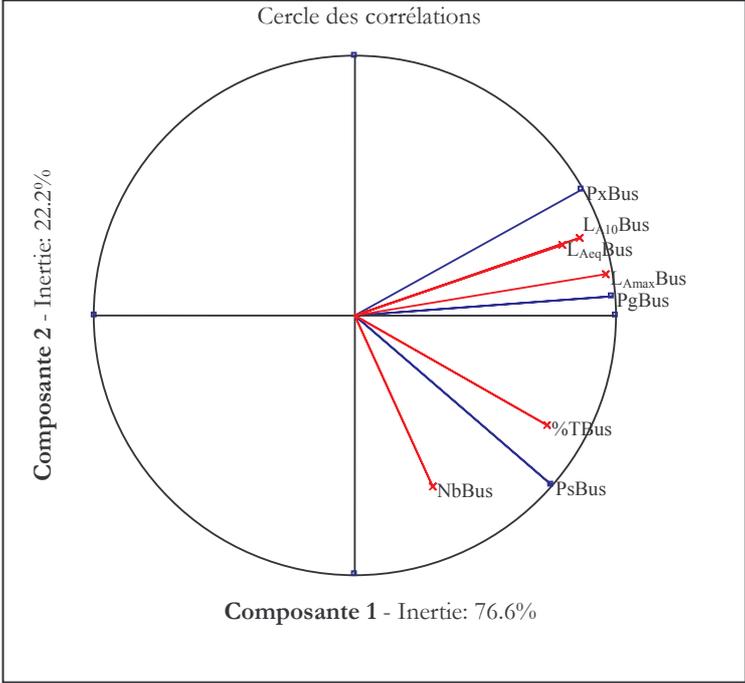


Plan Factoriel 1-2

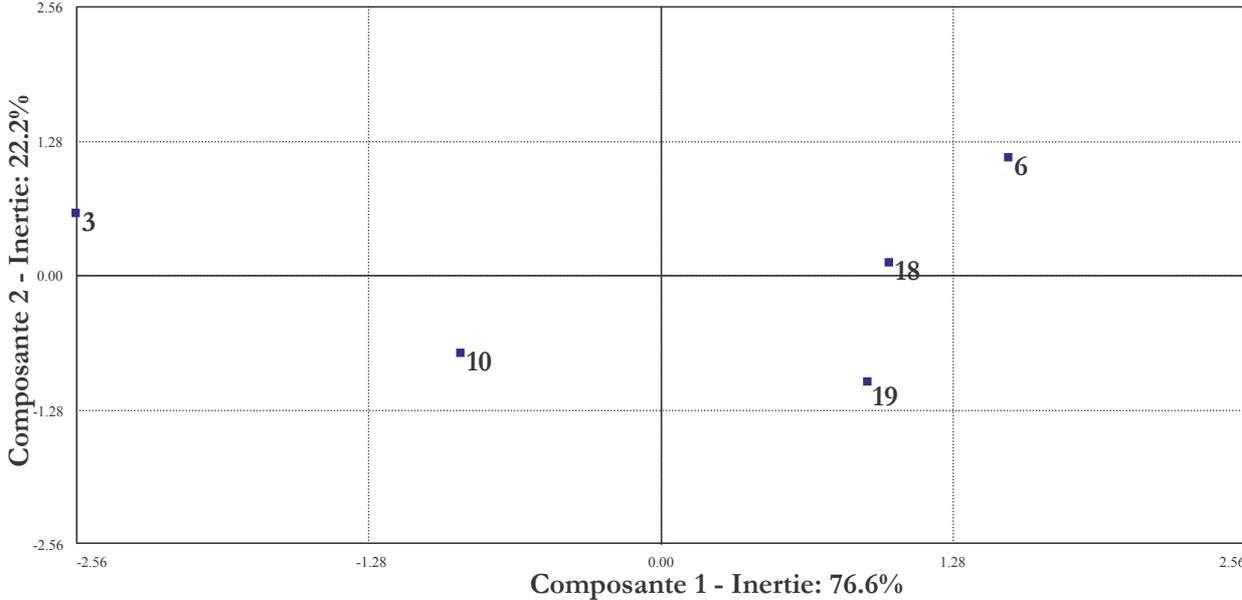


ANNEXE 3c : Représentations graphiques de l'ACP faite sur les stimuli contenant des bus.

Numéro de stimulus	3	6	10	18	19
Nombre de bus	1	1	1	1	2

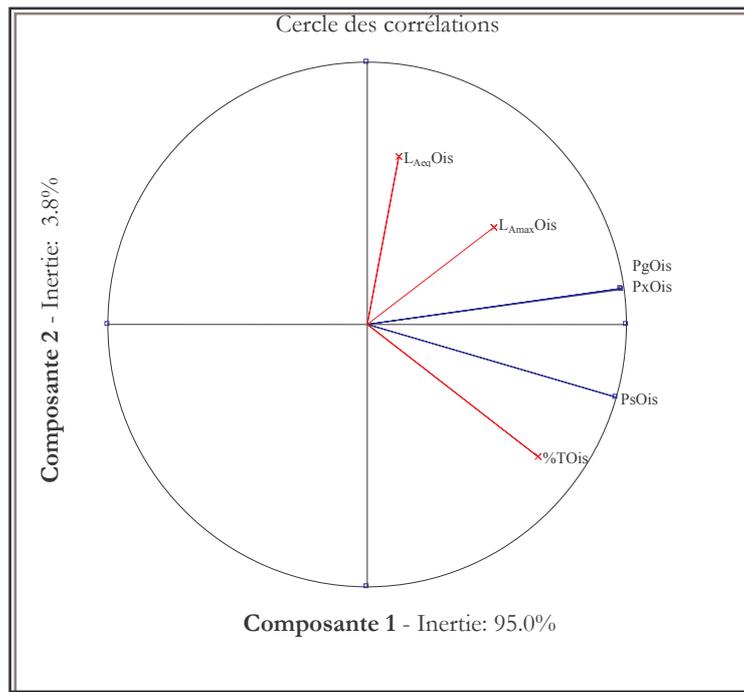


Plan Factoriel 1-2

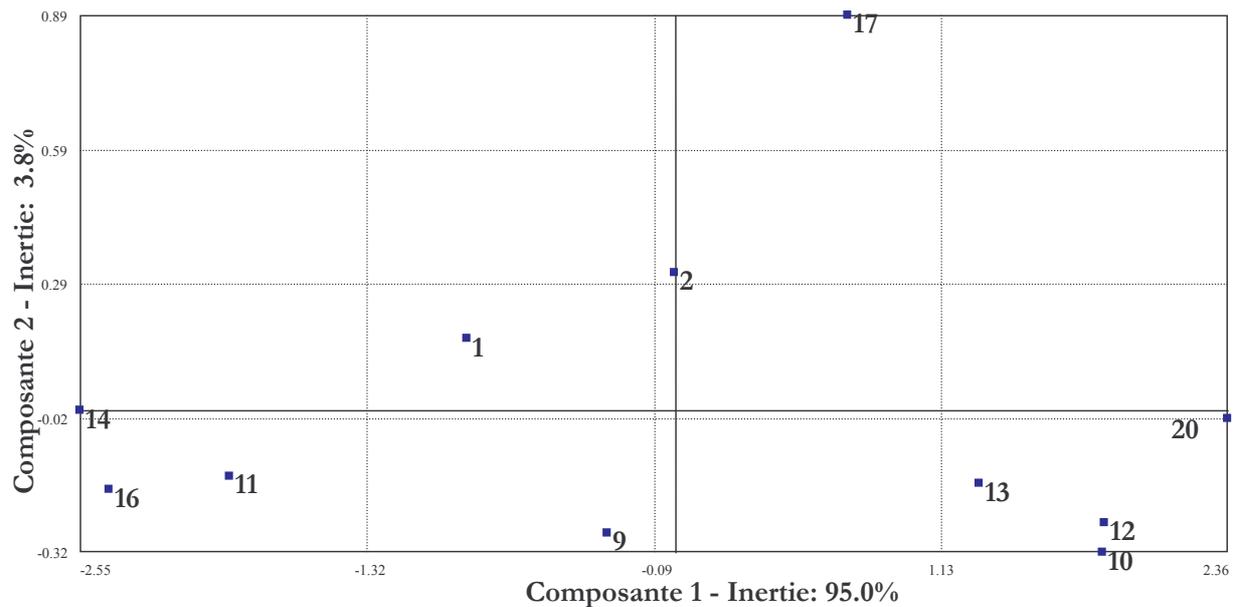


ANNEXE 3d : Représentations graphiques de l'ACP faite sur les stimuli contenant des oiseaux.

Numéro de stimulus	1	2	9	10	11	12	13	14	16	17	20
%T	44	25	85	73	15	60	83	28	35	28	76

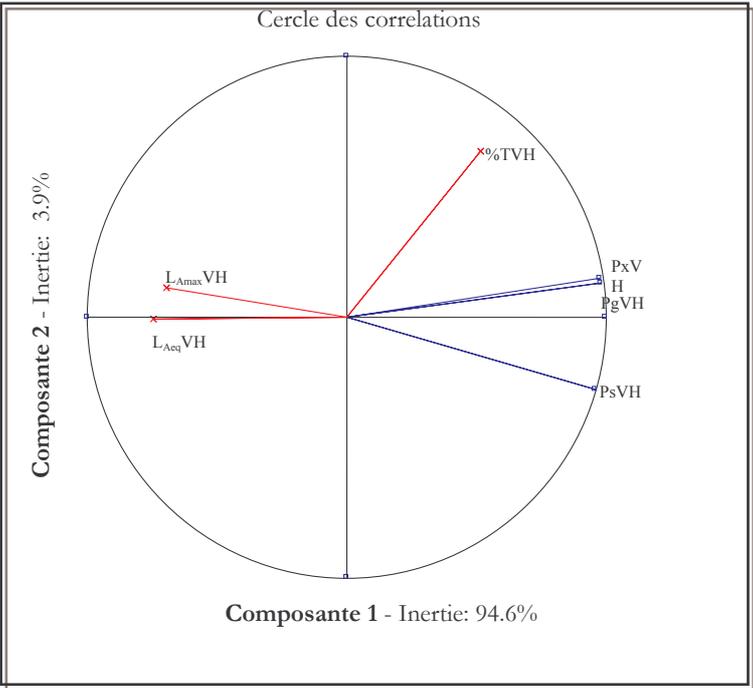


Plan Factoriel 1-2

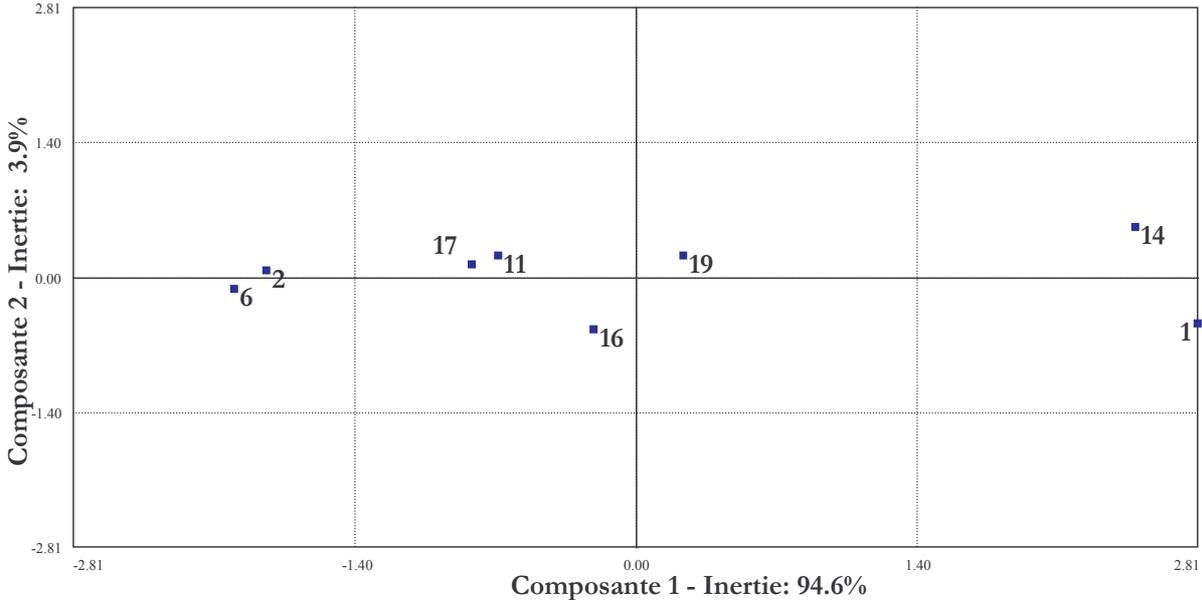


ANNEXE 3e : Représentations graphiques de l'ACP faite sur les stimuli contenant des voix d'adultes.

Numéro de stimulus	1	2	6	11	14	16	17	19
%T	7	9	4	5	28	9	15	15

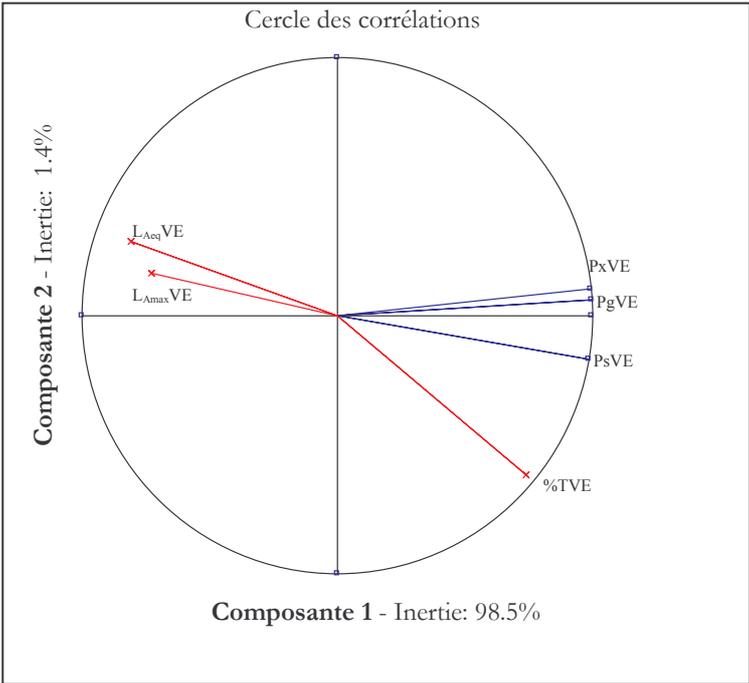


Plan Factoriel 1-2

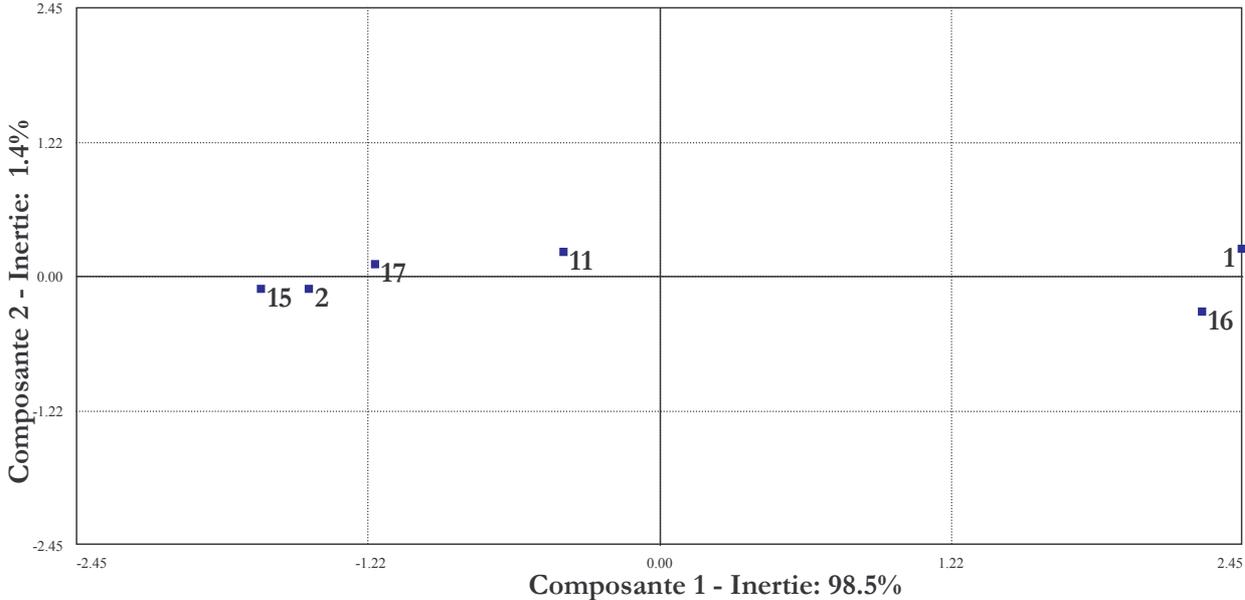


ANNEXE 3f : Représentations graphiques de l'ACP faite sur les stimuli contenant des voix d'enfants.

Numéro de stimulus	1	2	11	15	16	17
%T	16	7	7	5	55	5

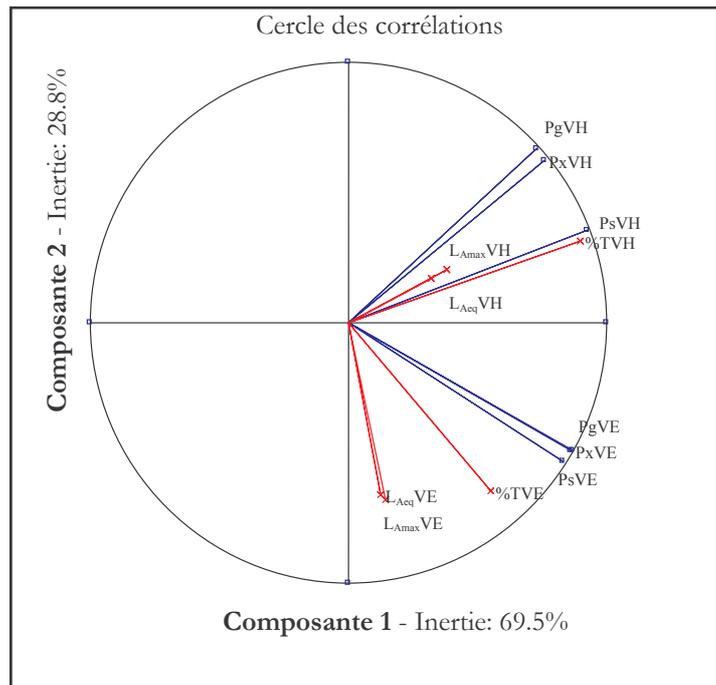


Plan Factoriel 1-2

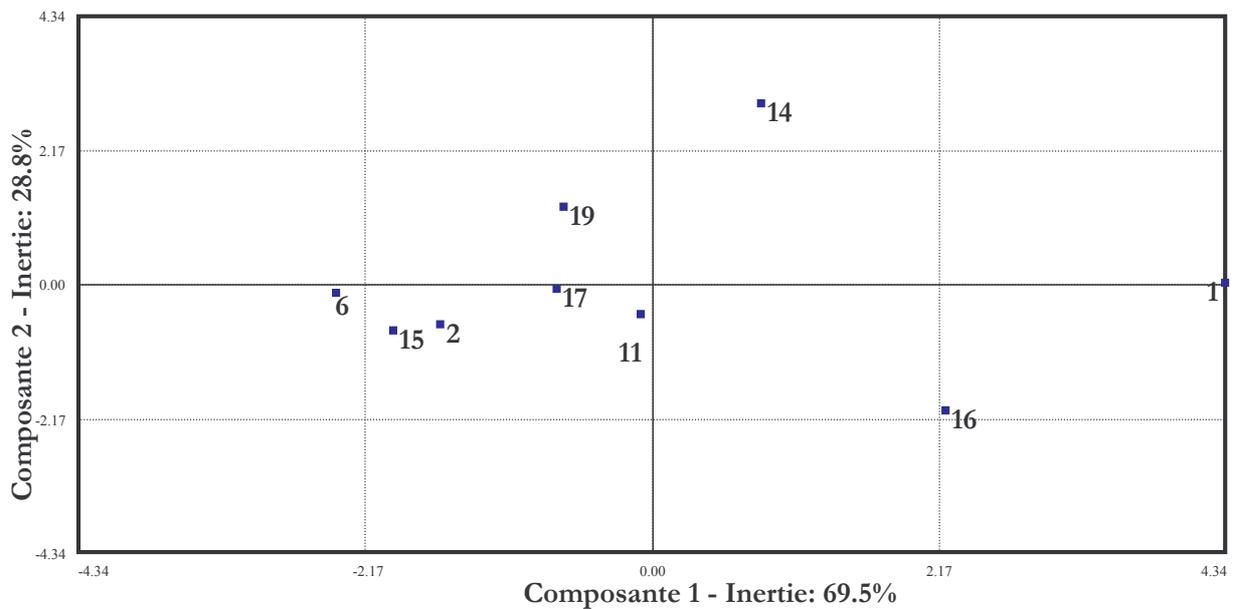


ANNEXE 3f : Représentations graphiques de l'ACP faite sur les stimuli contenant des voix.

Numéro de stimulus	1	2	6	11	14	15	16	17	19

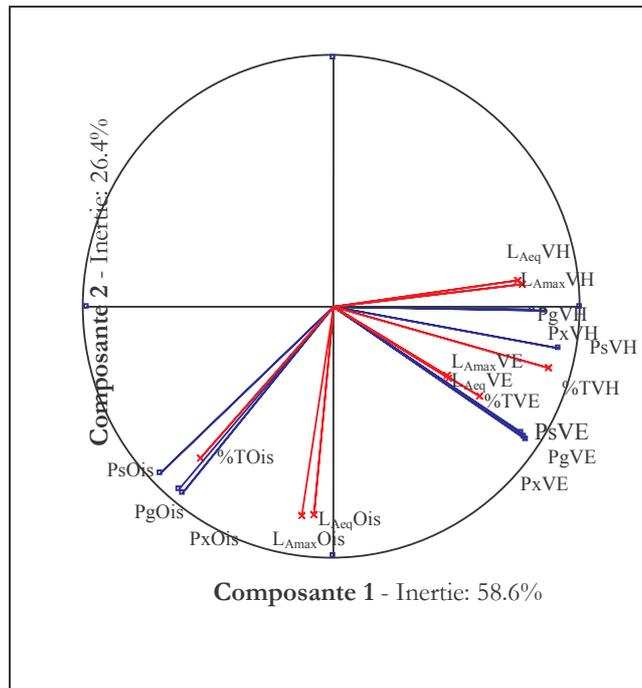


Plan Factoriel 1-2

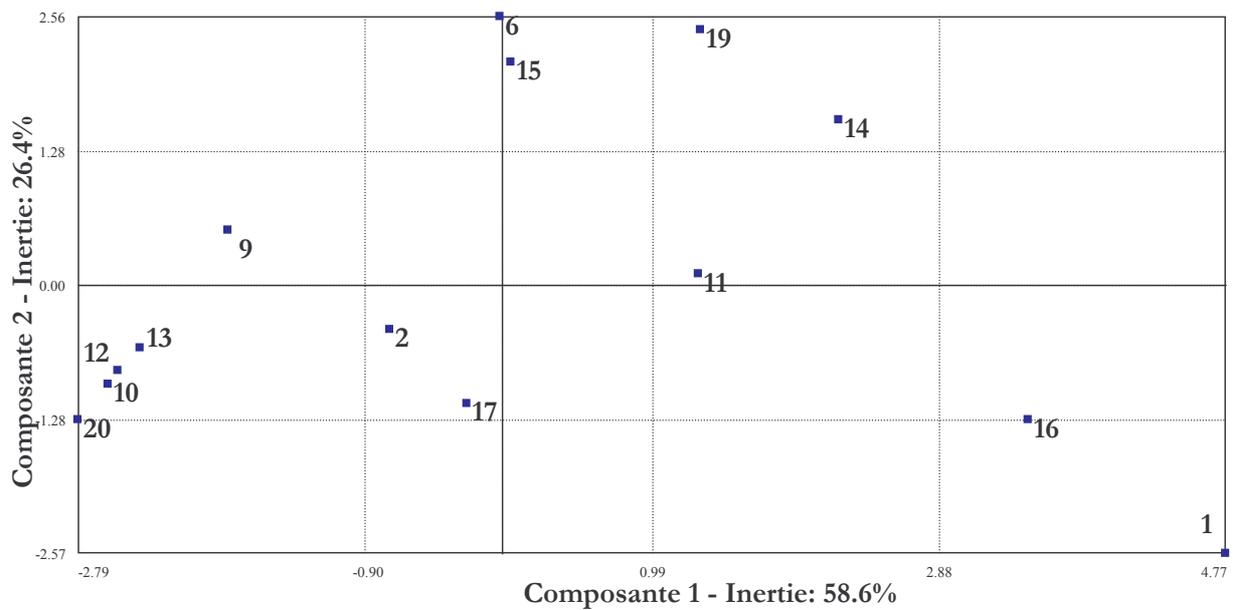


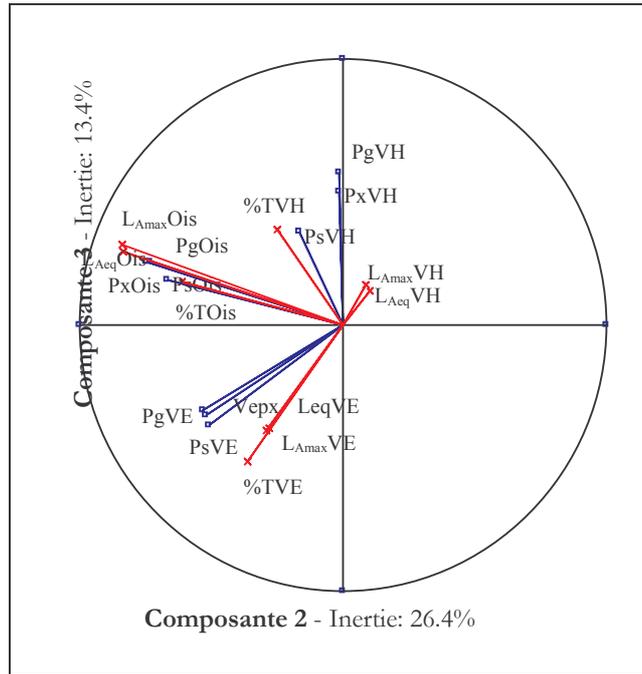
ANNEXE 3g : Représentations graphiques de l'ACP faite sur les stimuli contenant des voix et des oiseaux.

Numéro de stimulus	1	2	6	9	10	11	12	13	14	15	16	17	19	20
--------------------	---	---	---	---	----	----	----	----	----	----	----	----	----	----

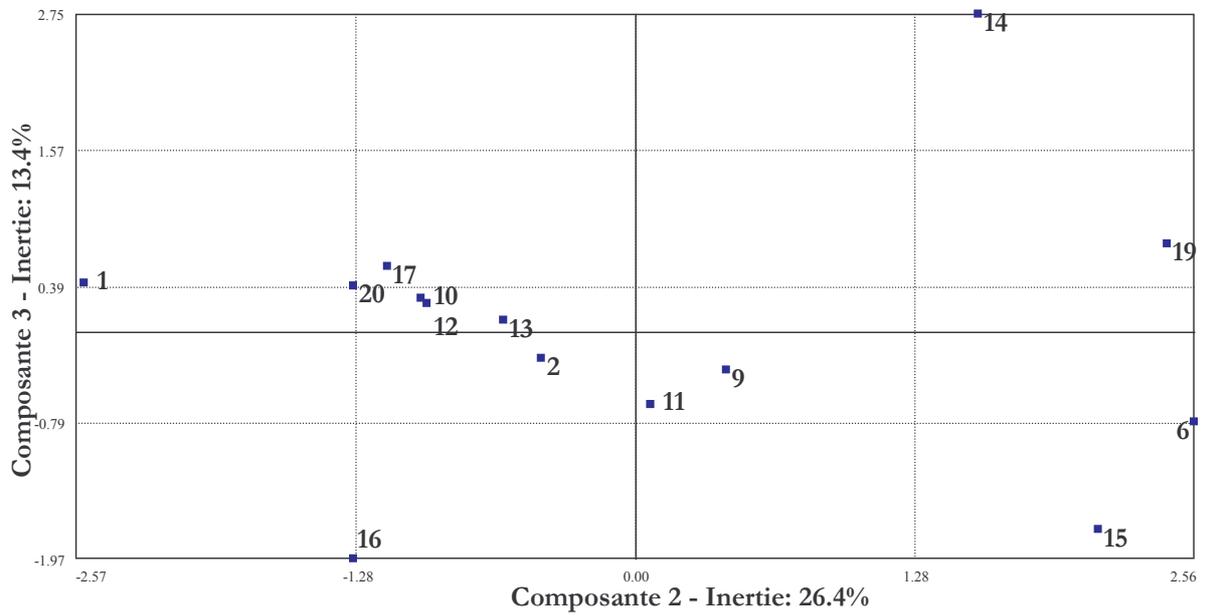


Plan Factoriel 1-2



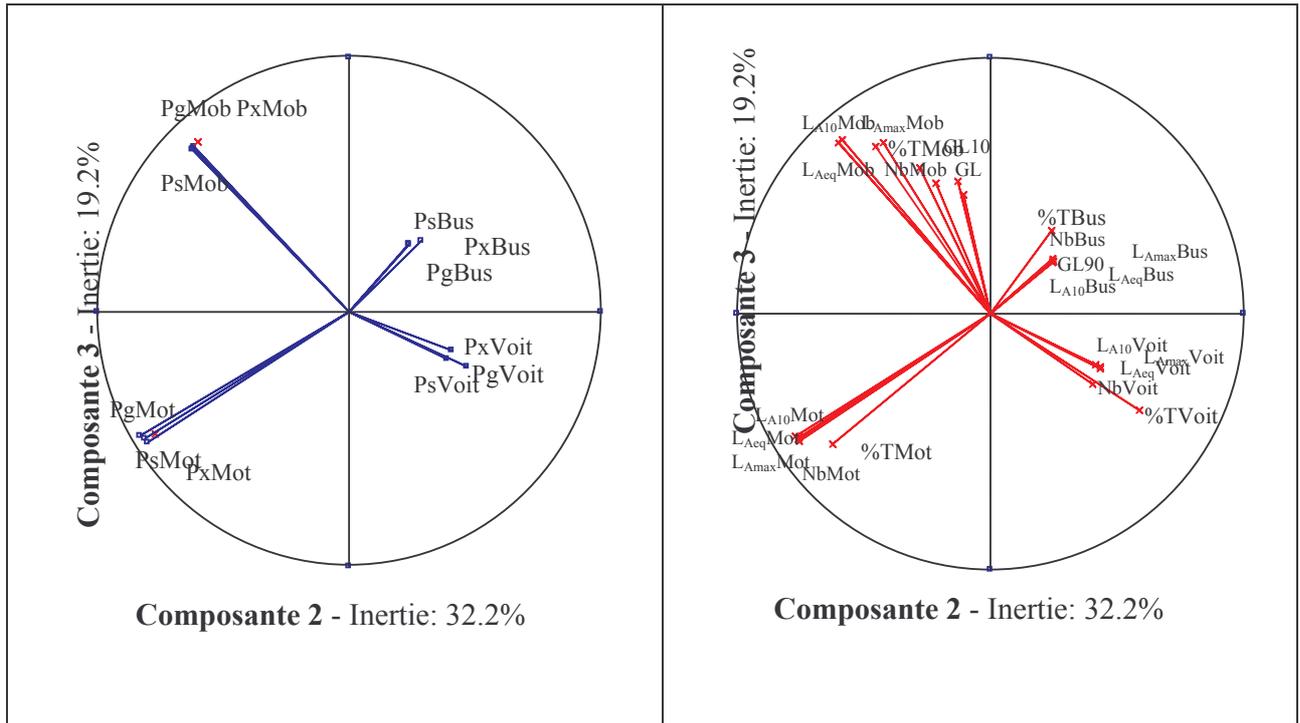


Plan Factoriel 2-3

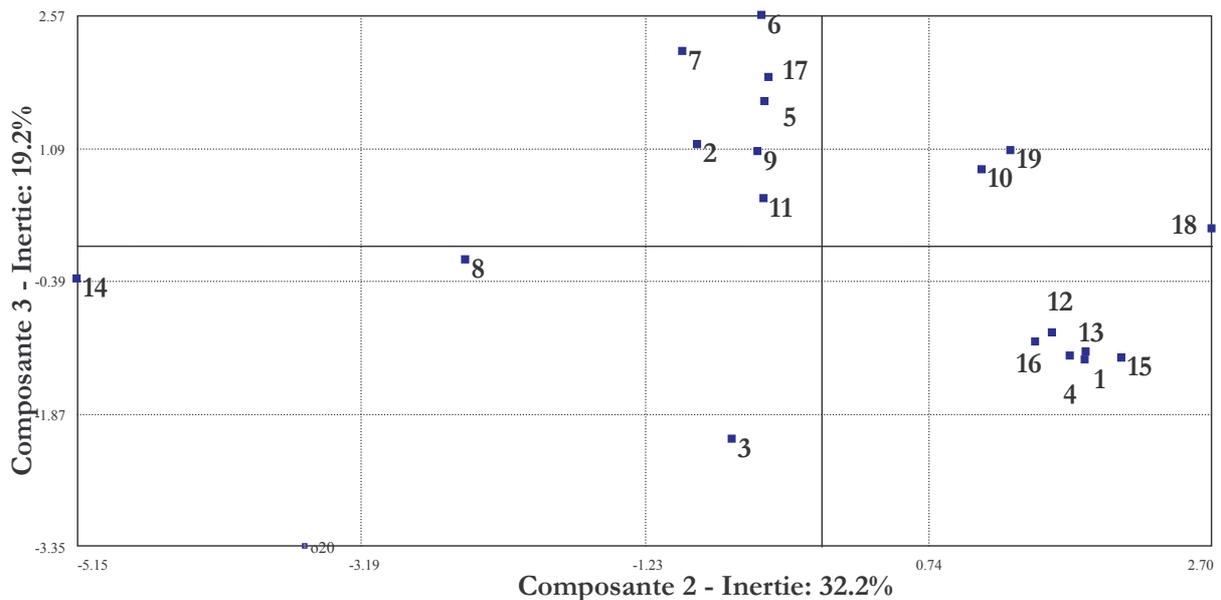


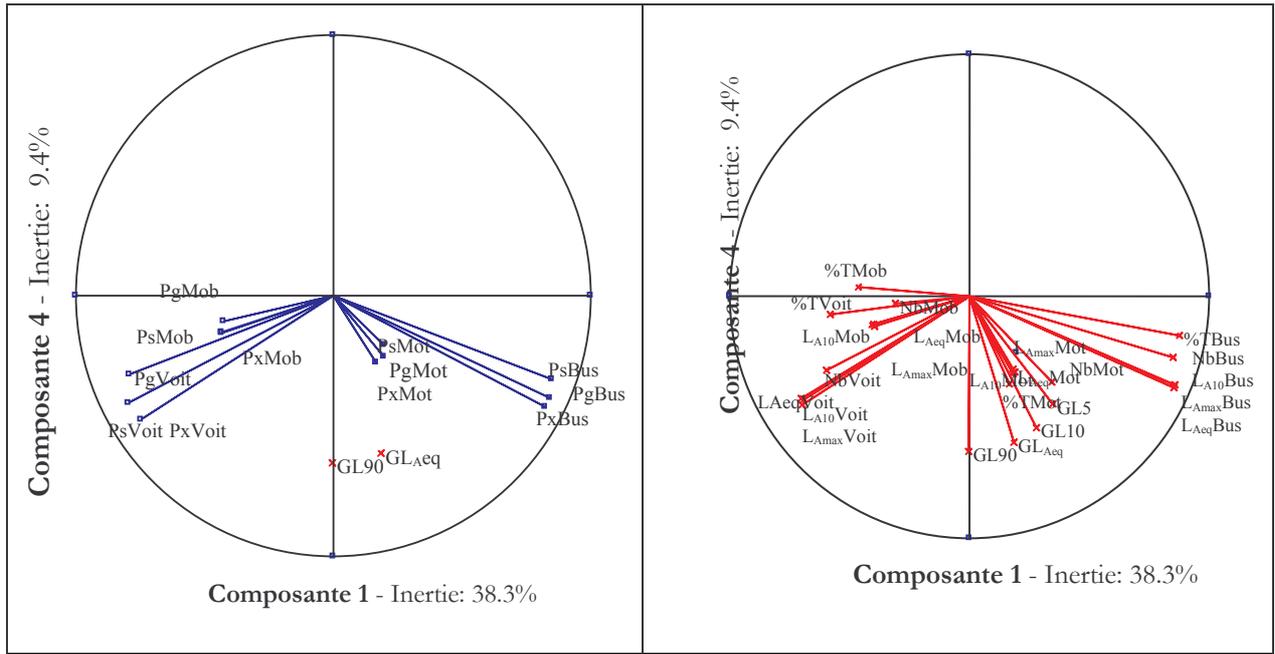
ANNEXE 3g : Représentations graphiques de l'ACP faite sur les stimuli contenant des véhicules.

Stimulus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Voit	4	2	1	5	1	0	0	3	1	0	4	1	3	0	4	1	2	2	0	0
Bus	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	2	0
Mob	0	2	0	0	1	1	2	1	1	0	1	0	0	1	0	0	2	0	0	0
Moto	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1

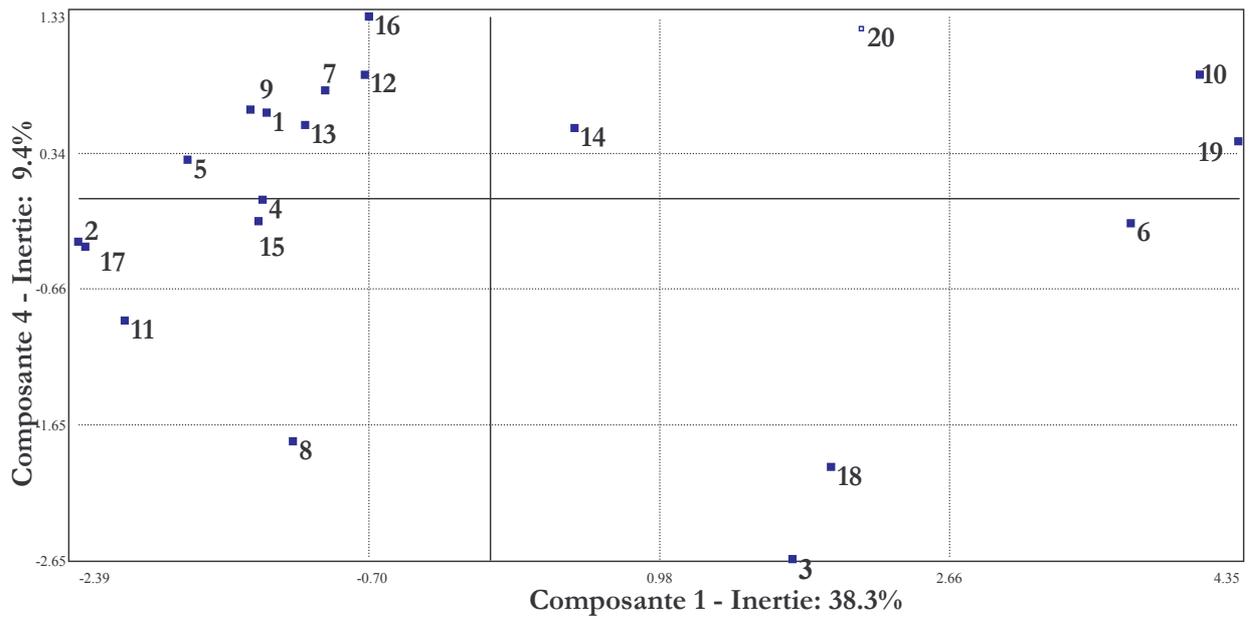


Plan Factoriel 2-3

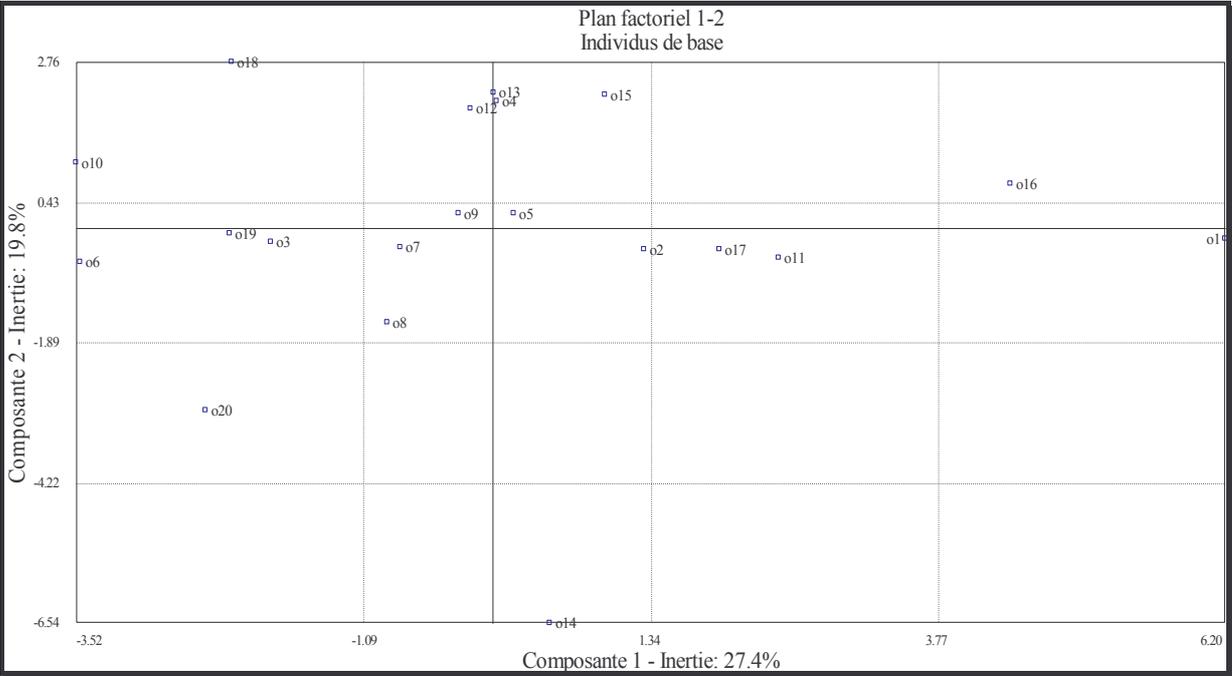
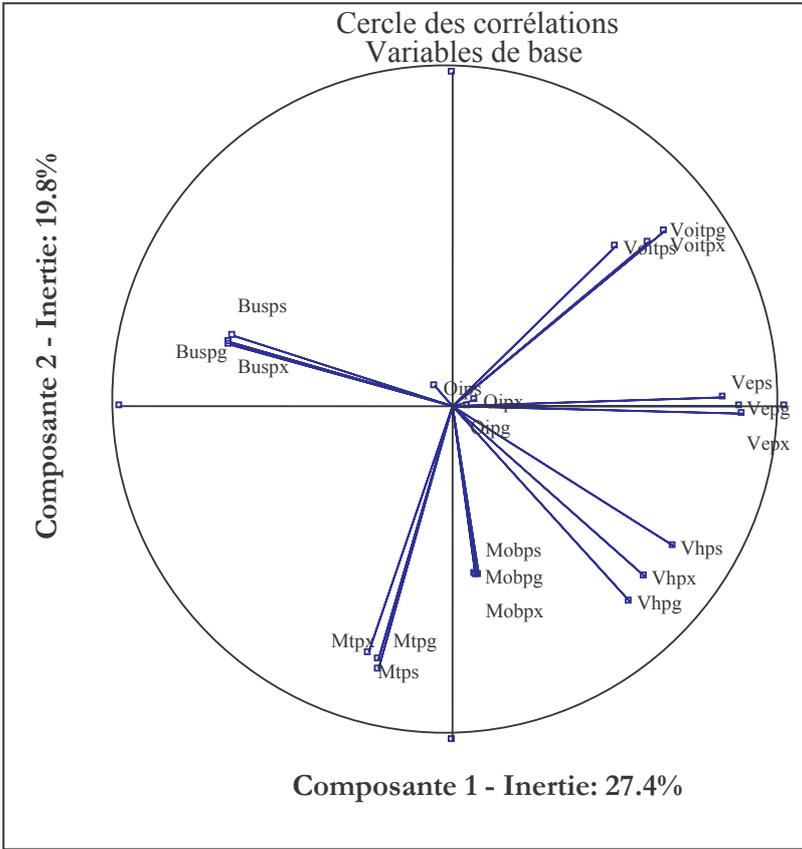


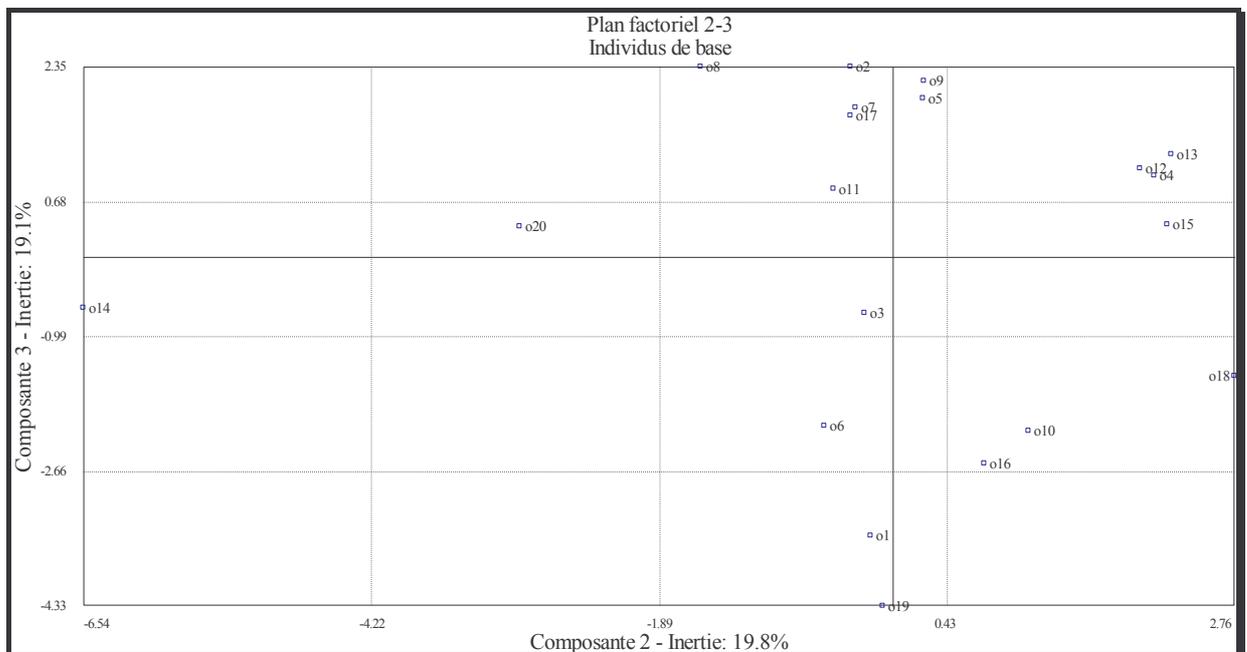
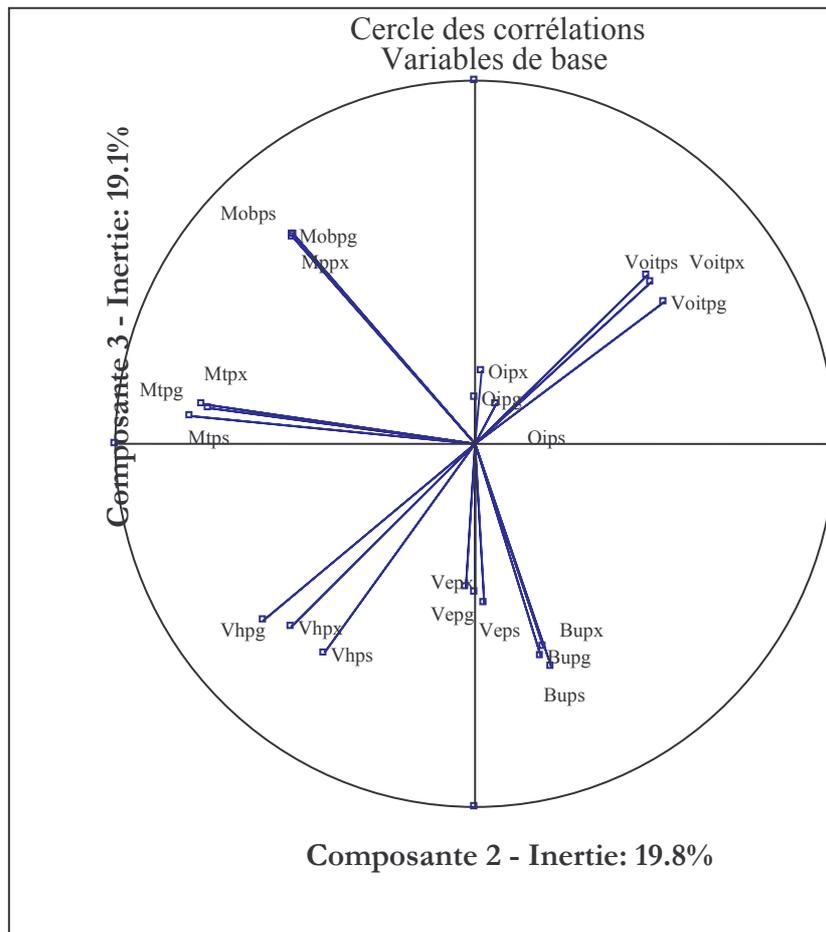


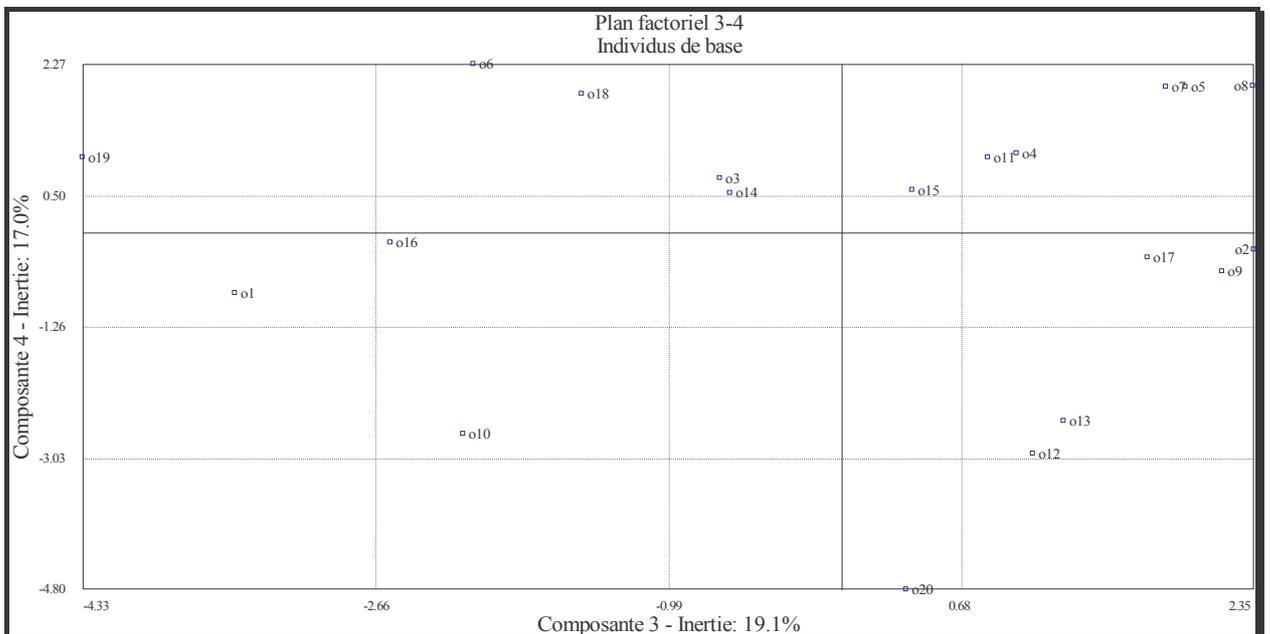
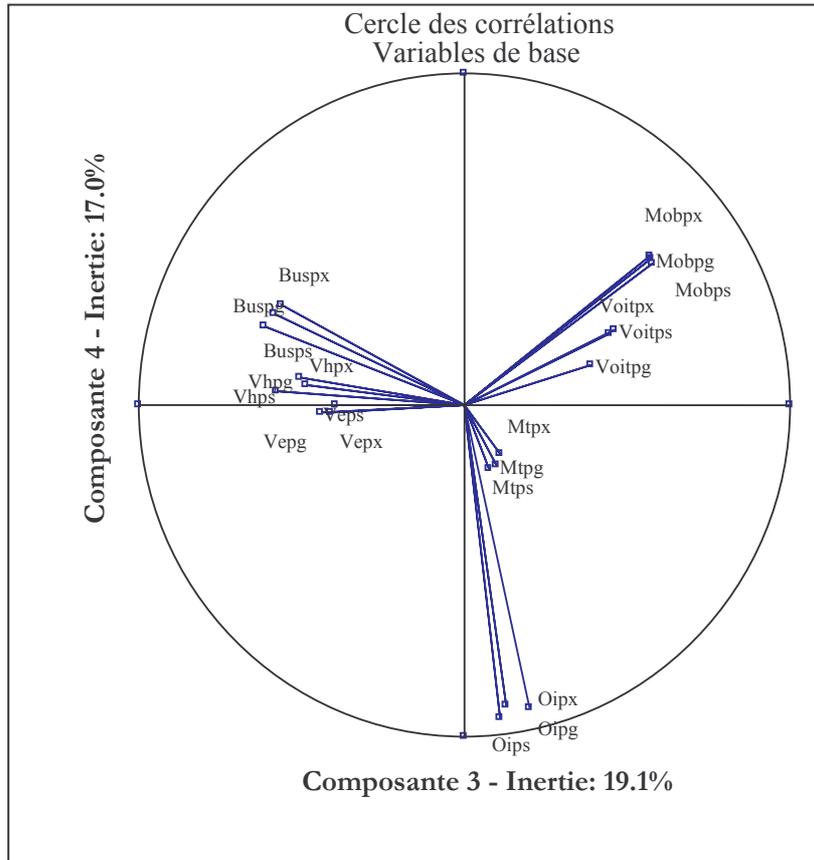
Plan Factoriel 1-4

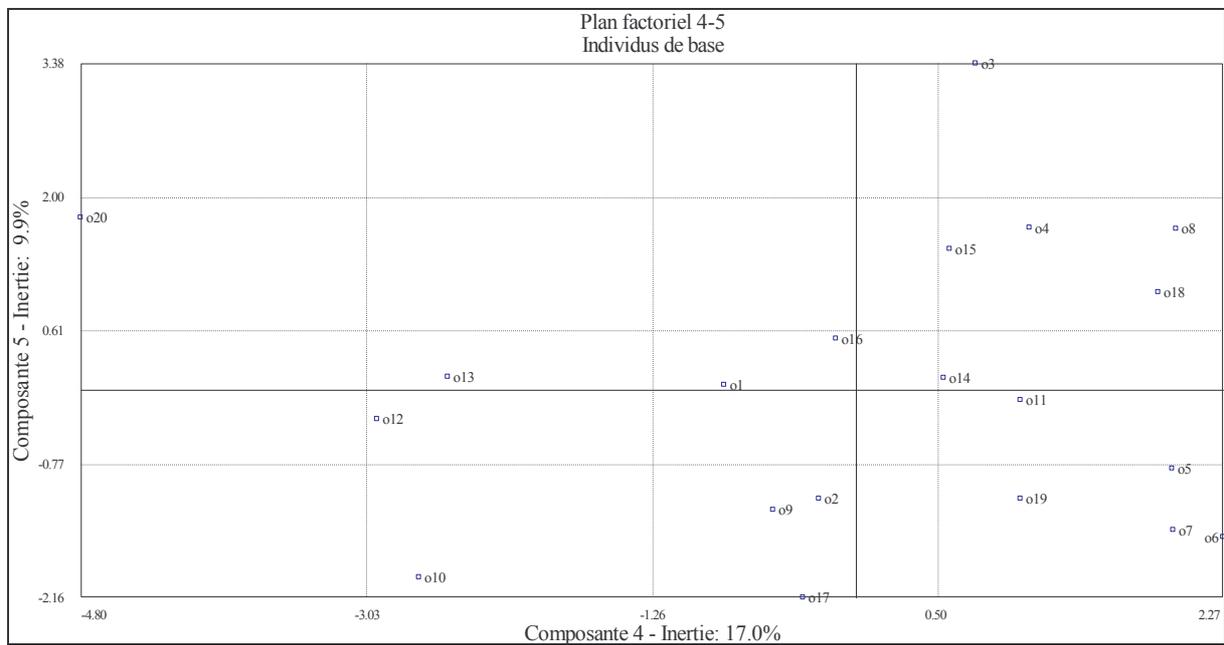
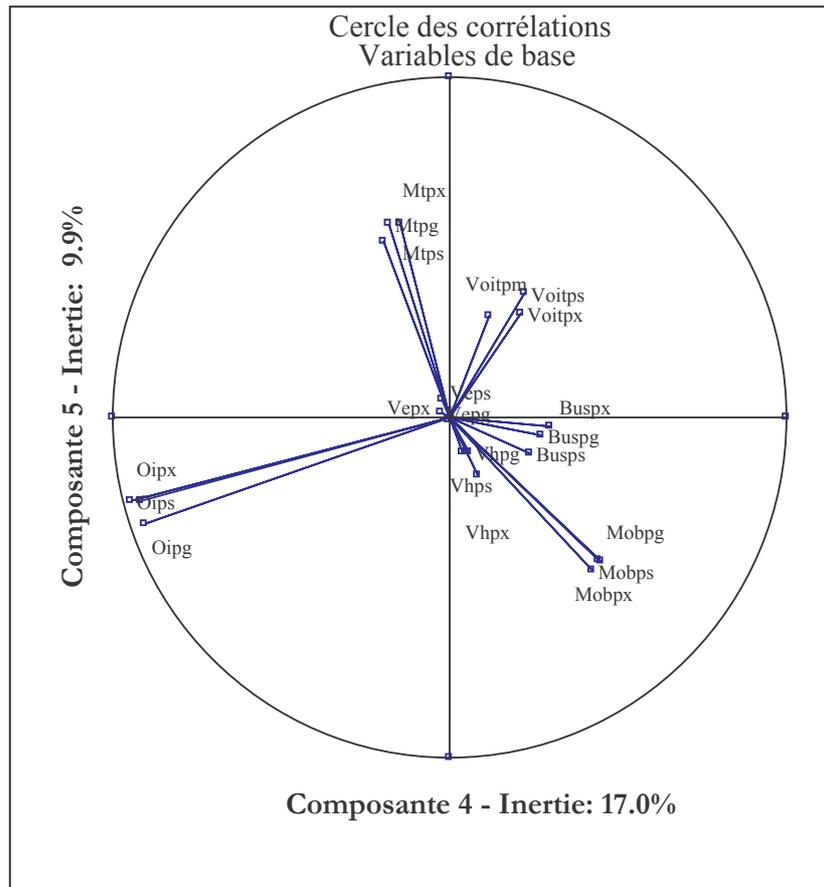


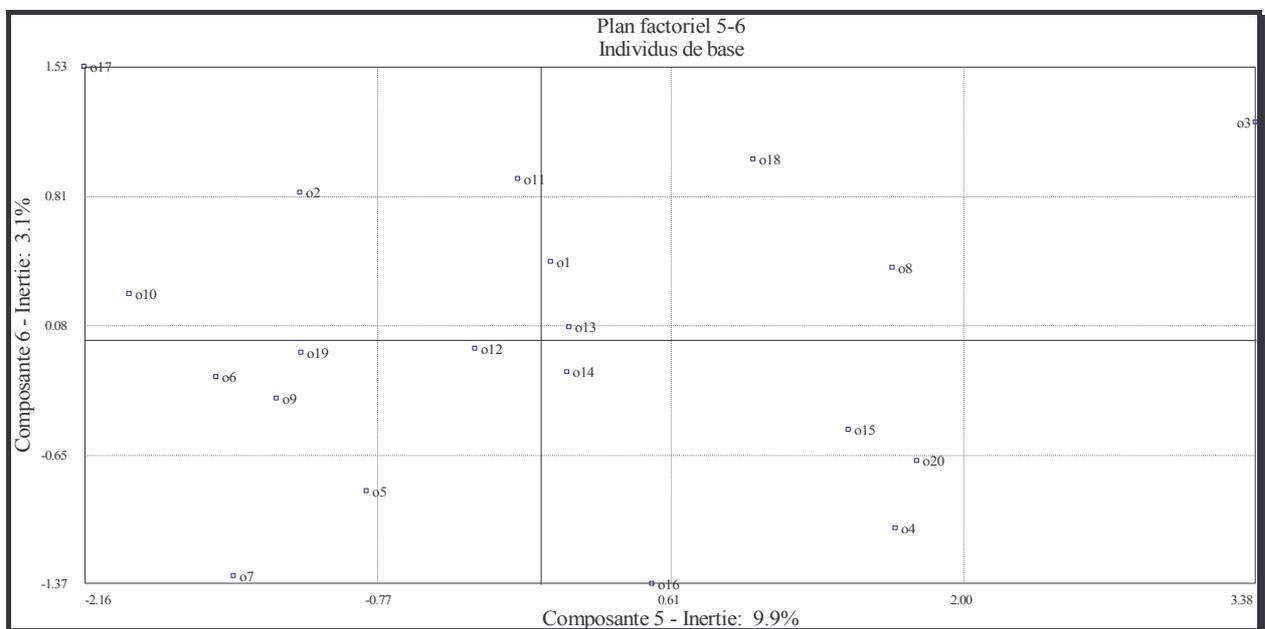
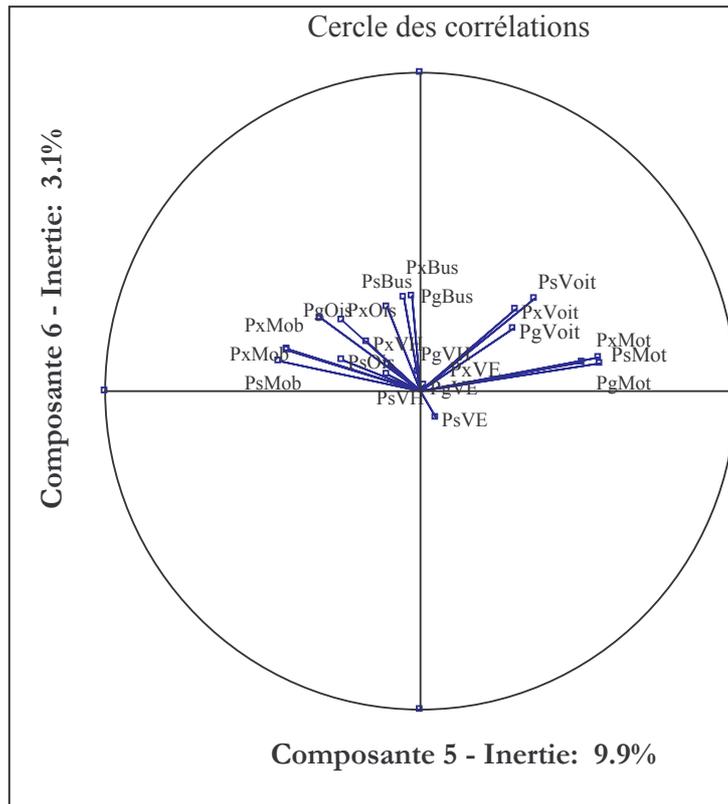
ANNEXE 3h : Représentations graphiques de l'ACP faite sur tous les stimuli.











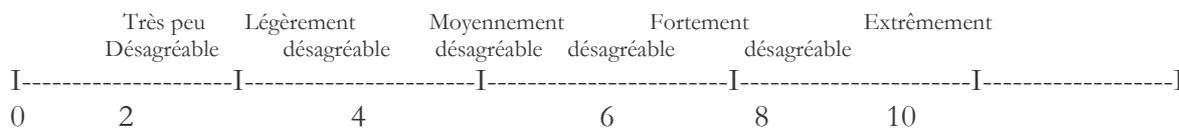
ANNEXE 4a : consigne du test 3.

Désagrément sonore perçu sur les séquences de 15 s.

Vous allez entendre vingt séquences d'une durée de 15 s chacune. Imaginez-vous un déplacement pédestre urbain tel que vous l'avez déjà vécu : vous avez un rendez-vous mais vous n'êtes pas pressés.

Pour chaque séquence écoutée, il vous est demandé de juger le désagrément que vous avez ressenti pendant l'écoute d'une séquence sur une échelle allant de « 1 » à « 10 ».

Consignes concernant l'échelle du désagrément :



Vous devez utiliser cette échelle pour évaluer l'importance de votre perception ou de votre sensation de désagrément sonore.

La note « 10 », le maximum, correspond à une sensation de désagrément extrême. C'est la plus forte perception ou sensation de désagrément sonore que vous avez déjà ressentie dans votre vécu. Si dans le cas contraire vous ne ressentez rien de désagréable, vous devez, bien sûr, mettre la note « 0 ».

Il est important que vous notiez ce que vous ressentez, et non pas ce que vous pensez qu'il est correct de dire. Soyez sincère et spontané, évitez au possible de réécrire votre réponse. N'essayez pas de vous souvenir de la note précédente. Souvenez-vous que vous devez commencer avec les graduations verbales pour ensuite choisir un nombre- le nombre que vous voulez ! – en rapport avec l'expression ou entre deux d'entre elles.

Pour vous familiariser avec les séquences, quatre d'entre elles vous sont présentées en introduction et vous aiderons à construire votre échelle. Attention par la suite, vous êtes limités à une seule écoute par séquence.

Bonne écoute

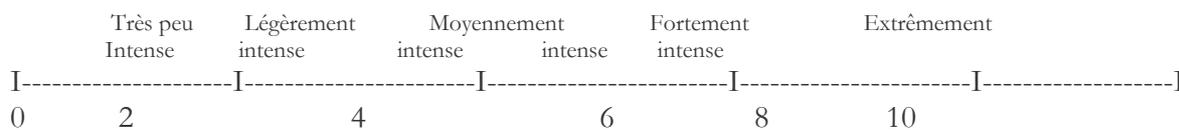
ANNEXE 4b : consigne du test 2

Niveau sonore perçu sur les séquences de 15 s.

Vous allez entendre vingt séquences d'une durée de 15 s chacune. Imaginez-vous un déplacement pédestre urbain tel que vous l'avez déjà vécu : vous avez un rendez-vous mais vous n'êtes pas pressés.

Pour chaque séquence écoutée, il vous est demandé de juger la force sonore (ou intensité subjective ou encore sonie) que vous avez ressentie sur une échelle allant de « 1 » à « 10 ».

Consignes concernant l'échelle de force sonore :



Vous devez utiliser cette échelle pour évaluer l'importance de votre perception ou de votre sensation de l'intensité sonore.

La note « 10 », le maximum, correspond à une sensation d'intensité extrême. C'est la plus forte perception ou sensation d'intensité sonore que vous avez déjà ressentie dans votre vécu.

Il est important que vous notiez ce que vous ressentez, et non pas ce que vous pensez qu'il est correct de dire. Soyez sincère et spontané, évitez au possible de réécrire votre réponse. N'essayez pas de vous souvenir de la note précédente. Souvenez vous que vous devez commencer avec les graduations verbales pour ensuite choisir un nombre- le nombre que vous voulez ! – en rapport avec l'expression ou entre deux d'entre elles.

Attention, vous êtes limités à une seule écoute par séquence.

Estimez seulement l'intensité subjective des séquences sonores et non pas le désagrément ou la gêne que pourraient vous causer les sons.

Bonne écoute

ANNEXE 4c : consigne du test in situ

Feuille d'explication fournie au Groupe n° 1,

Test auditif sur site

Renseignements :

Groupe : 1

Nom :

Prénom :

Adresse :

.....

...

.....

Principe du test:

Vous allez effectuer une promenade d'environ 15 min dans Paris. Vous allez être amené à traverser trois lieux distincts représentant trois ambiances sonores distinctes : un parc, une rue moyennement circulée et un marché. Chaque lieu a été divisé en séquences d'environ 1 minute. Il vous sera demandé, durant cette ballade, d'être sensible à l'environnement sonore, et de le juger selon des qualificatifs expliqués ci-dessous. Vous allez être divisés en trois groupes :

- un groupe notant chaque séquence (c'est-à-dire toutes les minutes)
- un groupe notant chaque ambiance (c'est-à-dire environ toutes les 5 min)
- un groupe notant la globalité de la ballade (c'est-à-dire à la fin)

Vous êtes assigné au groupe 1 notant toutes les séquences, c'est-à-dire que vous devez noter à la fin de chaque séquence la séquence que vous venez d'entendre.

Notation:

Vous allez devoir juger les trois qualificatifs suivants concernant les sources sonores que vous avez entendues:

- ***La prégnance*** : c'est l'importance, dans les sens où une source sonore s'est démarquée des autres, non pas à cause de sa durée, mais sur le fait qu'elle a attiré votre attention quand elle est apparue.
- ***La présence*** : c'est l'occurrence temporelle de la source. Cela revient à se demander si la source apparaît souvent durant la séquence.
- ***La proximité*** : c'est votre distance par rapport à la source (attention une petite note indique une source éloignée).

Les sources sonores que l'on va vous demander d'évaluer sont groupées par catégories et pré-inscrites sur la feuille de notation :

- les voitures
- les scooters et les cyclomoteurs
- les bus et les camions
- les voix
- les motos
- les chants d'oiseaux

Il vous sera également demandé de juger **à la fin de chaque séquence** la *force sonore ressentie* ainsi que le *côté agréable ou désagréable du son* entendu.

Les jugements se feront sous la forme d'une notation sur une échelle de « 0 » à « 10 ». Afin que chacun ait la même référence du maximum, la note « 10 » correspond à la sensation maximale que vous ayez rencontré durant votre expérience personnelle. Au contraire, la note « 0 » correspond à l'absence de sensation.

Pour le côté agréable ou désagréable du son, la notation se fera sur une échelle allant de « -5 » (très désagréable) à « +5 » (très agréable).

Il est important que vous notiez ce que vous ressentez, et non pas ce que vous pensez qu'il est correct de dire. Soyez sincère et spontané, évitez au maximum de réécrire votre réponse. Il est préférable que vous commenciez à choisir une graduation verbale pour ensuite choisir un nombre en rapport avec l'expression ou entre deux d'entre elles.

Veillez à bien garder en tête que c'est un test auditif d'évaluation d'ambiance sonore. Ne vous laissez pas influencer par ce que vous voyez. Ainsi, la proximité des sources doit être évaluée à l'oreille. Ceci est vrai aussi pour la reconnaissance des sources.

Bon test.

Feuille de notation pour l'ambiance *marché*,

Groupe n°1 (marché → rue → parc)																															
Ambiance marché																															
																		Nom :													
<i>il vous est demandé de noter chaque séquence</i>																															
		voitures			bus & camions			motos			mobylettes & scooters			voix			oiseaux			force sonore ressentie	agrément ou désagrément sonore										
		prégnance	présence	proximité	prégnance	présence	proximité	prégnance	présence	proximité	prégnance	présence	proximité	prégnance	présence	proximité	prégnance	présence	proximité												
ambiance "marché"	seq 1																														
	seq 2																														
	seq 3																														
	seq 4																														
	seq 5																														
échelle de notation :																															
prégnant : présent : proche :		<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">0</td> <td style="text-align: center;">2</td> <td style="text-align: center;">4</td> <td style="text-align: center;">6</td> <td style="text-align: center;">8</td> </tr> <tr> <td style="text-align: center;">-</td> <td style="text-align: center;">très peu</td> <td style="text-align: center;">légèrement</td> <td style="text-align: center;">moyennement</td> <td style="text-align: center;">fortement</td> <td style="text-align: center;">extrêmement</td> </tr> </table>																		0	2	4	6	8	-	très peu	légèrement	moyennement	fortement	extrêmement	
		0	2	4	6	8																									
-	très peu	légèrement	moyennement	fortement	extrêmement																										
force sonore ressentie :		<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">0</td> <td style="text-align: center;">2</td> <td style="text-align: center;">4</td> <td style="text-align: center;">6</td> <td style="text-align: center;">8</td> </tr> <tr> <td style="text-align: center;">^</td> <td style="text-align: center;">très peu intense</td> <td style="text-align: center;">légèrement intense</td> <td style="text-align: center;">moyennement intense</td> <td style="text-align: center;">fortement intense</td> <td style="text-align: center;">extrêmement intense</td> </tr> </table>																		0	2	4	6	8	^	très peu intense	légèrement intense	moyennement intense	fortement intense	extrêmement intense	
		0	2	4	6	8																									
^	très peu intense	légèrement intense	moyennement intense	fortement intense	extrêmement intense																										
agrément ou désagrément sonore :		<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">-5</td> <td style="text-align: center;">-3</td> <td style="text-align: center;">-1</td> <td style="text-align: center;">0</td> <td style="text-align: center;">+1</td> <td style="text-align: center;">+3</td> </tr> <tr> <td style="text-align: center;">+</td> <td style="text-align: center;">très désagréable</td> <td style="text-align: center;">désagréable</td> <td style="text-align: center;">neutre</td> <td style="text-align: center;">agréable</td> <td style="text-align: center;">très agréable</td> </tr> </table>																		-5	-3	-1	0	+1	+3	+	très désagréable	désagréable	neutre	agréable	très agréable
		-5	-3	-1	0	+1	+3																								
+	très désagréable	désagréable	neutre	agréable	très agréable																										

ANNEXE 5 : échelle de Borg

Instructions to the centiMax (Borg CR100) scale

Use this scale to describe how strong your perception or feeling is. It can be exertion, pain, annoyance or something else. One hundred (100) "Maximal" is a very important level of intensity. It's the strongest perception or feeling you've ever had. It can nevertheless be possible to feel or imagine something even stronger. Thus, the existence of "Absolute maximum" higher up and without any number and marked with a point "•". If your feeling is stronger than "100", you can use a higher number. Usually, the intensity isn't that strong that you reach "100", and seldom need you to use numbers below "2"- "3". If you are sure that you don't feel anything you should, off course, use "0". The scale works roughly as a percentage scale.

Look at the verbal anchors, Use these and chose a number close to the anchor. If your feeling or perception is "Very weak", choose a number around "6", if it is "Moderate" choose a number in the area of "20" to "30", have in mind that Moderate is about "25" and thus weaker than "Average" or "Middle". If the perception is "Strong" or "Powerful" (it feels "Heavy" or "Hard"), it's around "50". Notice that "strong" is about 50 percent i.e. half of "Max". Is the feeling "Very strong", choose a number from "60" to "80" depending on how strong it is?

It is important that you actually say what you are feeling, and not what you think what you are ought to say. Be as sincere and spontaneous as you want and try not underrating or overrating. Remember that you shall start with the verbal anchors and then chose a number -- any number you want! -- by an expression or between two of them.

ANNEXE 6 : Liste des opérateurs utilisés par EDS

Abs (x)	MFCC0 (x, MFCC_NCOEFFS)
Arcsin (x)	MFCC (x, MFCC_NCOEFFS)
Autocorrelation (x)	Min (x)
BarkBands (x, n)	Multiplication (x, MULTIPLICATION_VALUE)
Bartlett (x)	Nth (x, int INDEX)
Blackman (x)	PeakPos (x)
BpFilter (x, BPFILTER_FC1, BPFILTER_FC2)	Peaks (x)
Correlation (x, CORRELATION_SIGNAL)	Percentile (x, PERCENTILE_VALUE)
Chroma (x)	PitchBands (x, PERCENTILE_VALUE)
dB (x)	PitchBands (x, NBANDS)
Derivation (x)	Pitches (x)
Division (x, DIVISION_VALUE)	Power (x, POWER_VALUE)
Envelope (x, ENVELOPE_SIZE)	Range (x)
Fft (x)	RemoveSilentFrames (x)
FilterBank (x, n)	RHF (x)
Flatness (x)	RMS (x)
Hamming (x)	SpectralCentroid (x)
Hann (x)	SpectralDecrease (x)
Hanning (x)	SpectralFlatness (x)
HFC (x)	SpectralKurtosis (x)
Holes (x)	SpectralRollOff (x)
HpFilter (x, HPFILTER_FC)	SpectralSkewness (x)
Integration (x)	SpectralSpread (x)
Inverse (x)	Split (x, SPLIT_SIZE)
IQR (x)	SplitOverlap (x, SPLIT_SIZE, OVERLAP [0 1])
Length (x)	Sqrt (x)
Log10 (x)	Square (x)
LpFilter (x, LPFILTER_FC)	Sum (x)
Max (x)	Triangle (x)
MaxPos (x)	Variance (x)
Mean (x)	ZCR (x)
Median (x)	12Tones (x)
MelBands (x, NBANDS)	

ANNEXE 7 : Définitions de quelques opérateurs utilisés par EDS.

Opérateurs	Définitions
ADSR: Attack Decay Sustain Release	Caractéristiques de l'enveloppe temporelle d'un son: Attaque Descente Soutien Relache
Bark	Echelle de fréquences tenant compte de la perception auditive humaine
Brightness	Brillance
Bandwidth	Largeur de bande: gamme des fréquences du signal
Harmonicité	Caractérise le fait qu'un signal possède des pics de résonance forts à des intervalles de fréquence réguliers
Log attack time	Log du temps d'attaque (voir ADSR)
Loudness	Sonie (volume perçu)
Mel	Echelle de fréquences tenant compte de la perception auditive humaine
MFCC: Mel Frequency Cepstral Coefficient	Coefficients cepstraux: ils représentent la forme globale du spectre avec un faible nombre de coefficients pertinents d'un point de vue perceptif
Pitch	Fréquence perçue
RMS: Root Mean Square	Racine de la moyenne des carrés: mesure de l'énergie moyenne du signal
Spectral Centroid (Brightness)	Centroïde spectral: fréquence barycentre de l'énergie du spectre (moment d'ordre 1), elle mesure la brillance d'un son
Spectral Decrease	Décroissance spectrale: mesure la décroissance de l'amplitude spectrale (voir Spectral slope), suivant des critères perceptifs
Spectral Flatness (whiteness)	Platitude du spectre (blancheur): mesure la proximité du son avec un bruit blanc
Spectral Flux (delta spectral magnitude)	Flux spectral: mesure le taux de variation du contenu spectral
Spectral Kurtosis	Aplatissement spectral: mesure l'aplatissement du spectre par rapport à une fonction gaussienne (moment d'ordre 4)
Spectral Roll-off	Point de roulement du spectre: fréquence sous laquelle on trouve la majorité de l'énergie spectrale (typiquement 95%); c'est une approximation de la fréquence de coupure entre les parties sinusoïdale et bruit du signal.
Spectral Sharpness	Equivalent du centroïde spectral, mais calculé en utilisant l'échelle de Bark.
Spectral Skewness	Dissymétrie du spectre: mesure la dissymétrie du spectre par rapport à une distribution gaussienne (moment d'ordre 3)
Spectral Smoothness	Douceur du spectre: elle mesure la continuité de l'enveloppe spectrale, et est liée à la perception de différentes sources sonores
Spectral Slope	Pente spectrale: mesure la décroissance de l'amplitude spectrale avec la fréquence
Spectral Spread	Étendue spectrale: mesure l'étendue de l'énergie spectrale autour du centroïde (moment d'ordre 2)
Spectral Variation	Variation spectrale: mesure la variation du spectre d'une trame d'analyse à la suivante, à partir de la corrélation de leurs fonctions d'amplitude

ZCR: Zero-Crossing Rate	Taux de passage à zéro: sur des signaux non-bruités, donne une approximation de la fréquence fondamentale; sur des signaux quelconques, donne une approximation de la quantité de bruit présente dans le signal
HFC	High-Frequency Content : Contenu spectral en hautes fréquences
IQR	Inter Quartile Range : estimateur de la diffusion ou l'étalement d'un ensemble de données. Equivalent à un écart type réalisé sur 50% des données (de 25 à 75%)