



Synthèse - Reproduction - Perception des Sons Instrumentaux et Environnementaux : Application au Design Sonore

Nicolas Misdariis

► To cite this version:

Nicolas Misdariis. Synthèse - Reproduction - Perception des Sons Instrumentaux et Environnementaux : Application au Design Sonore. Mécanique des matériaux [physics.class-ph]. Conservatoire national des arts et métiers - CNAM, 2014. Français. NNT : 2014CNAM0955 . tel-01136661

HAL Id: tel-01136661

<https://theses.hal.science/tel-01136661>

Submitted on 27 Mar 2015

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.

THÈSE présentée par :

Nicolas MISDARIIS

soutenue le : 11 décembre 2014

pour obtenir le grade de : **Docteur du Conservatoire National des Arts et Métiers**

Discipline/ Spécialité : Mécanique / Acoustique

Synthèse – Reproduction – Perception des Sons Instrumentaux et Environnementaux

Application au Design Sonore

THÈSE dirigée par :

Mr Alexandre GARCIA

Professeur des Universités (Cnam Paris, Laboratoire de Mécanique des Structures et des Systèmes Couplés)

RAPPORTEURS :

Mr Bruno GAZENGEL

Professeur des Universités (Université du Maine (LAUM), UMR CNRS 6613)

Mr Jean-François PETIOT

Professeur des Universités (École Centrale de Nantes - IRCCyN - UMR CNRS 6597)

JURY :

Mr Alexandre GARCIA

Mr Bruno GAZENGEL

Mr Jean-François PETIOT

Mr Eric BAVU

Maître de Conférences (Cnam Paris, Laboratoire de Mécanique des Structures et des Systèmes Couplés)

Mr René CAUSSE

Chargé de Recherche (HDR), Responsable de l'équipe Acoustique Instrumentale à l'UMR STMS Ircam-CNRS-UPMC

Mr Brian KATZ

Chargé de Recherche (HDR) au LIMSI - CNRS

*« Je suis toujours en train de chercher ;
je pense que nous sommes maintenant sur le point de trouver »*

John Coltrane, 1965

Remerciements

La rédaction de ce document m'a donné l'occasion de me remémorer toutes les personnes avec qui j'ai collaboré – ou que j'ai pu simplement croiser – depuis que je travaille à l'Ircam. Je voudrais profiter de cette page pour les remercier toutes, sans exception, pour les échanges que nous avons eus, les idées que nous avons élaborées et l'amitié que nous avons partagée. Même s'il demeure anonyme – la liste des noms aurait été trop longue, le risque d'oubli trop important ... –, ce vœu n'en reste pas moins sincère et réellement adressé à chacun.

Cependant, parmi ces personnes, il y en a certaines que je souhaiterais sortir de cet anonymat et remercier plus particulièrement :

- Hugues Vinet, directeur scientifique, qui m'accorde sa confiance depuis mes débuts à l'Ircam ;
- René Caussé, Steve McAdams, Olivier Warusfel, Louis Dandrel, Patrick Susini, responsables scientifiques qui, à un moment donné, m'ont accueilli dans leur équipe de recherche, m'ont transmis leurs connaissances et m'ont permis d'accéder à des responsabilités de plus en plus grandes ;
- François Nicolas, Andrea Cera, Oliver Schneller, Valerio Sannicandro, Sébastien Gaxie, compositeurs que j'ai eu l'honneur et la chance d'accompagner dans un ou plusieurs de leurs projets artistiques ;
- Olivier Houix, Guillaume Lemaitre, Mondher Ayari, Jean-Julien Aucouturier, Emmanuel Ponsot, Maxime Carron, Eric Boyer, collègues avec qui je partage actuellement les réflexions et réunions sur les différents projets d'étude de l'équipe.

Dans une temporalité plus réduite et dans le contexte du doctorat, j'adresse également mes profonds remerciements à Alexandre Garcia, référent et directeur de thèse, qui m'a très amicalement conseillé et efficacement aidé depuis le début de mes démarches, ainsi qu'à Bruno Gazengel et Jean-François Petiot qui ont accepté la tâche essentielle, mais chronophage, de rapporteur.

Enfin, pour l'éternité, j'envoie un extraordinaire merci à ma famille et tous mes amis, surtout ceux du mois d'août 2014 sur qui j'ai pu compter sans faille et sans qui ce projet aurait pu ne pas tout à fait exister.

Résumé

Ce mémoire présente une composition d'études et de travaux de recherche orientés autour de trois grandes thématiques : la synthèse, la reproduction et la perception des sons, en considérant à la fois les sons de nature musicale mais aussi environnementale. Il vise en outre un champ d'application, le design sonore, qui implique globalement la création intentionnelle de sons du quotidien. La structure du document est conçue selon un schéma relativement uniforme et comporte, pour chaque partie, une présentation générale de la thématique apportant des éléments théoriques et des données relatives à l'état de l'art, suivie de développements spécifiques permettant de converger vers les sujets d'étude propres à chaque thème – explicitement, formalisme modal dans la synthèse par modélisation physique, pour la partie "Synthèse" ; mesure et contrôle de la directivité des instruments de musique, pour la partie "Reproduction" ; timbre et identification des sources sonores, pour la partie "Perception" – puis d'une présentation détaillée des travaux personnels relatifs à chacun des sujets, le cas échéant, sous la forme d'un article publié. Ces divers éléments de connaissance et d'expérience proposent donc une contribution personnelle et originale, volontairement inscrite dans un cadre de recherche élargi, pluridisciplinaire et appliqué.

Mots-clés : synthèse sonore, modélisation physique, reproduction sonore, rayonnement acoustique, directivité, perception sonore, timbre, sons environnementaux, design sonore, sonification, interface homme-machine.

Abstract

This dissertation presents a composition of studies and research works articulated around three main topics : synthesis, reproduction and perception of sounds, considering both musical and environmental sounds. Moreover, it focuses on an application field, the sound design, that globally involves the conception of intentional everyday sounds. The document is based on a rather uniform structure and contains, for each part, a general presentation of the topic which brings theoretical elements together with an overview of the state-of-the-art, followed by more precise developments in order to focus on the specific matters related to each topic – in detail, modal formalism in sound synthesis by physical modeling, for the "Synthesis" section ; measurement and control of musical instruments directivity, for the "Reproduction" section ; timbre and sound sources identification, for the "Perception" section – and then followed by a detailed presentation of the personal works related to each matter, in some cases, in the form of published papers. Then, these several elements of knowledge and experience offer a personal and original contribution, deliberately put in a broad, multidisciplinary and applied framework.

Keywords : sound synthesis, physical modeling, sound reproduction, acoustic radiation, directivity, auditory perception, timbre, environmental sounds, sound design, sonification, computer-human interface.

Table des Matières

REMERCIEMENTS.....	3
RESUME.....	5
ABSTRACT	6
TABLE DES MATIERES.....	7
LISTE DES TABLEAUX.....	12
LISTE DES FIGURES	13
LISTE DES ANNEXES	16
INTRODUCTION.....	18
PARTIE I. SYNTHESE	22
I.1 PROBLEMATIQUE	22
I.2 SYNTHESE PAR MODELISATION PHYSIQUE	25
I.3 LE FORMALISME MODAL	27
I.4 DESCRIPTION DE MODALYS	29
I.4.1 Modalys : Eléments.....	30
I.4.2 Modalys : Architecture.....	31
I.4.3 Modalys : Programmation.....	32
I.5 DEVELOPPEMENT D'UN MODELE D'ACCORDEON DANS MODALYS	33
I.5.1 L'accordéon	33
I.5.1.1 Introduction	33
I.5.1.2 Etat de l'art	34
I.5.1.3 Fonctionnement	34
I.5.1.4 Approche expérimentale	36
I.5.2 Implémentation dans Modalys.....	37
I.5.2.1 La connexion 'reed-free'	37
I.5.2.2 Construction et évaluation du modèle	38
I.5.3 Conclusion de l'étude.....	39
I.6 ETUDE DE L'ACOUSTIQUE DE L'ANCHE LIBRE	40
I.6.1 Préambule.....	40
I.6.2 Publication [Misdariis et al., Congrès Français d'Acoustique, 2000]	40
I.6.2.1 Introduction	40
I.6.2.2 Le son de l'accordéon : caractéristiques majeures	41
I.6.2.3 Caractérisation physique de l'instrument	41
I.6.2.4 Caractérisation de l'écoulement.....	43
I.6.2.5 Modèle physique de la vibration de l'anche.....	44
I.6.2.6 Conclusion – Perspectives	45
I.7 CONCLUSION	45
PARTIE II. REPRODUCTION	48
II.1 PROBLEMATIQUE.....	48
II.2 ETUDE DU RAYONNEMENT ACOUSTIQUE.....	53
II.2.1 Modélisation du rayonnement du haut-parleur : le piston plan.....	54
II.2.2 Description du champ acoustique sur la base des harmoniques sphériques.....	56
II.3 CONTROLE DU RAYONNEMENT ACOUSTIQUE	57
II.3.1 Etat de l'art	57
II.3.2 Etat des lieux.....	65
II.3.2.1 Système multi haut-parleurs pour simulation de source omnidirectionnelle	65
II.3.2.2 Système multi haut-parleurs pour contrôle et reproduction du rayonnement.....	65

<i>II.3.3 Système multi haut-parleurs à directivité contrôlable : projet Timée.....</i>	<i>67</i>
II.3.3.1 Préambule	67
II.3.3.2 Publication [Misdariis et al., Journées d'Informatique Musicale, 2002]	68
II.3.3.2.1 Introduction	68
II.3.3.2.2 Méthodes de reproduction	68
II.3.3.2.2.1 Approche "statique"	68
II.3.3.2.2.2 Approche "modale"	69
II.3.3.2.3 Résultats	71
II.3.3.2.3 Une nouvelle instrumentalisation de la musique mixte	72
II.3.3.2.3.1 Problématique	72
Possibilités sonores et musicales du dispositif	73
II.3.3.2.3.2 Premières performances	73
II.3.3.2.4 Une nouvelle simulation de l'instrument de musique	74
II.3.3.2.4.1 Contrôle indépendant du timbre et de la directivité	74
II.3.3.2.4.2 Vers une synthèse sonore "augmentée"	75
II.3.3.2.5 Conclusion	76
II.4 ETUDE DU RAYONNEMENT APPLIQUE A LA REPRODUCTION D'UN INSTRUMENT	76
II.4.1 Description de l'étude	77
II.4.2 Caractérisation des données	79
II.4.2.1 Dispositif expérimental	79
II.4.2.2 Mesure de la source	80
II.4.2.3 Mesure des cibles	82
II.4.3 Modélisation des données	83
II.4.4 Résultats	85
II.4.5 Conclusion	87
II.5 EXTENSIONS DU PROJET	87
II.5.1 La Synthèse Additive Spatiale	87
II.5.1.1 Analyse et reproduction des propriétés spectrales et spatiales du violon	89
II.5.1.1.1 Préambule	89
II.5.1.1.2 Publication [Vos et al., Journ. Acoust. Soc. Nether., 2003]	89
II.5.1.1.2.1 Introduction	89
II.5.1.1.2.2 Theory	90
II.5.1.1.2.2.1 Measurements	90
II.5.1.1.2.2.2 Reproduction	90
II.5.1.1.2.3 Measurement	91
II.5.1.1.2.3.1 Set-up	91
II.5.1.1.2.3.2 Spectrum	93
II.5.1.1.2.4 Analysis	94
II.5.1.1.2.4.1 Frequency Analysis	94
II.5.1.1.2.4.2 Directivity analysis	95
II.5.1.1.2.5 Reproduction	96
II.5.1.1.2.6 Conclusions	97
II.5.1.2 Mesure et contrôle du rayonnement appliquée à la synthèse sonore du violon	98
II.5.1.2.1 Préambule	98
II.5.1.2.2 Publication [Misdariis et al., Congrès Français d'Acoustique, 2006]	98
II.5.1.2.2.1 Introduction	98
II.5.1.2.2.2 Outils de l'étude	99
II.5.1.2.2.2.1 Synthèse sonore	99
II.5.1.2.2.2.2 Synthèse du rayonnement	100
II.5.1.2.2.2.2.1 Généralités	100
II.5.1.2.2.2.2.2 Implémentation et Propriétés	100
II.5.1.2.2.3 Objet de l'étude : le violon	101
II.5.1.2.2.3.1 Présentation	101
II.5.1.2.2.3.2 Expérience préliminaire	102
II.5.1.2.2.3.3 Dispositif expérimental	102
II.5.1.2.2.3.3.1 Mesures vibratoires	103
II.5.1.2.2.3.3.2 Mesures acoustiques	103
II.5.1.2.2.3.4 Premiers résultats	103
II.5.1.2.2.4 Perspectives	104
II.5.1.2.2.5 Conclusion	105
II.5.2 Evaluation perceptive du contrôle de la directivité	105
II.5.2.1 Cadre général	106
II.5.2.2 Objectifs	107
II.5.2.3 Expérience	107

II.5.2.3.1 Stimuli.....	107
II.5.2.3.1.1 Plaque frappée.....	108
II.5.2.3.1.2 Voix parlée/chantée.....	109
II.5.2.3.2 Apparatus.....	110
II.5.2.3.3 Plan d'expérience.....	110
II.5.2.3.4 Protocole.....	110
II.5.2.4 Résultats et Analyses.....	112
II.5.2.4.1 Moyennes et écarts-type.....	113
II.5.2.4.2 Définition des outliers.....	113
II.5.2.4.3 Analyses de la variance.....	114
II.5.2.5 Conclusion et Perspectives.....	116
II.6 CONCLUSION.....	117
PARTIE III. PERCEPTION.....	120
III.1 PROBLEMATIQUE.....	120
III.2 PERCEPTION SONORE : TIMBRE ET IDENTIFICATION DES SOURCES.....	128
III.2.1 Introduction.....	129
III.2.2 Etude du timbre.....	130
III.2.3 Identification et classification des sources.....	134
III.3 PARADIGMES EXPERIMENTAUX.....	138
III.3.1 Test de dissemblance.....	138
III.3.2 Test de catégorisation.....	139
III.3.3 Test de différentiel sémantique.....	141
III.4 TIMBRE DES SONS MUSICAUX : DEVELOPPEMENT D'UN MOTEUR DE RECHERCHE (PROJET SOL).....	142
III.4.1 Préambule.....	142
III.4.2 Publication [Misdariis et al., International Congress on Acoustics, 1998a].....	144
III.4.2.1 Theoretical Basis.....	144
III.4.2.2 Construction of the Distance Model.....	145
III.4.2.3 Application of the Model to a Real Database.....	145
III.5 EVALUATION OBJECTIVE DE LA QUALITE PERÇUE D'UN INSTRUMENT DE MUSIQUE.....	146
III.5.1 Préambule.....	146
III.5.2 Publication [Caussé et al., Congrès Français d'Acoustique, 1997].....	147
III.5.2.1 Introduction.....	147
III.5.2.2 L'évaluation subjective des instruments.....	147
III.5.2.2.1 Les tests perceptifs.....	147
III.5.2.2.2 L'analyse multidimensionnelle.....	148
III.5.2.3 La caractérisation physique des instruments.....	148
III.5.2.3.1 Approche morphologique.....	148
III.5.2.3.2 Approche acoustique.....	149
III.5.2.4 Relation critères subjectifs / paramètres physiques.....	149
III.5.2.5 Conclusion et perspectives.....	150
III.5.3 Publication [Jossic et al., Congrès Français d'Acoustique, 2014].....	150
III.5.3.1 Introduction.....	150
III.5.3.2 Système de contrôle actif numérique.....	151
III.5.3.2.1 Fonction de transfert en boucle fermée.....	151
III.5.3.2.2 Phase en boucle ouverte.....	152
III.5.3.2.3 Description du contrôle numérique.....	152
III.5.3.3. Approche perceptive de l'étude.....	154
III.5.3.3.1 Approche psychophysique et écologique de l'étude.....	154
III.5.3.3.2 choix des tests.....	154
III.5.3.4 Protocole expérimental.....	155
III.5.3.4.1 Espace des paramètres et configurations.....	155
III.5.3.4.2 Sujets.....	155
III.5.3.4.3 Déroulement de la session.....	155
III.5.3.5 Résultats.....	156
III.5.3.5.1 MDS INDSCAL globale.....	156
III.5.3.5.2 Analyse des données verbales.....	157
III.5.3.5.2.1 Champs lexicaux.....	157
III.5.3.5.2.2 Consensus des musiciens.....	157
III.5.3.5.3 Corrélation des verbalisations et des paramètres physiques.....	157
III.5.3.5.3.1 Effet de l'augmentation de gain à phase constante.....	158
III.5.3.5.3.2 Effet de la phase à gain maximal.....	159

III.5.3 6. Conclusion	160
III.6 TIMBRE DES SONS ENVIRONNEMENTAUX : META-ANALYSE	160
III.6.1 <i>Préambule</i>	160
III.6.2 <i>Publication</i> [Misdariis et al., Journ. Audio, Speech and Music Process., 2010b]	161
III.6.2.1 Introduction	161
III.6.2.2 Primary Studies	162
III.6.2.2.1 Studies A (A1, A2): car interior	164
III.6.2.2.2 Study B: interior air-conditioning units	164
III.6.2.2.3 Study C: car horns	165
III.6.2.2.4 Study D: car door closing	165
III.6.2.2.5 Comparisons and discussion	166
III.6.2.3 Meta-processing: complementary experimental investigations	167
III.6.2.3.1 Experiment 1: Free sorting task on the initial corpora	168
III.6.2.3.1.1 Method	168
III.6.2.3.1.2 Results	168
III.6.2.3.2 Experiment 2: Forced-choice sorting task on an extended corpus	170
III.6.2.3.2.1 Method	171
III.6.2.3.2.2 Results	171
III.6.2.3.3 Discussion	172
III.6.2.4 Meta-processing: modeling the description structure	172
III.6.2.4.1 Continuous level: unifying the perceptual space dimensions	173
III.6.2.4.1.1 MDS analyses compatibility	173
III.6.2.4.1.2 <i>Motor</i> category	173
III.6.2.4.1.3 <i>Instrument-like</i> category	177
III.6.2.4.1.4 <i>Impact</i> category	179
III.6.2.4.1.5 Discussion	181
III.6.2.4.2 Categorical level: building an automatic classifier	181
III.6.2.4.2.1 Specificities of the categories	182
III.6.2.4.2.2 Classification modeling tool: the multinomial logistic regression	182
III.6.2.4.2.3 Model selection	183
III.6.2.4.2.4 Model validation	184
III.6.2.4.2.5 Discussion	185
III.6.2.4.3 Summary	185
III.6.2.5 Conclusion	186
III.7 CONCLUSION	187
PARTIE IV. DOMAINES D'APPLICATION	190
IV.1 DESIGN SONORE	190
IV.2 ETUDE ET CONCEPTION DE SONS DE CLIGNOTANT	194
IV.2.1 <i>Préambule</i>	194
IV.2.2 <i>Publication</i> [Langlois et al., Forum Acusticum, 2005]	195
IV.2.2.1. Introduction	195
IV.2.2.2. Sound Quality Study	195
IV.2.2.2.1 Results	196
IV.2.2.2.2 Conclusion	197
IV.2.2.3. Sound Design	197
IV.2.2.3.1 Method	198
IV.2.2.3.2 Conclusion	198
IV.2.2.4. Technological Impact on Sound Design	198
IV.2.2.4.1 Method	198
IV.2.2.4.2 Results	199
IV.2.2.4.3 Conclusion	200
IV.2.2.5 Conclusion	200
IV.3 STRATEGIES DE SONIFICATION D'UN CLAVIER D'ATM	201
IV.3.1 <i>Préambule</i>	201
IV.3.2 <i>Publication</i> [Susini et al., Journ. Multimodal User Interfaces, 2012]	201
IV.3.2.1 Introduction	201
IV.3.2.1.1 Auditory displays: mapping sounds and information	202
IV.3.2.1.2 Causality and Naturalness	203
IV.3.2.1.3 Sonic Interactions: mapping sounds and actions	203
IV.3.2.1.4 How to evaluate an interface ?	204
IV.3.2.1.5 Goals of the study	205
IV.3.2.1.6 Outline	205
IV.3.2.2. Design of the sonified ATM interface	206

IV.3.2.2.1 The ATM interface.....	206
IV.3.2.2.2 Creation of the sonic feedback types.....	206
IV.3.2.2.3 Selection of the feedback sounds.....	207
IV.3.2.3. Rating the sonic feedback.....	208
IV.3.2.3.1 Three scales of evaluation.....	208
IV.3.2.3.2 Method.....	208
IV.3.2.3.3 Results.....	209
IV.3.2.3.3.1 Ratings of naturalness in steps 0 and 1.....	209
IV.3.2.3.3.2 Analysis of the pre-experimental ratings (step1).....	210
IV.3.2.3.3.3 comparison between the pre- and post- experimental ratings (step 3).....	212
IV.3.2.3.3.4 Direct comparisons (step 4).....	212
IV.3.2.3.3.5 Validation (step 5).....	214
IV.3.2.3.4 Discussion.....	215
IV.3.2.4. Conclusion.....	216
IV.4 NAVIGATION DANS UN MENU HIERARCHIQUE.....	217
IV.4.1 <i>Préambule</i>	217
IV.4.2 <i>Publication</i> [Langlois et al., IHM conférence, 2010].....	218
IV.4.2.1 Introduction.....	218
IV.4.2.2 Description du guidage sonore.....	218
IV.4.2.2.1 Sons de confirmation d'action (« feedback sounds »).....	220
IV.4.2.2.2 Sons hiérarchiques.....	221
IV.4.2.2.3 Spearcons et synthèse vocale.....	221
IV.4.2.2.4 Extraits musicaux prototypiques.....	221
IV.4.2.3 Evaluation.....	222
IV.4.2.3.1 Protocole d'évaluation.....	222
IV.4.2.4 Résultats.....	222
IV.4.2.4.1 Qualité des librairies hybrides et musèmes.....	222
IV.4.2.4.2 Efficacité du guidage sonore {étape 6}.....	223
IV.4.2.4.3 Avis post conduite des participants {étape 7}.....	226
IV.4.2.5 Discussion.....	226
IV.4.2.6 Conclusions et perspectives.....	229
IV.4.3 <i>Extension de l'étude</i>	229
IV.4.3.1 Rappel du cadre général.....	229
IV.4.3.2 Nouveau protocole d'évaluation.....	230
IV.4.3.2.1 Cadre et objectif.....	230
IV.4.3.2.2 Dispositif expérimental.....	231
IV.4.3.2.3 Stimuli.....	231
IV.4.3.2.4 Plan d'expérience.....	232
IV.4.3.2.5 Procédure.....	233
IV.4.3.2.6 Premiers résultats.....	233
IV.4.3.2.7 Discussion.....	235
IV.4.3.3 Conclusion.....	235
IV.5 CONCLUSION.....	236
CONCLUSION.....	240
BIBLIOGRAPHIE.....	242
ANNEXES.....	266
ANNEXE 1. RAYONNEMENT ACOUSTIQUE DU PISTON PLAN CIRCULAIRE.....	267
ANNEXE 2. RESOLUTION DE L'EQUATION D'ONDE EN COORDONNEES SPHERIQUES.....	269
ANNEXE 3. MESURES DE LA SOURCE GSR.....	271
ANNEXE 4. MESURES DE LA DIRECTIVITES DU SAXOPHONE.....	273
ANNEXE 5. MESURES DE LA DIRECTIVITE DU COR DE BASSET.....	275
ANNEXE 6 : COMPLEMENTARY DATA AND INITIAL RESULTS RELATED TO THE 4 PRIMARY STUDIES.....	277
ANNEXE 7 : ILLUSTRATION OF THE GRAPHICAL USER INTERFACES USED IN EXPERIMENTS 1 AND 2.....	283
ANNEXE 8 : DETAILS OF ACOUSTIC FEATURES CALCULATION.....	285

Liste des tableaux

Tableau 1. Fréquences propres du violon (Hz) – recoupement biblio.....	104
Tableau 2. Séquence des phonèmes parlés (droite) et chantés {min. : Lab2 ; maj. : Lab3} (gauche).....	109
Tableau 3. Modalités des variables et facteurs expérimentaux.....	112
Tableau 4. Couplage des facteurs expérimentaux système de diffusion (D) et mode de rayonnement (R) en un facteur complexification du système de diffusion (C).	115
Tableau 5. Récapitulatif des résultats des 2 analyses de contraste selon les stimuli (Pl., plaques ; Vo., voix) et les échelles d'évaluation (E) – les contrastes significatifs sont notés *.....	116
Table 6. Correlation between perceptual dimensions and physical parameters (Krimphoff et al. [4,5])	145
Table 7. Table recapitulating methodological context and main results for studies A1, A2, B, C, and D (see corresponding sections above and Appendix 6 for further details).....	166
Table 8. Experiment 2 – distribution of the sounds in the experimental categories.....	172
Table 9. Correlations between acoustic features and dimensions of the <i>motor</i> category / studies A1, A2, B ($df = 14, 12, 17$, respectively, $** p < 0.01$).....	174
Table 10. Coefficients of the linear combination defining Complex brightness for studies A1, A2 and B.	176
Table 11. Correlations between acoustic features and dimensions of the <i>instrument-like</i> category / study C ($df = 20$, $** p < 0.01$).....	178
Table 12. Correlations between acoustic features and dimensions of the <i>impact</i> category / study D ($df = 10$, $** p < 0.01$).....	179
Table 13. LR value for each spectral/temporal feature pair.	184
Table 14. Results of the predicting tool based on SSp/ED features, after 100 runs of a 70%-learning/30%-predicting loop on the 207-sound perceptually extended corpus (Experiment 2).....	185
Table 15. Different sounds used to sonify ATM keyboard – HF: high-freq; MF: medium-freq; LF: low-freq.	207
Table 16. Pre-experimental ratings for the subjects in the two groups (G1: normal mode of operation; G2: defective mode of operation) on the three scales: average and standard deviation (in brackets)	211
Table 17. Correlation matrix of pre- and post-experimental ratings (N = 90) for the naturalness (N), the usability (U), and pleasantness (P) scales. ***: $p < 0.0001$, **: $p < 0.01$	212
Tableau 18. Evaluation de la cohérence son/fonction	223

Liste des figures

Figure 1. Illustration d'un générateur élémentaire de type oscillateur à lecture de table [Roads, 1996].	23
Figure 2. Illustration de la connection (interaction) 'frotter' dans Modalys.	30
Figure 3. Schéma de l'anche libre de l'accordéon	35
Figure 4. Schéma de la "musique" de l'accordéon	35
Figure 5. Schéma de la "platine" et de l'anche libre de l'accordéon	35
Figure 6. Dispositifs expérimentaux de mesure de l'anche d'accordéon	36
Figure 7. Mouvement d'une anche libre fixée sur une platine en plexiglas	36
Figure 8. Modélisation de l'accordéon : données mesurées sur le dispositif expérimental	38
Figure 9. Modélisation de l'accordéon : résultats issus du modèle de synthèse	38
Figure 10. Accordéon – Evolution de la pression acoustique	41
Figure 11. Accordéon – Spectre du son tenu (de l'accordéon)	41
Figure 12. Accordéon – Evolution des signaux de pression et de déplacement (mesure)	42
Figure 13. Accordéon – Fréquence d'oscillation en fonction de P_0	43
Figure 14. Accordéon – Evolution des signaux de pression et de déplacement (simulation)	44
Figure 15. Illustration de la force électromagnétique (force de Laplace)	49
Figure 16. Illustration de la force électrostatique (loi de Coulomb)	49
Figure 17. Détails d'un haut-parleur électrodynamique	50
Figure 18. Détails d'un haut-parleur électrostatique	50
Figure 19. Illustration du principe de la théorie Duplex de Rayleigh. (ITD et ILD)	51
Figure 20. Illustration du principe des HRTFs (<i>Head-Related Transfer Functions</i>)	52
Figure 21. Illustration du théorème de Huyghens	52
Figure 22. Illustration du principe de décomposition en harmoniques sphériques	53
Figure 23. Rayonnement du piston plan – données du problème	54
Figure 24. Diagrammes de directivité d'un piston plan en fonction de θ et ka	55
Figure 25. Système de coordonnées sphériques	56
Figure 26. Partie réelle des harmoniques sphériques pour les 4 premiers degrés	57
Figure 27. Mesure de la fonction de transfert entre 2 directions d'émission d'un instrument	58
Figure 28. Mesure de la directivité du trombone (gauche) et du violon (droite)	58
Figure 29. Mesure du rayonnement des instruments selon Garcia [Garcia, 1983]	59
Figure 30. Mesure du rayonnement acoustique par une méthode indirecte [Weinreich, 1997]	59
Figure 31. Mesure du rayonnement du violon par la méthode indirecte (données)	60
Figure 32. <i>DTC loudspeaker</i> – Dispositif expérimental élaboré par G. Weinreich pour restituer le rayonnement caractéristique d'un signal de violon [Weinreich, 1997b]	61
Figure 33. Dispositif de mesure du rayonnement selon Bresciani [Bresciani, 1996]	61
Figure 34. Mesure de la directivité de la clarinette pour le doigté "tous trous fermés" [Bresciani, 1996]	62
Figure 35. Expérience de Bresciani – Reproduction du trombone par un haut-parleur dodécaédrique à directivité contrôlée [Bresciani, 1996]	62
Figure 36. Source icosaoédrique à 120 voies développée par le CNMAT	66
Figure 37. Source icosaoédrique	66
Figure 38. Source dodécaédrique	66
Figure 39. Source cubique à 6 voies – gauche – et cubique à 12 x2 voies (+ 1 caisson BF) – droite – développée par l'Ircam	67
Figure 40. Reproduction du rayonnement (approche "statique")	69
Figure 41. Implémentation du traitement du signal (approche statique)	69
Figure 42. Approche "modale" : figures de directivité (gauche) et nouvelle implémentation (droite)	70
Figure 43. Approche "modale" : prototype multi-voies	70
Figure 44. Prototype multi-voies : égalisation en puissance de H0 et H1	71
Figure 45. Reproduction de H0 à 500 Hz (gauche), 2000Hz (centre) et 10000 Hz (droite)	71
Figure 46. Reproduction de H1 à 500 Hz (gauche), 2000Hz (centre) & 10000 Hz (droite)	71
Figure 47. Prototype multi-voies : énergie frontale (gauche), énergies frontale/latérale (droite)	72
Figure 48. Prototype multi-voies : morphing de directivité (gauche), contrepoint spatial (droite)	73
Figure 49. Prototype multi-voies : portrait cubiste	73
Figure 50. Mesures et modèles de spectres de puissance d'instruments : clavecin (gauche), flûte (droite)	75
Figure 51. Global Source Radiator (GSR) et nomenclature des haut-parleurs	77
Figure 52. Cor de Basset (gauche) – Saxophone tenor (droite)	78

Figure 53. Dispositif de mesure chambre anéchoïque.....	78
Figure 54. Dispositif de mesure du rayonnement en chambre sourde	79
Figure 55. Surface de Voronoï sur la sphère et pondération selon une ligne de microphones	80
Figure 56. Champ acoustique mesuré pour le HP n° 7 du GSR	81
Figure 57. Singularités de rayonnement du HP n° 7 du GSR.....	81
Figure 58. Dispositif pour la mesure du saxophone (gauche) et du cor de basset (droite)	82
Figure 59. Directivité du saxophone mesurée pour un doigté "ouvert" (Si medium, 246 Hz).....	83
Figure 60. Inversion du système MIMO pour reproduction des harmoniques sphériques par la source GSR	84
Figure 61. Mesures de directivité de la source GSR reproduisant les harmoniques sphériques à 1 kHz.....	85
Figure 62. Influence du contrôle de directivité sur le spectre omnidirectionnel de la source GSR	86
Figure 63. Répartition des phases pour la reproduction des harmoniques sphériques avec la GSR.....	86
Figure 64. Expérience de séparation d'un son de hautbois (McAdams et Reynolds, 1983)	88
Figure 65. Dispositif multicanal pour la synthèse Spatio-Operationnelle Spectrale (SOS)	88
Figure 66. Reproduction of an original field with loudspeakers.	90
Figure 67. The reproduction signal paths of two different signals S_1 and S_2	91
Figure 68. Schematic picture of the placement of the piezo-electric bar.	91
Figure 69. The violin in the anechoic room.	92
Figure 70. The relative spectrum of the violin.	93
Figure 71. Some shapes of the directivity of the violin.	94
Figure 72. Spectrum of the coefficients of the violin showing its omnidirectional behaviour for frequencies up to 650Hz.	95
Figure 73. The violin spectrum (black , upper line) and the power of the decomposition in only zeroth and first order harmonics (red, lower line).	96
Figure 74. Schematic representation of signal processing.....	97
Figure 75. Détail des étapes de la <i>Synthèse Additive Spatiale</i>	99
Figure 76. Synthèse du rayonnement (principe).....	100
Figure 77. Synthèse du rayonnement (contrôle de la directivité en fonction de la fréquence).....	101
Figure 78. Spectre de puissance du violon mesuré (noir) et reconstruit (rouge),	102
Figure 79. Violon - dispositif expérimental de mesure	103
Figure 80. Exemple de réponse impulsionnelle (en temporel et en fréquentiel)	103
Figure 81. Décomposition du champ de pression rayonné par le violon ;	104
Figure 82. Procédure générale de <i>Synthèse Additive Spatiale</i>	105
Figure 83. Système simplifié (4 haut-parleurs) et système complet (Timée multi-voies)	106
Figure 84. Etude perceptive du rayonnement : plan d'expérience	110
Figure 85. Tête artificielle pour enregistrements binauraux et position par rapport à la source (Espro).	111
Figure 86. Etude perceptive du rayonnement : premiers résultats	113
Figure 87. Analyse en Composantes Principales (ACP) sur les réponses des sujets	113
Figure 88. ANOVA#1 pour les plaques : effet des facteurs principaux sur la variable dépendante (E).....	114
Figure 89. ANOVA#1 pour les voix : effet des facteurs principaux sur la variable dépendante (E)	114
Figure 90. ANOVA#1 pour les plaques : effet du type de plaques sur la variable dépendante.....	115
Figure 91. ANOVA#1 pour les voix : effet du mode de rayonnement (R) sur le système de diffusion (D) en fonction de la variable dépendante	115
Figure 92. ANOVA#2, pour les voix : effet du facteur complexification du système de diffusion (C) sur les modalités du stimulus (S) en fonction de la variable dépendante.....	116
Figure 93. Théorie de la Gestalt : vase de Rubin (gauche) – image du dalmatien (droite)	122
Figure 94. Hypothèse expérimentale de la psychophysique.....	123
Figure 95. Principales étapes du processus de traitement de l'information.....	124
Figure 96. Résultat des travaux de Helmholtz sur le timbre interprétés par E. Leipp	131
Figure 97. Espace des timbres d'après Grey [Grey, 1977]	132
Figure 98. Espace des timbres d'après McAdams et coll. [McAdams et al., 1995]	132
Figure 99. Taxonomie de Gaver pour la classification des sons environnementaux.....	137
Figure 100. Structure des interactions de la classe des « solides » – extrait de [Houix, 2012]......	137
Figure 101. Représentation graphique d'un arbre ultramétrique.....	141
Figure 102. Exemple d'un profil (portrait) sémantique.....	142
Figure 103. Interface du moteur de recherche par similarité de la base SOL.	143
Figure 104. Spectral Centroid vs RMS level of Bb0, staccato trombone sound	146
Figure 105. Schéma simplifié du système de contrôle numérique monté sur le système tube+anche	151

Figure 106. Dispositif expérimental pour mesure de la fonction de transfert en boucle fermée du système	151
Figure 107. Fonction de transfert en boucle fermée du système avec contrôleur numérique.....	152
Figure 108. Phase en boucle ouverte de la fonction de transfert du système	152
Figure 109. Filtre passe-bande dans Max/MSP utilisé pour le système de contrôle numérique.....	153
Figure 110. FT mesurée avec différents contrôles appliqués à la deuxième résonance de l'instrument. ...	153
Figure 111. Espace des paramètres pour les configurations testées dans notre expérience.	155
Figure 112. MDS INDSCAL de l'ensemble des matrices de dissemblances des 7 sujets.	156
Figure 113. Aspects musicaux (harmoniques, facilité, ouverture) jugés avec variation de gain pour $\phi = 0$	158
Figure 114. Evolution des aspects musicaux lors de l'augmentation du gain, à phase constante $\phi = \pi$. .	159
Figure 115. Effet de la phase pour les configurations à gain maximal, en comparaison, avec la configuration sans contrôle.	159
Figure 116. Schematic representation of the proposed 2-level organisation structure.....	167
Figure 117. Experiment 1, dendrogram resulting from the cluster analysis.	169
Figure 118. Linear regression between dim. 1 and HNR.....	175
Figure 119. Linear regression between dim. 1 and HNR.....	175
Figure 120. Linear regression between dim. 2 and Complex Brightness.....	176
Figure 121. Linear regression between dim. 2 and Complex Brightness.....	176
Figure 122. Linear regression between dim. 3 and Perceptual Spectral Spread	177
Figure 123. Linear regression between dim. 3 and Loudness,.....	177
Figure 124. Linear regression between dim. 1 and Roughness.....	178
Figure 125. Linear regression between dim. 2 and Perceptual Spectral Centroid	179
Figure 126. Linear regression between dim. 3 and Perceptual Spectral Spread	179
Figure 127. Linear regression between dim. 1 and Perceptual Spectral Centroid – <i>impact</i> class / study D.	180
Figure 128. Linear regression between dim. 2 and Cleanness indicator – <i>impact</i> class / study D.....	180
Figure 129. Linear regression between dim. 3 and Sound level.....	181
Figure 130. Tree computed from the categorization data	196
Figure 131. Two-axis preference mapping computed from the "sound quality" projection data.....	197
Figure 132. Frequency response of examples of buzzers (red and blue lines) and loudspeaker (black line)	199
Figure 133. Sound characteristics (in italics) produced by the PWM signal.....	200
Figure 134. The graphical interface of the ATM used in the study.	206
Figure 135. Different steps of the experimental procedure.....	209
Figure 136. Pre-experimental ratings of naturalness (N, step 1), averaged across the 45 participants.	210
Figure 137. Pre-experimental ratings on the three scales for the sounds at the three level of naturalness.	211
Figure 138. Direct comparisons of pre- and post-experimental assessments on 3 scales for the 2 groups	213
Figure 139. Step 5. Ratings on the two scales (Usability, Pleasantness) for the sounds at the two levels of naturalness (High, Medium), and for the normal and the defective mode of operations.	215
Figure 140. Arborecence des menus et principe d'utilisation des 3 librairies (LIB1, 2, 3) de sons.....	219
Figure 141. Temps de réalisation des tâches (moyenne +/- un écart type).	224
Figure 142. Confort perçu (moyenne +/- un écart type).....	225
Figure 143. Détournement de regard perçu (moy. +/- un écart type).....	225
Figure 144. Dispositif expérimental et dimensionnement.....	231
Figure 145. Cible de la tâche primaire (T1) : <i>smiley</i> (gauche) et distracteur (droite).....	231
Figure 146. Interface de navigation (nouvelle version) utilisée pour la tâche secondaire (T2)	232
Figure 147. Schéma de principe du protocole expérimental	233
Figure 148. Temps de réaction de la tâche T1.	234
Figure 149. Temps de navigation au cours de la tâche secondaire (T2)	234
Figure 150. Pourcentage de temps passé dans chaque zone d'attention	235

Liste des annexes

Annexe 1. Rayonnement acoustique du piston plan circulaire	267
Annexe 2. Résolution de l'équation d'onde en coordonnées sphériques	269
Annexe 3. Mesures de la source GSR	271
Annexe 4. Mesures de la directivité du saxophone	273
Annexe 5. Mesures de la directivité du cor de basset	275
Annexe 6. Complementary data and initial results related to the four primary studies	277
<i>Data related to study A1</i>	277
<i>Data related to study A2</i>	278
<i>Data related to study B</i>	279
<i>Data related to study C</i>	280
<i>Data related to study D</i>	281
Annexe 7. Illustration of the experimental graphical user interfaces used in experiments 1 and 2	283
<i>Screenshot of Experiment 1 - GUI</i>	283
<i>Screenshot of Experiment 2 - GUI</i>	283
Annexe 8. Details of acoustic features calculation	285
<i>RMS value</i>	285
<i>Loudness</i>	285
<i>Harmonic emergence feature</i>	285
<i>Spectral centroid</i>	285
<i>Spectral spread</i>	285
<i>Complex Brightness</i>	286
<i>Roughness</i>	286
<i>Cleanness indicator</i>	286

Introduction

Introduction

Ce document présente une sélection structurée de travaux accomplis au cours des 15 dernières années passées dans le Département Recherche et Développement de l'Ircam (Institut de Recherche et Coordination Acoustique / Musique) devenu entre-temps UMR STMS (Sciences et Technologies de la Musique et du Son) Ircam-CNRS-UPMC. Il représente de manière non linéaire mon parcours professionnel de chercheur, effectué au sein de plusieurs équipes de ce laboratoire, et dans différents champs de l'acoustique : acoustique instrumentale, acoustique des salles, psychoacoustique ou perception et cognition musicales. Il représente également ma contribution à un domaine d'application – et de recherche – tout juste émergeant il y a quinze ans et qui s'est largement développé depuis ce temps, notamment à l'Ircam : le design sonore.

La structure globale de ce document est théoriquement construite sur le modèle du chemin acoustique « émission – propagation – réception » et s'articule donc autour de trois grands axes : la **synthèse**, la **reproduction** et la **perception** des sons, en couvrant un corpus d'étude très large allant des sons musicaux aux sons environnementaux. L'aspect volontairement ouvert de cette thématique est motivé, d'une part, par la réalité de mes travaux, et d'autre part, par les nombreuses similitudes ou interactions existant entre ces différents objets de recherche. Enfin, dans une quatrième partie, le document aborde la question de la recherche en **design sonore** par l'intermédiaire de plusieurs études illustrant, chacune à leur manière, les articulations possibles entre la connaissance scientifique ou technologique dans le domaine du son (musical ou environnemental), et la création sonore appliquée.

Le chapitre relatif à la synthèse expose la problématique générale en présentant un aperçu des grandes approches de synthèse sonore numérique, en détaillant ensuite plus précisément l'une d'entre elles, la synthèse par modélisation physique, et enfin en se focalisant sur l'un des formalismes de modélisation : le formalisme modal. Ces propos liminaires permettent d'introduire le cadre dans lequel s'est effectué mes travaux sur la modélisation de l'accordéon, à savoir le moteur de synthèse sonore *Modalys*. L'architecture et la syntaxe du logiciel sont explicitées et mises en regard du travail de recherche et de développement sur le fonctionnement de l'anche libre d'accordéon. Les premiers résultats obtenus y sont ensuite présentés ainsi que l'extension de l'étude qui a principalement consisté à améliorer l'expérience de visualisation et de mesure de l'écoulement en amont et en aval de l'anche ainsi que la détermination de l'effet d'inertie du fluide, principal responsable de son mouvement.

Le chapitre relatif à la reproduction part d'un bref historique sur les techniques de restitution – et d'enregistrement – du son, présente ensuite l'élément-clé de cette problématique, à savoir le haut-parleur et les différentes techniques qui permettent de recréer une image sonore cohérente aux oreilles d'un auditeur, pour arriver enfin à la question centrale qui a alimenté un certain nombre de mes travaux : la directivité des sources et son contrôle par des moyens électroacoustiques. Des éléments théoriques concernant le rayonnement acoustique et la représentation du champ sonore ainsi qu'un état de l'art sur la directivité des instruments de musique et les technologies associées, permettent de poser le cadre d'un projet mené pendant plusieurs années et qui a abouti à des résultats tangibles autant sur le plan scientifique

qu'artistique : la mesure de la directivité et son contrôle par un réseau de haut-parleurs dit "sphérique" – et baptisé *Timée* dans ses applications musicales, par le compositeur en recherche initialement associé au projet (François Nicolas).

Parti d'une réflexion assez large et relativement générique sur la question, cet axe d'étude a ensuite opéré un double mouvement de développement : une focalisation sur l'étude et la reproduction du rayonnement d'instruments donnés (saxophone, cor de basset), et une extension vers la mise en oeuvre d'un nouveau paradigme de synthèse associant un moteur de synthèse sonore (*Modalys*) et un moteur de synthèse de directivité (*Timée*), conférant ainsi aux sons de synthèse des propriétés spatiales propres à la source modélisée. Cette approche, appelée *Synthèse Additive Spatiale*, a été instanciée sur plusieurs modèles (violon, plaque frappée) ; elle a, en outre, fait l'objet d'une première tentative d'évaluation perceptive dans le cadre plus général d'une étude sur la perception du rayonnement acoustique.

Le chapitre relatif à la perception propose, en premier lieu, de replacer cette notion par rapport aux champs respectifs de la cognition et, plus largement, des sciences cognitives. Afin de fixer le cadre théorique des travaux présentés par la suite, les principaux fondements de la psychologie (behaviorisme, théorie de la Gestalt, psychophysique, psychologie cognitive) sont ensuite passés en revue sous l'angle – autant que possible – de la modalité auditive. Dans la même optique, la question de la représentation cognitive du monde, et notamment, la notion de catégories est également abordée par le biais des trois grandes théories relatives aux mécanismes de catégorisation (théories classique, prototypique et théorie des exemplaires).

Cela étant – et toujours dans l'idée de donner un cadre formel à mes travaux dans le domaine –, la notion de perception sonore est ensuite plus spécifiquement considérée, d'une part, en détaillant les études sur le timbre musical ainsi que sur l'identification des sources sonores en général, et d'autre part, en présentant les principaux paradigmes expérimentaux utilisés pour étudier les mécanismes de perception sonore.

Enfin, une série de travaux personnels relatifs à cette thématique est présentée et alimente certaines des grandes questions soulevées dans les sections précédentes : le timbre des sons musicaux, avec le développement d'un moteur de recherche fonctionnant sur un principe de similarité perceptive ; la perception de la qualité d'un instrument de musique impliquant la boucle de rétroaction musicien-instrument (expérience en contexte de jeu) et, pour l'une d'entre elles, explorant les possibilités innovantes de la technologie du contrôle actif ; et enfin, une méta-analyse du timbre d'un corpus élargi de sons environnementaux qui met en évidence une structure à la fois catégorielle et continue décrite et modélisée par des descripteurs acoustiques calculés sur le signal sonore.

Le chapitre relatif aux applications en design sonore s'ouvre par une réflexion autour de cette discipline – émergente il y a vingt ans mais dorénavant bien établie – et la manière dont elle tente de se définir en théorie, notamment avec les concepts posés par R. M. Schafer, ou en pratique. En replaçant le design sonore dans le champ plus large, mais néanmoins naturel, du design et en considérant certaines de ses composantes étymologiques et historiques, des éléments de définition sont ensuite proposés et aboutissent à considérer le design sonore comme un art sonore appliqué mettant en oeuvre un processus de création interdisciplinaire, contrainte et ciblée, et s'appliquant sur des terrains de natures relativement diverses (industrie, architecture, urbanisme, multimédia, marketing, cinéma, etc.).

Ma contribution à cette discipline en termes de recherche et d'application s'illustre ici par trois études qui s'inscrivent globalement dans la thématique des Interfaces Homme-Machine (IHM). La première concerne le signal sonore du clignotant et s'intéresse à la manière dont le design sonore peut apporter des réponses nouvelles, c'est-à-dire inouïes, à cette signalétique pourtant très balisée et normalisée ; elle met en évidence également l'impact d'une conception sonore innovante sur les avancées de la technologie en termes de diffusion audio à l'intérieur de l'habitacle automobile. La seconde étude explore l'influence d'une propriété sonore – en l'occurrence le caractère naturel – sur l'ergonomie et l'agrément d'utilisation d'une interface de type clavier numérique sonifiée, c'est-à-dire munie d'un retour sonore ; l'hypothèse globale posée étant que les sons naturels favorisent l'utilisation d'un tel système. Enfin, la troisième étude se place également dans le contexte de l'habitacle automobile et propose différentes approches de sonification d'un processus de navigation dans la structure hiérarchique d'un système multimédia ; ces approches sont construites en combinant des principes de base avec des techniques de synthèse sonore par modélisation physique et sont évaluées subjectivement en termes fonctionnels et esthétiques. Une extension de ce travail propose, quant à elle, de répondre à la question fonctionnelle – notamment par rapport à l'attention nécessaire lors de la tâche de conduite automobile – en utilisant un procédé de mesure oculométrique qui permet de repérer de manière objective la position et le mouvement du regard du conducteur.

Par ailleurs, la plupart des travaux qui composent ce document ayant fait l'objet d'articles ou de communications scientifiques – publiés dans des revues ou des actes de congrès –, je me suis permis, avec l'accord du directeur de mon école doctorale (ED Abbé Grégoire) et de mon directeur de thèse, d'insérer *in extenso* ces **publications** dans le corps du texte – donc, pour certaines, en anglais –, lorsque cela me semblait pertinent pour le développement des idées et moyennant un préambule de remise en contexte. D'un point de vue typographique, ces parties sont repérées par un titre commun surligné : **Publication**, ainsi que par une police de caractère spécifique : corps de l'article publié, qui permettent de les identifier facilement au cours de la lecture.

Première partie

Synthèse

PARTIE I. Synthèse

I.1 Problématique

La synthèse sonore est un procédé technique permettant de produire un signal sonore à partir d'une combinaison d'éléments de base. Ces éléments peuvent être soit de nature matérielle (*hardware*) soit de nature logicielle (*software*) et distinguent une synthèse dite **analogique**, d'une synthèse dite **numérique** (*digital synthesis*).

D'un point de vue organologique, il est vraisemblable de faire coïncider les débuts de la synthèse sonore avec l'expansion de l'électricité puis de l'électronique, et l'apparition d'une nouvelle famille instrumentale : les instruments électro-mécaniques puis électriques ou électroniques et enfin numériques.

Le premier instrument de ce type – pouvant, de fait, être considéré comme le premier spécimen d'instrument de synthèse – est le Telharmonium¹, construit par Taddeus Cahill, en 1895. Cet instrument, entièrement mécanique et électrique, met en œuvre le principe de la roue phonique constituée d'une roue crantée en rotation devant une bobine électrique produisant ainsi un signal dont la fréquence dépend de la vitesse de rotation et du nombre de crans de la roue. Ce principe sera repris plus tard notamment dans la conception et le développement des orgues Hammond.

Les instruments électro-analogiques – électriques ou électroniques –, quant à eux, utilisent plusieurs circuits électriques pour produire des sons puis les modifier, les moduler et les amplifier ; on commence alors à parler de **synthétiseurs analogiques**. Dans ce cas, les oscillateurs physiques (mécaniques ou acoustiques) sont d'abord remplacés par des composants électriques (lampes, transistors, etc.), puis par des circuits électroniques.

Structurellement, les synthétiseurs analogiques sont modulaires, composés de blocs reliés entre eux selon un schéma de câblage (externe ou intégré) appelé *patch*. Les blocs correspondent à des fonctions primaires de la synthèse, principalement :

- les oscillateurs, ou générateurs de signal (sinusoïde, dent de scie, carré, triangle, bruit blanc ou rose, etc.), assurés par des unités de type VCO (*Voltage-Controlled Oscillator*) ;
- les filtres (passe-bas, passe-haut, passe-bande ou coupe-bande), assurés par des unités de type VCF (*Voltage-Controlled Filter*) ;
- les générateurs d'enveloppe ADSR (*Attack – Decay – Sustain – Release*) ;
- l'amplificateur assuré par une unité de type VCA (*Voltage-Controlled Amplifier*) ;
- les modulateurs, essentiellement basse fréquence, assurés par des unités de type LFO (*Low-Frequency Oscillator*) mais pouvant être également d'autres types (glissement, mixage, inversion, réduction, modulation en anneau).

La **synthèse numérique** apparaît, quant à elle, avec l'essor de l'informatique et de l'audio-numérique, c'est-à-dire avec la conversion du signal analogique en signal numérique. Cette transition, fondée sur le principe de l'échantillonnage et ses technologies associées (convertisseurs), marque également le passage d'une

¹ <http://sonhors.free.fr/>

² <http://www.csounds.com>

représentation continue des signaux vers un monde discret – échantillonné –, construit sur le système de description binaire et son unité, le *bit*.

Historiquement, les premières expériences de synthèse numérique sont celles effectuées en 1957 par Max Mathews et son équipe (entre autres, H. McDonald, J. R. Pierce, N. Guttman) du Bell Telephone Laboratories. C'est à cette date, que le premier programme informatique produisant du son est créé. Ce programme (Music I) est le début d'une série d'implémentations (Music II, III, V, ... N) qui définissent les prémisses de la synthèse sonore par ordinateur. A l'occasion de l'une de ces étapes (Music III, 1960), M. Mathews – et son collègue J. Miller – établissent le concept fondamental de **générateur élémentaire** (*Unit Generator*) qui représente, d'après Curtis Roads, l'un des développements les plus significatifs dans la conception des langages de synthèse sonore numérique [Roads, 1996].

Par définition, un générateur élémentaire est un module de génération (oscillateur, ...) ou de modification (filtre, ...) du signal qui, assemblé à d'autres dans une structure appelée également *patch*, constitue un instrument de synthèse (voir figure 1). On notera ici, la similitude de forme et de terminologie avec la synthèse analogique décrite précédemment, sachant que, dans ce cas précis, les connections et les constituants sont tous instanciés sous forme logicielle. Ce formalisme permet, en outre, de décrire la plupart des architectures de synthèse.

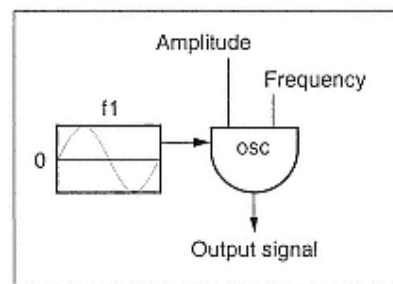


Figure 1. Illustration d'un générateur élémentaire de type oscillateur à lecture de table [Roads, 1996]. La forme d'onde est stockée dans une mémoire et lue en fonction des paramètres (amplitude, fréquence).

Cela étant, dans le domaine de la synthèse sonore numérique, il est d'usage de distinguer deux grandes approches qui se différencient notamment par les moyens mis en œuvre pour produire les sons :

- l'approche de type **signal** consiste à s'intéresser à la structure du signal, ses propriétés morphologiques à la fois du point de vue temporel et fréquentiel. Cette démarche est issue de l'approche analogique historique et s'appuie sur des techniques d'analyse de type transformée de Fourier qui permettent de détailler la composition spectrale du signal ;
- l'approche de type **modélisation physique** consiste à s'intéresser à la structure du mécanisme producteur du son, la physique de ses composants, à la fois du point de vue mécanique et acoustique. Cette approche est issue des avancées de la science et de la technologie (informatique) et s'appuie généralement sur des démarches expérimentales permettant de comprendre le fonctionnement des systèmes.

L'exemple que l'on prend habituellement pour illustrer cette différence de "philosophie" est celui de la clarinette. Elle pourra, en effet, être modélisée de deux manières différentes : i/ soit selon ses caractéristiques spectrales, prédominance des harmoniques impaires par rapport aux harmoniques paires (approche signal) ; ii/ soit

par son mode de fonctionnement, couplage non-linéaire entre une anche excitatrice et un tube cylindrique résonateur, alimenté par une pression d'alimentation (approche modélisation physique).

Par ailleurs, une autre typologie des nombreuses techniques de synthèse sonore numérique peut également être proposée sur une base plus précise des différents algorithmes mis en jeu :

- synthèse par **lecture d'échantillons**, partant du principe du stockage d'un élément de signal dans une mémoire sur laquelle est effectuée une opération de lecture paramétrique (vitesse, incrément). Elle regroupe la synthèse granulaire, la synthèse par lecture de table et l'échantillonnage – énumérées ici dans l'ordre croissant de la taille de mémoire considérée.
 - la synthèse par lecture de table est la manière la plus ancienne et la plus simple de générer du son au moyen d'un ordinateur (cf. "Music I" de M. Mathews). La forme d'onde stockée en mémoire, sur quelques périodes, est généralement canonique (sinus, carrée, triangle, etc.) mais peut aussi être quelconque.
 - la synthèse par échantillonnage peut être considérée comme une technique de lecture de table à plus grande échelle. Un ensemble d'échantillons correspondant aux possibilités mélodiques et expressives (nuances) d'un instrument est stocké puis recombinaison par le synthétiseur en fonction des paramètres de jeu.
 - la synthèse par granulation (ou granulaire) peut aussi être considérée comme une technique de lecture de table mais à micro-échelle. Elle est basée sur la production d'une grande quantité d'événements sonores (les grains) d'une durée variant entre quelques dizaines et quelques centaines de millisecondes. La notion de grain sonore est à relier à l'idée de « quantum de son » développée par le physicien Dennis Gabor (1946).
- synthèse par **algorithmes abstraits**, partant du principe d'une série d'opérations formelles sur des composants unitaires (le plus souvent, des générateurs élémentaires). Elle regroupe principalement les synthèses par modulation – en anneau, d'amplitude, de fréquence ou de largeur d'impulsion – mais aussi, la distorsion non linéaire ou les approches de type procédurale.
 - la synthèse par modulation d'amplitude – respectivement, en anneau – consiste à moduler l'amplitude d'un signal (onde porteuse) par un autre signal (onde modulante), le signal modulant étant unipolaire – resp., bipolaire.
 - La synthèse par modulation de fréquence – respectivement, de phase – consiste à moduler la fréquence – resp., la phase – d'un signal (onde porteuse) par un autre signal (onde modulante).
 - la synthèse par modulation de largeur d'impulsion (PWM – *Pulse Width Modulation*) consiste à moduler les paramètres d'un train d'impulsions rectangulaires (largeur / amplitude de l'impulsion, intervalle inter-impulsions relié à la fréquence fondamentale).
- synthèses par **modèles de signaux**, partant du principe d'une construction du signal à partir d'une série d'opérations simples (addition, soustraction) sur des composants unitaires. Elle regroupe essentiellement la synthèse additive et soustractive.

- la synthèse additive consiste à additionner des formes d'onde élémentaires pour créer une forme d'onde complexe. Les formes élémentaires étant le plus souvent des générateurs élémentaires de type oscillateurs sinusoïdaux.
- la synthèse soustractive est le procédé quasiment inverse de la synthèse additive ; elle consiste à retrancher des zones d'énergie à partir du spectre large bande d'un bruit (blanc, rose, ...) ou d'un train d'impulsions. C'est une approche que l'on peut qualifier de source/filtre, les composants mis en jeu étant ici des "générateurs élémentaires" de type filtre.
- synthèse par **modèles physiques**, partant du principe – comme mentionné plus haut – que les propriétés sonores visées sont issues d'une source physique dont on cherche à modéliser le mécanisme. Elle regroupe différents types de synthèses : MSW, Karplus-Strong, guide d'onde ou Hiller-Ruiz, Cordis Anima et synthèse modale (voir § I.2).

I.2 Synthèse par modélisation physique

La synthèse par modélisation physique repose fondamentalement sur la connaissance théorique de la physique des systèmes mécaniques, ou acoustiques, producteurs de son. Elle s'est intéressée, en premier lieu, au cas des instruments de musique grâce, notamment, aux travaux de recherche de Helmholtz [Helmholtz, 1860], Lord Rayleigh [Rayleigh, 1945] Pierce [Pierce, 1992], Olson [Olson, 1952], Fletcher et Rossing [Fletcher et al., 1998] ou Benade [Benade, 1990], pour n'en citer que quelques-uns. Cependant, nécessitant d'importantes ressources de calcul, cette technique n'a réellement pris son essor que dans les années 80 avec le développement de l'informatique et la constante progression des puissances de traitement numérique des données. Avec la généralisation du concept de calcul distribué (*grid computing*), elle est aujourd'hui tout à fait opérationnelle et peut rivaliser avec les autres techniques de synthèse – moins lourdes en nombres d'opérations mathématiques par seconde (*flops*) – notamment dans le domaine de l'implémentation temps-réel.

A ce titre, la synthèse par modélisation physique présente des avantages et des inconvénients. Ses avantages résident essentiellement dans le fait que le modèle physique possède, par nature, les caractéristiques intrinsèques de la source réelle ce qui a pour conséquence de produire des sons et des comportements généralement plus réalistes que dans une synthèse de type signal. Par exemple, le système tube cylindrique + anche simple du cas précédent sur la clarinette contient naturellement la spécificité spectrale de cette famille d'instruments (prépondérance des harmoniques impaires sur les harmoniques paires). De même, les phénomènes transitoires ou le saut de palier d'un cuivre, sont inclus dans l'évolution du modèle en fonction de ses paramètres d'excitation (notamment, augmentation de la pression d'alimentation).

En outre, la nature physique et intuitive des paramètres de contrôle peut également être vue comme un avantage par rapport à une approche signal, lorsqu'il s'agira, par exemple, de définir la synthèse en termes de grandeurs géométriques (longueur, diamètre, position, etc.), structuraux (densité, module d'Young, etc.) ou énergétiques (valeur d'une force, etc.) plutôt qu'en termes de fréquence / amplitude d'un oscillateur ou facteur de qualité d'un filtre.

Cela étant, cette dernière propriété de la modélisation physique peut aussi s'avérer être un obstacle à la composition musicale, la relation entre paramètres physiques du

modèles et propriétés du son qu'il produit n'étant pas directe et pouvant aboutir à des problèmes d'inférence et de prédictibilité du résultat. Ce dernier point, associé à un relatif problème de puissance de calcul – tendant à devenir négligeable au fil des décennies, comme mentionné plus haut – constitue, a priori, les seuls inconvénients majeurs d'une approche de synthèse par modélisation physique.

En termes de perspective, la synthèse par modélisation physique peut être envisagée à la fois du point de vue scientifique mais aussi artistique.

En effet, le premier objectif affirmé de cette technique est la simulation au plus proche des phénomènes physiques régissant les instruments de musique – ou autres mécanismes sonores – afin de mieux en comprendre le fonctionnement (« [...] *the closer the simulation, the better understood the system is* » [Roads, 1996]).

Cependant, les potentialités d'une telle approche peuvent facilement être étendues au domaine artistique, dans la mesure où elle permet, a priori – et dans une certaine mesure –, la création de systèmes physiquement impossibles, produisant des sons potentiellement inouïs, et permettant donc à la pensée ou à l'esprit créatif d'un compositeur de s'émanciper et de visiter des territoires musicaux et sonores jusqu'alors inexplorés. Dans la limite des lois de la physique – c'est-à-dire notamment, avec l'énergie nécessaire pour mettre en vibration de telles configurations irréalistes –, il est donc a priori possible de faire sonner une corde en bois ou en verre, de jouer d'une membrane de plusieurs dizaines de mètres de diamètre ou d'une clarinette de plusieurs mètres de long, et mieux encore, de construire des instruments hybrides quasi "extra-terrestres" comme, par exemple, celui imaginé par M. D. Bonnet dans sa thèse : un gong traversé par une corde dont la position varie au cours du temps [Bonnet, 2002].

Par ailleurs, de manière plus détournée – et avec toutes les approximations inhérentes à cette application –, elle peut aussi avoir un intérêt historique et muséographique dans la mesure où elle peut permettre de recréer le fonctionnement et le rendu sonore d'instruments anciens, voire disparus, à partir de relevés géométriques ou de plans de fabrication.

En termes de paradigme, la synthèse par modélisation physique est basée sur deux notions fortes, issues d'une représentation schématique du monde des instruments de musique : l'**excitation** et la **résonance**.

Dans le formalisme général, l'excitation comprend à la fois l'élément excitateur (par exemple, l'anche de clarinette ou l'archet du violon) ainsi que la fonction de couplage avec le résonateur (par exemple, la fonction collé/glissé – *stick/slip* –, dans le cas de la corde de violon frottée par l'archet); elle est généralement de nature non-linéaire. La résonance comprend, quant à elle, simplement l'élément résonateur auquel est couplée l'excitation (par exemple, le tube de clarinette ou la corde du violon – ou plus précisément, l'ensemble corde+chevalet+caisse du violon); le résonateur est le plus souvent régi par des lois linéaires, sauf dans le cas spécifique de modélisation de phénomènes non-linéaires (par exemple, le cuivrage des instruments à vent de type cuivre [Msallam et al., 2000]).

Nous verrons un peu plus tard que certains formalismes – en particulier, le formalisme modal – différencient l'élément excitateur de la fonction excitatrice et considèrent donc le modèle de base formé de trois éléments : l'excitateur, le résonateur tous deux décrits de manière linéaire, et l'interaction qui renferme les non-linéarités du système.

D'un point de vue méthodologique, la synthèse par modélisation physique peut être vue comme le résultat d'un processus comportant plusieurs étapes : comprendre la physique de l'instrument au moyen d'expériences d'observation et de mesure (analyse), formaliser les principes fondamentaux de son fonctionnement au moyen d'un ensemble d'équations mathématiques décrivant la dynamique du problème (modélisation), exprimer le système d'équations en choisissant des méthodes numériques capables d'approximer les opérateurs mathématiques mis en jeu (discrétisation) puis résoudre le système de manière incrémentale afin d'obtenir la valeur du signal à chaque pas d'échantillonnage (synthèse).

Sur cette base, au cours de l'histoire de la synthèse par modélisation physique, différentes approches ont été adoptées et ont donné naissance à plusieurs travaux et méthodes de synthèse. En reprenant la dualité entre « Ondes et Modes » exprimée par Chaigne et Kergomard dans leur récent ouvrage sur la physique des instruments de musique, [Chaigne et al., 2008], celles-ci peuvent cependant être classées en deux grandes catégories (voir [Roads, 1996 ; Russ, 2008] pour un développement complet de ces différentes méthodes) :

- d'un point de vue « Ondes » – c'est-à-dire en décomposant les phénomènes acoustiques en réflexions successives des ondes –, les travaux de McIntyre, Schumacher et Woodhouse (et la synthèse MSW associée) qui correspond à un modèle en ondes progressives incluses dans une boucle de rétroaction [McIntyre et al., 1983], puis les modèles simplifiés de Karplus-Strong [Karplus et al., 1983] et enfin leur généralisation à la synthèse par guides d'ondes [Smith, 1992].
- d'un point de vue « Modes » – c'est-à-dire en considérant les phénomènes acoustiques du point de vue modal – les travaux initiaux de Hiller et Ruiz [Hiller et al., 1971] qui modélisent le comportement dynamique de la structure en la décomposant en systèmes élémentaires (masse-ressort-amortisseur) puis leur application dans l'environnement Cordis Anima [Cadoz, 1984] et enfin l'alternative constituée par la synthèse modale [Adrien, 1988 ; 1991].

I.3 Le formalisme modal

La synthèse modale se présente donc comme une alternative au paradigme masse-ressort-amortisseur. Elle part fondamentalement de deux principes :

- la **modularité** qui postule qu'un objet vibrant produisant du son peut être représenté par un ensemble de sous-structures dont le nombre est, par ailleurs, beaucoup moins important que dans le cas de cellules masse-ressort-amortisseur.
- la **superposition** qui conçoit le mouvement instantané d'une structure comme la somme de mouvements élémentaires, appelés modes.

Ainsi, chaque sous-structure possède des modes de vibration mis en jeu lors de l'excitation, spécifiques à la structure et dépendant de ses caractéristiques propres. Les données modales se présentent sous un même formalisme comprenant :

- les fréquences (d'oscillation),
- les coefficients d'amortissement (de l'énergie),
- les déformées (de la structure).

Ces informations sont, a priori, suffisantes pour simuler le comportement statique ou dynamique d'une structure, lorsque celle-ci est excitée par une force extérieure. Par ailleurs, d'un point de vue uniquement spectral, les fréquences modales peuvent être

apparentées aux résonances de la structure et, de ce fait, sont assimilables aux fréquences des partiels obtenues par mesure ou traitement du signal (modèle de résonances) ; le coefficient d'amortissement d'un mode (σ) étant directement lié au facteur de qualité à -3 dB (Q) de la résonance correspondante par la relation suivante :

$$\sigma = \omega / 2Q \quad (\text{Eq. 1})$$

où ω désigne la pulsation du mode. Cette analogie permet, dans certaines implémentations, d'établir un pont entre le formalisme modal et le domaine signal.

Fondamentalement, la description modale d'une structure continue conduit, en premier lieu, à une description similaire à celle du paradigme masse-ressort-amortisseur. En effet, toute structure physique peut être considérée comme un système à plusieurs degrés de liberté, le nombre de degrés de liberté étant fonction de la résolution spatiale de la discrétisation. Dans un cas idéal non dissipatif, le mouvement peut alors être décrit par l'équation générique suivante :

$$[M] \cdot \{\ddot{y}\} + [K] \cdot \{y\} = \{f^{ext}\} \quad (\text{Eq. 2})$$

où $[M]$, $[K]$ sont respectivement les matrices de masse et de raideur, $\{y\}$ le déplacement des nœuds du maillage de la structure et $\{f^{ext}\}$ le vecteur de forces extérieures appliquées à chacun d'eux.

La détermination des modes propres de la structure nécessitent, dans un premier temps, de considérer les vibrations libres du systèmes donc avec $\{f^{ext}\} = \{0\}$.

L'hypothèse est ensuite faite d'une solution de type sinusoïdale $\{y\} = \{y_0\} \cdot e^{i\omega t}$, ce qui conduit à l'équation :

$$([K] - \omega^2 [M]) \cdot \{y\} = 0 \quad (\text{Eq. 3})$$

Le modèle spatial ($[M]$, $[K]$) est converti en modèle modal ($\{\omega\}, \{\psi\}$) en considérant les solutions du système complet :

$$\begin{cases} \det([K] - \omega^2 [M]) = 0 \\ ([K] - \omega_i^2 [M]) \cdot \{\psi^i\} = 0 \end{cases} \quad (\text{Eq. 4})$$

qui donnent, à une constante multiplicative près, le résultat pour le $i^{\text{ème}}$ mode sous la forme ($\{\omega_i\}, \{\psi^i\}$), respectivement pulsation et coordonnées du vecteur propre, c'est-à-dire, valeurs des déformations en chaque point du maillage de la structure.

On peut alors démontrer que l'ensemble des vecteurs propres $\{\psi\}$ forme une base orthogonale dans laquelle les solutions de l'Eq. 3 peuvent être décomposées de la manière suivante :

$$\{y\} = [\Psi] \cdot \{\psi\} \Leftrightarrow \begin{cases} y_1 = \Psi_1^1 \cdot \psi^1 + \dots + \Psi_1^N \cdot \psi^N \\ \dots \\ y_N = \Psi_N^1 \cdot \psi^1 + \dots + \Psi_N^N \cdot \psi^N \end{cases} \quad (\text{Eq. 5})$$

Ce qui met en évidence la propriété de superposition, à savoir que le mouvement du $k^{\text{ième}}$ point de la structure est le résultat de la contribution des N modes considérés dans la description modale de départ.

Quant à la dernière donnée modale – le coefficient d'amortissement –, il n'apparaît uniquement que si l'on considère un cas de figure dissipatif, c'est-à-dire en considérant les pertes d'énergie ayant lieu au cours du mouvement. Dans ce cas, le mouvement de la structure s'écrit à l'aide de l'Eq. 2 modifiée comme suit :

$$[M] \cdot \{\ddot{y}\} + [B] \cdot \{\dot{y}\} + [K] \cdot \{y\} = \{f^{ext}\}, \quad (\text{Eq. 6})$$

où $[B]$ représente la matrice d'amortissement.

De plus, en première approximation, dans le cas d'un modèle simplifié, l'amortissement est considéré proportionnel, c'est-à-dire que la matrice d'amortissement $[B]$ est une combinaison linéaire des matrices de masse $[M]$ et de raideur $[K]$:

$$[B] = a \cdot [M] + b \cdot [K] \quad (\text{Eq. 7})$$

Ainsi modélisé, l'amortissement présente l'avantage non négligeable de préserver la valeur des fréquences et déformées modales, comme si la structure était idéale, c'est-à-dire sans pertes.; l'effet de l'amortissement se répercutant uniquement dans l'expression de la décroissance des modes au cours du temps.

En pratique, les données modales peuvent être obtenues de trois manières différentes, selon la nature de l'objet à modéliser :

- calculées numériquement à partir de la solution analytique du problème, si la structure est simple et homogène (par exemple, une corde idéale, fixée aux deux extrémités) ;
- mesurées expérimentalement par des techniques d'analyse modale, si la structure est complexe et/ou inhomogène (par exemple, une table d'harmonie de piano) ;
- estimées numériquement à partir d'un code de calcul par éléments finis, si la structure est inexistante et/ou virtuelle.

I.4 Description de Modalys

Modalys est un logiciel de synthèse sonore par modélisation physique dont le principe de fonctionnement repose sur le formalisme modal [Adrien, 1988 ; Eckel et al., 1995 ; Vergez et al., 2000; Bensoam et al., 2001]. Comme explicité au paragraphe précédent, ce formalisme permet de décrire le comportement vibratoire d'une structure au moyen de données modales (fréquences, amortissements et déformées) dépendant de ses caractéristiques physiques : géométrie, matériau, etc. D'un point de vue formel, Modalys reprend également le paradigme organologique de l'instrument excitation/résonance en différenciant néanmoins dans l'excitation, la source excitatrice et l'interaction.

Ainsi, une bibliothèque d'objets excitateurs, résonateurs et interactions permet à Modalys de fonctionner comme un environnement modulaire en offrant des possibilités d'assemblage entre sous-structures élémentaires. Par nature, Modalys possède donc toutes les caractéristiques de l'approche de la synthèse sonore par modélisation physique (contrôle intuitif, comportements intrinsèques, ...) auxquelles s'ajoutent les

avantages liés à l'approche modale, notamment en terme de construction et d'hybridation des structures ; il en possède également les inconvénients dus aux hypothèses de linéarité et au coût de calcul qui, dans certains cas – et malgré l'augmentation progressive des puissances de calcul – reste encore élevé.

1.4.1 Modalys : Éléments

Modalys est souvent comparé à un atelier de lutherie virtuelle ; il permet de disposer d'un ensemble potentiellement infini de corps physiques, les **objets**, à partir desquels un instrument de musique – ou tout autre corps sonore – peut être construit. L'objet est défini par ses propriétés physiques : type de géométrie (tube, corde, plaque, etc.), dimensions spatiales (longueur, rayon, etc.) et propriétés du matériau (densité, modules d'élasticité, etc.). Une fois définis pour une géométrie donnée, ces paramètres physiques sont transcrits en données modales (voir § 1.3). Il est important de noter que, à ce stade, la fonction de l'objet (excitation ou résonance) n'est pas prise en compte dans sa définition et que son comportement est considéré linéaire ; les non-linéarités étant localisées dans la description de la connexion (cf. ci-dessous).

De fait, Modalys exploite la spécificité du formalisme modal qui impose une description uniforme des données modales, quelque soit la nature de la structure. Un tube acoustique, une corde ou une membrane auront le même schéma de données – à la dimensionnalité de la géométrie près – et pourront donc être théoriquement connectés ensemble pour former une structure complexe (par exemple, une corde tendue sur un chevalet fixé à une table d'harmonie de violon), ou bien une combinaison exciteuse-résonateur a priori incongrue (p. ex., fixer une anche simple sur une corde).

Le deuxième élément constitutif de l'environnement Modalys s'exprime en termes de **connections** (ou interactions) : l'instrument n'est pas seulement défini par la nature des structures qui le composent mais aussi par la manière dont ces structures sont assemblées entre elles. Les connections peuvent être vues comme des liens physiques qui mettent les structures en interaction mutuelle, exprimant ainsi la manière dont elles échangent leur énergie au cours du temps. Les interactions physiques peuvent être de nature très différentes (frapper, frotter, glisser, fixer, etc.) et, dans le formalisme propre à Modalys, concentrent toutes les non-linéarités de la description du modèle (voir, par exemple, la fonction collé/glissé de l'archet de violon sur la corde – figure 2).



Figure 2. Illustration de la connexion (interaction) 'frotter' dans Modalys.
 f_n et f_t sont resp. les forces normale et tangentielle au point de contact ;
 v_1 et v_2 les vitesses respectives des deux objets en contact

Cela étant, deux autres typologies d'éléments font également partie de la nomenclature de Modalys : les accès et les contrôleurs. Les **accès** correspondent en quelque sorte à des points d'interface qui permettent de relier les objets (ou les contrôleurs) entre eux et rendent ainsi possible les échanges d'énergie – soit en interne

entre les objets, soit entre un objet et une source d'énergie externe. Les **contrôleurs**, quant à eux, peuvent être vus comme des générateurs de valeurs qui permettent de faire varier les éléments du modèle au cours du temps (p. ex., l'amplitude d'une force externe, la position d'un accès, etc.). Différents types de contrôleurs peuvent ainsi être déclarés (constante, impulsion, enveloppe, etc.), voire implémentés au moyen de fonctions mathématiques, ou bien encore extraits – en temps-réel ou hors temps-réel – d'un capteur externe (p. ex., dans le cas d'une performance gestuelle).

De fait, la première fonction du contrôleur est d'ordre musical car elle permet de spécifier l'évolution des variables physiques externes correspondant à l'énergie apportée à un modèle et donc de piloter la synthèse sonore en simulant le jeu de l'instrumentiste.

Ce dernier point pose ouvertement la question du **geste** instrumental et de la manière dont il peut être modélisé dans une démarche globale de synthèse par modélisation physique.

En effet, à la différence du monde réel où la distinction est clairement faite entre instrument et instrumentiste – et plus généralement, source et agent, ou bien encore objet et action –, dans le monde virtuel de la synthèse, prendre en compte le corps de l'instrumentiste (doigts, lèvres, etc.) signifie le modéliser en tant que corps physique interagissant avec l'instrument. Par exemple, le fait de créer un violon par modélisation physique signifie d'assembler ses éléments constitutifs (cordes, touche, chevalet, table, dos) et de le coupler avec un modèle d'évolution des doigts du violoniste. Il en résulte un contrôle du modèle à plusieurs niveaux : une description expressive de la performance – ou données gestuelles de premier plan – (type et nature du frottement, du vibrato, etc.) mais aussi une description donnant accès à toutes les informations non-expressives du geste – ou données gestuelles d'arrière-plan – (type de doigté, de corde sélectionnée, etc.).

Définir un contrôle pertinent dans Modalys – et par extension dans tout autre moteur de synthèse par modélisation physique – nécessite donc de traiter en même temps ces différents niveaux de description, explicites et implicites (voir, pour illustration de ces propos, l'extrait d'une pièce pour violon conçu avec Modalys par Richard Dudas).

1.4.2 Modalys : Architecture

En résumé, les structures fondamentales de Modalys sont : les objets, les connections et les contrôleurs. Cette partie présente la manière dont ces trois structures sont implémentées et dont le signal sonore est finalement calculé.

Le principe opératoire de Modalys est un calcul échantillon par échantillon. A chaque échantillon, la liste des objets instanciés est parcourue et les algorithmes de synthèse modale sont exécutés pour chacun d'eux afin d'obtenir, ou de réévaluer, l'état vibratoire courant de la structure complète. A partir de cette information, l'ensemble des connections instanciées est à son tour parcouru pour examiner si cet état vibratoire implique l'activation – ou la désactivation – d'une interaction. (par exemple, si la distance entre deux objets est nulle). Dans ce cas, les équations d'interaction sont résolues pour obtenir les forces qui seront imprimées aux objets lors du prochain pas de temps (échantillon). Malgré un nombre de calculs potentiellement élevé (pour des structures complexes, par exemple), l'approche par échantillon est, a priori compatible avec un calcul et donc une synthèse temps-réel. Les récents développements de Modalys (N. Ellis, R. Piéchaud) montrent en effet la faisabilité de ce type d'implémentation et, par

la même, la richesse des possibilités offertes par un contrôle temps-réel d'un moteur de synthèse sonore par modélisation physique.

Cette procédure diffère, par exemple, d'un moteur de synthèse comme *Csound*² qui effectue ses calculs par bloc et non par échantillon. Dans *Csound* les blocs sont appelés « blocs de contrôle » (*control blocs*) et sont considérés comme des entités vis-à-vis du calcul. Ils contiennent un nombre défini d'échantillons et sont traités avec un pas de temps différent de la fréquence d'échantillonnage du signal audio. La procédure complète de traitement est constituée de chaînes (*chains*) comportant les instruments instanciés pour réaliser la synthèse. La chaîne est alors parcourue à chaque bloc et génère le son résultant en fonction des instruments actifs au cours de ce pas de temps. L'avantage de ce calcul par bloc réside essentiellement dans une économie de coût de calcul ; son inconvénient majeur, en revanche, est qu'il génère une latence liée à la taille des blocs de contrôle.

1.4.3 Modalys : Programmation

Modalys offre principalement trois possibilités de langage de programmation : un langage textuel (*Lisp*³), un langage de programmation graphique (*Max/MSP*⁴) et au sein d'un environnement de Composition Assistée par Ordinateur (*OpenMusic*⁵).

Le langage textuel a historiquement été le premier moyen de programmer l'environnement Modalys. Il utilise le langage *Lisp* – et initialement l'un de ses "dialectes", le *Scheme* –, largement reconnu dans le domaine de l'informatique musicale. Ainsi, tous les éléments et fonctions de Modalys sont imbriqués en *Lisp* et y apparaissent comme des objets additionnels à la syntaxe courante. A titre d'exemple, les commandes typiques de Modalys sont présentées au travers de l'exercice de construction d'un instrument de type xylophone (pour une présentation exhaustive de la syntaxe, voir le tutoriel de Modalys (J. Morrison, D. Waxman, F. Iovino, R. Dudas)).

En première approximation, le xylophone est constitué d'une barre en bois (résonateur) frappée par une mailloche (excitateur). Ces deux objets devront être définis, en même temps que la connexion 'frapper', la position initiale des deux objets ainsi que le mouvement de la mailloche au cours du temps.

Pour créer des objets, Modalys dispose d'une fonction primitive dont les arguments principaux sont le type de géométrie et les propriétés du matériau constituant l'objet ; cette déclaration peut ainsi s'écrire de la manière suivante – la liste des propriétés étant optionnelle et, le cas échéant, définie par défaut :

```
>> (define my-xylo (make-object 'rect-free-bar (length 0.3) (density 300)))
```

L'objet mailloche se définit de la même manière (éventuellement, avec un autre type d'objet) et la connexion de type 'frapper' peut être envisagée. Au préalable, un accès sur chacun des deux objets en interaction potentielle doit être défini, constituant ainsi un point de contact entre l'excitateur et le résonateur. Pour la barre, cette déclaration peut s'écrire de la manière suivante – en posant que l'accès est fixe (*const*), situé respectivement à 60% et 70% des deux dimensions principales de la barre (longueur, largeur) et que le contact s'effectue perpendiculairement à l'objet (*normal*) :

```
>> (define my-xylo-hit (make-access my-xylo (const .6 .7) 'normal))
```

² <http://www.csounds.com>

³ <http://common-lisp.net>

⁴ <http://cycling74.com>

⁵ <http://repmus.ircam.fr/openmusic/>

Une déclaration similaire sur l'objet mailloche (définissant la variable *my-stick-hit*) permet ainsi de définir l'interaction mutuelle entre les deux objets – où les deux derniers arguments numériques font référence aux positions initiales respectives de la barre et de la mailloche :

```
>> (make-connection 'strike my-xylo-hit my-stick-hit 0 .1)
```

Une fois la connexion instanciée, la modélisation de l'instrument en lui-même peut être considérée comme achevée et il ne reste plus qu'à définir la manière dont il va être "jouer", c'est-à-dire mis en mouvement. Pour cela, la position de la mailloche est définie au moyen du profil temporel de son évolution et la définition préalable d'un accès sur l'objet permettant de recevoir cette énergie cinétique. La déclaration peut se faire de la manière suivante – où la variable *my-stick-base* se réfère à cet accès et où le profil est défini par une liste de valeurs à deux entrées : temps et position :

```
>> (make-connection 'position my-stick-base (make-controller 'envelope 1  
      (list (list 0.00 .1) (list 0.05 -.0001) (list 0.10 .1) (list 5.00 .1))))
```

L'exécution de ces éléments de script, s'effectue par la dernière commande suivante, permettant de générer le signal de synthèse pour une durée donnée (ici, 5 sec.) :

```
>> (run 5)
```

Les deux autres modes de programmation de Modalys sont étroitement liés aux environnements qui les hébergent (à savoir, Max/MSP et OpenMusic) et ne seront donc pas détaillés dans ce document. Il est néanmoins intéressant de noter que le portage de Modalys dans l'environnement Max/MSP a permis une utilisation temps-réel de ce moteur de synthèse et a grandement élargi ses possibilités d'utilisation autant sur le plan musical que scientifique.

I.5 Développement d'un modèle d'accordéon dans Modalys

Le but de cette étude est la conception d'un modèle permettant de générer des sons d'accordéon et, le cas échéant, de s'en éloigner en agissant sur les paramètres physiques accessibles. La tâche comprend deux étapes : la compréhension de l'instrument réel au moyen d'une recherche bibliographique et expérimentale, puis l'implémentation d'un modèle physique dans le formalisme modal et compatible avec la structure du moteur de synthèse Modalys (voir § I.4) . En outre, ce travail s'effectue dans le cadre d'une collaboration artistique avec le compositeur Luis Naon⁶, et le réalisateur en informatique musicale associé au projet, Tom Mays.

I.5.1 L'accordéon

I.5.1.1 Introduction

L'accordéon est un instrument à anche libre. Les deux anches identiques, associées à une même note, sont commandées indépendamment par les flux d'air entrant et sortant générés par les mouvements alternatifs du soufflet (poussé / tiré). Cette étude s'intéresse aux phénomènes aérodynamiques liés à la vibration de l'anche. Des observations et mesures expérimentales sont effectuées sur plusieurs anches d'accordéon et pour différentes valeurs de pression d'alimentation. Parmi les données collectées : la pression acoustique en champ proche (*via* des microphones placés au-

⁶ <http://brahms.ircam.fr/luis-naon>

dessus et en-dessous de l'anche), le déplacement et la vitesse de l'anche (grâce à un capteur capacitif et un vibromètre laser), le profil géométrique de l'anche, la visualisation du mouvement de l'anche par stroboscopie et la visualisation du flux d'air à la sortie du dispositif.

A partir de ces résultats, les premiers éléments des principaux mécanismes physiques responsables du son de l'anche libre sont explicités; ils sont implémentés dans un modèle compatible avec le formalisme modal afin d'être intégrés dans l'environnement de synthèse Modalsys.

I.5.1.2 Etat de l'art

La recherche bibliographique a permis de faire le point sur les travaux scientifiques menés autour du fonctionnement de l'accordéon. Peu de publications traitent directement de cette problématique. La plus ancienne, qui s'avère être aussi l'une des plus complètes tant sur la question physique qu'historique, organologique ou musicologique, est le bulletin du Groupe d'Acoustique Musicale (G.A.M.) de 1972 [Monichon et al., 1972]. Quelques travaux plus récents font néanmoins état de nouvelles recherches : l'influence de la résonance de cavité sur la vibration de l'anche libre [Tonon, 2009], ou bien une méthode (brevetée) pour reproduire de manière électronique le son produit par un accordéon [Bruti, 2005].

Cela étant, le champ de cet état de l'art s'est également élargi aux "cousins" de l'accordéon : harmonica, harmonium, etc. En effet, cette famille d'instruments a fait l'objet de travaux de recherche plus nombreux, pour certains plus récents, et sont susceptibles d'aider à la compréhension de l'instrument visé : l'harmonica étudié par Cuesta, Valette et Millot [Millot, 2001], ou bien l'acoustique des instruments à anche libre étudiée par Cottingham et couvrant aussi bien les instruments occidentaux (harmonica, concertina, accordéon, ...) que les instruments asiatiques, dont notamment les différents orgues à bouches (le *sheng* chinois, le *sho* japonais ou le *khaen* thaïlandais ou laotien) [Cottingham, 2007; 2011].

I.5.1.3 Fonctionnement

Sur la base de la description de Leipp dans la section « Eléments d'anatomie, de physiologie et d'acoustique de l'accordéon » du bulletin du G.A.M. [Monichon et al., 1972 – p. 12], voici comment l'on peut définir le principe de fonctionnement de l'accordéon.

L'élément excitateur de l'instrument est une **anche simple**, le plus souvent métallique, et libre (*free reed*), c'est-à-dire qu'aucun obstacle ne vient contraindre sa vibration ; contrairement à la clarinette, par exemple, qui bat sur le plateau du bec – et que l'on qualifie donc d'anche battante (*beating reed*). En outre, l'anche d'accordéon est dite "forte" car c'est elle qui détermine la fréquence de jeu émise ; dans le cas de la clarinette, l'anche est dite "faible" car sa fréquence de résonance est nettement supérieure à la fréquence de jeu et c'est le tube de l'instrument (résonateur) qui pilote la fréquence du son émis.

De ce fait, chaque note de l'accordéon correspond en théorie à une anche ; en réalité, il y en a plusieurs à cause des deux modes d'action du soufflet (poussé et tiré) et des différents registres de jeu (figure 3).

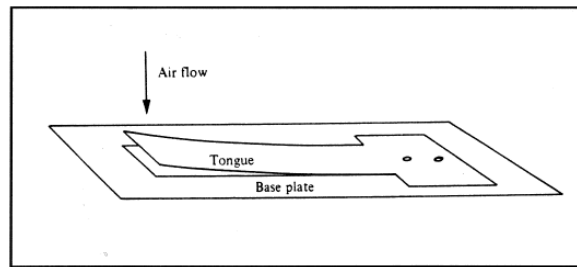


Figure 3. Schéma de l'anche libre de l'accordéon

Le résonateur est constitué d'une cavité alvéolaire, généralement en bois, et propre à chaque anche ; toutes les cavités d'un même registre sont regroupées ensemble pour former un sommier que l'on nomme "**musique**" (voir figure 4). Les musiques sont disposées à l'intérieur du soufflet et sont connectées, à leur base, au mécanisme d'actionnement des touches de sorte que l'enfoncement d'une touche provoque l'ouverture d'une fenêtre dans la cavité correspondante permettant ainsi l'écoulement du flux d'air et donc la mise en vibration de l'anche concernée.

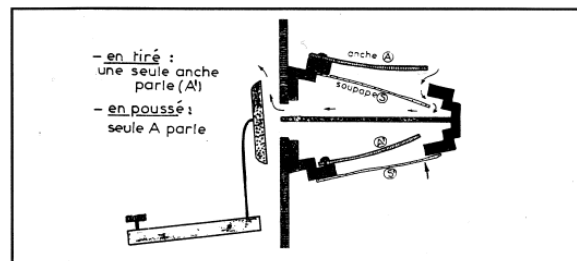


Figure 4. Schéma de la "musique" de l'accordéon

La connexion exciteur/résonateur s'effectue en fixant l'anche sur une plaquette de métal, d'épaisseur non négligeable, appelée "**platine**" qui est à son tour fixée sur l'une des faces de la cavité. La platine présente une découpe dont la géométrie s'ajuste le plus précisément possible à celle de l'anche afin de minimiser au maximum les fuites d'air lorsque l'anche se situe dans son épaisseur ; en revanche, lorsque l'anche est au-dessus de la platine, l'ouverture ainsi créée – par laquelle s'écoule le flux d'air – est appelée la "lumière".

En outre, chaque platine possède deux anches identiques montées en inverse afin de permettre les deux modes d'action du soufflet : poussé ou tiré. Pour cela, les deux anches sont en plus équipées de languettes de cuir qui sont fixées sur la platine et qui font office de soupape : suivant le mode de jeu – c'est-à-dire suivant le sens du courant d'air –, la languette se plaque rendant ainsi muette l'anche correspondante, ou au contraire se décolle permettant ainsi la vibration (voir figure 5).

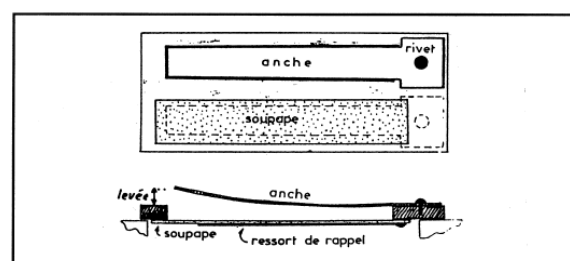


Figure 5. Schéma de la "platine" et de l'anche libre de l'accordéon

1.5.1.4 Approche expérimentale

Dans un premier temps, le prêt d'un accordéon démontable nous a permis d'appréhender concrètement les divers éléments constituant l'instrument.

Au-delà de cette première approche, la mesure in situ étant impossible sans détériorer gravement l'instrument (perçage du soufflet, etc.), nous avons élaboré avec l'aide de l'Atelier de Mécanique de l'Ircam (Alain Terrier) un montage expérimental simple constitué d'une cavité alimentée en air comprimé, surmontée d'une platine refabriquée aux dimensions et sur laquelle est rivetée une anche isolée de sa platine d'origine (voir figure 6). Ce système nous a permis d'effectuer des mesures de pression en amont et en aval de l'anche ainsi que des mesures de déplacement grâce à un dispositif mis au point lors d'une précédente étude (l'anche métallique est utilisée comme seconde armature d'un capteur capacitif dont on a enlevé la capsule).

En outre, l'emploi d'un matériau transparent, notamment pour réaliser la platine, a rendu possible la visualisation des mouvements de l'anche.

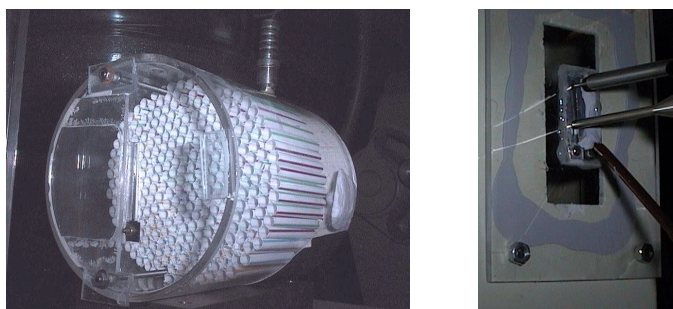


Figure 6. Dispositifs expérimentaux de mesure de l'anche d'accordéon.

Boîtier alimenté en air comprimé sur lequel est fixée la platine+anche (gauche) ; détails de la mesure du déplacement de l'anche au moyen d'un capteur capacitif (droite).

Les données recueillies lors de ces expériences sont directement exploitables : elles confirment les notions relatives au fonctionnement de l'instrument recueillies dans la littérature, elles constituent des éléments de comparaison quant à la forme de l'onde de pression acoustique dans le champ proche de l'anche et enfin, sous la forme d'un enregistrement vidéo, elles fournissent un outil d'observation du mouvement d'une anche libre (voir figure 7).

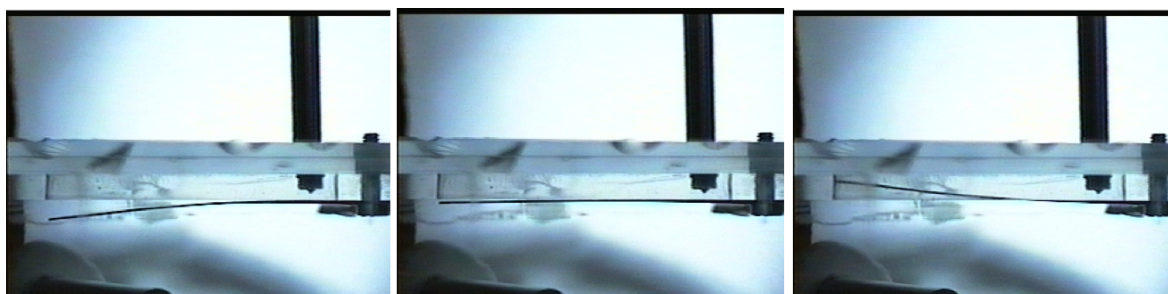


Figure 7. Mouvement d'une anche libre fixée sur une platine en plexiglas (extrait d'une prise de vue stroboscopique)

Ainsi, malgré quelques défauts de non-linéarité dans certains dispositifs de mesure, nous avons pu, par exemple, constater la caractéristique quasi-sinusoïdale du déplacement de l'anche, ou bien observer le fait que l'anche effectue une grande partie de son mouvement à l'intérieur de la platine (voir figure 8, milieu).

I.5.2 Implémentation dans Modalys

I.5.2.1 La connexion 'reed-free

Dans un premier temps, et avec l'exigence de s'insérer dans la temporalité du projet musical de L. Naon, notre démarche pour créer l'objet accordéon a été guidée par un souci d'efficacité avec, comme objectif, un résultat valide et exploitable dans l'environnement Modalys.

Dans cette optique, le travail de développement s'est concentré en priorité sur le couplage résonateur/excitateur. C'est en effet dans cette partie du modèle que sont implantées toutes les caractéristiques intrinsèques à chaque type d'interaction. Pour les instruments à vent, par exemple, on y trouve les principales formes de non-linéarité intervenant dans les mécanismes d'oscillations auto-entretenues ; comme, par exemple, dans le cas de la clarinette :

- l'existence de deux états de fonctionnement successifs (fermé/ouvert), suivant la position de l'anche ;
- la contribution non linéaire de la pression acoustique dans l'embouchure et du déplacement de l'anche à l'établissement du débit acoustique à l'entrée du résonateur. Pour la clarinette, cette relation entre débit (U), pression (P) et déplacement de l'anche (x) a été établie expérimentalement par Backus [Backus, 1963] :

$$U = B \cdot P^{\frac{2}{3}} \cdot x^{\frac{4}{3}} \quad , \text{ où } B \text{ est la constante de Backus} \quad (\text{Eq. 8})$$

C'est d'ailleurs à partir du code-source définissant la connexion 'reed de type anche de clarinette que la nouvelle connexion 'reed-free est implémentée en y adaptant le principe de fonctionnement de l'anche libre. En fait, les modifications majeures portent sur la description physique des deux états de fonctionnement et notamment celle de la position "fermée" :

- pour la clarinette, au cours d'une période de vibration de l'anche, cet état – très court – se caractérise par une vitesse nulle de l'anche et une force de réaction du bec avec lequel elle est en contact ;
- par contre, sur une période de vibration de l'anche d'accordéon, cet état – correspondant au moment où l'anche se trouve dans l'épaisseur de la platine – couvre un temps beaucoup plus long. En outre, du fait qu'il s'agit d'une anche libre, sa vitesse n'est jamais nulle (sauf quand elle s'inverse) et dans l'état "fermé", étant données les tolérances très serrées de l'assemblage anche/platine, on considère en première approximation que le débit en aval de l'anche, c'est-à-dire à l'entrée du résonateur, est nul ou tout du moins constant (si l'on décide de prendre en compte les fuites latérales).

Des mesures synchronisées effectuées sur le dispositif expérimental confirment relativement bien cette hypothèse : on observe en effet un palier quasi-constant de la pression acoustique externe à chaque fois que le déplacement de l'anche se fait vers l'intérieur (voir figure 8).

Par ailleurs, l'analyse des données générées par la connexion 'reed-free (voir figure 9) fait apparaître de bonnes similarités avec les mesures expérimentales – malgré quelques artefacts de simulation – tant au niveau du mouvement de l'anche que de la forme d'onde produite par le système excitateur. Ces résultats nous encouragent donc à

intégrer telle quelle la nouvelle connexion dans la bibliothèque Modalys et à l'associer aux objets déjà existant pour développer l'objet accordéon.

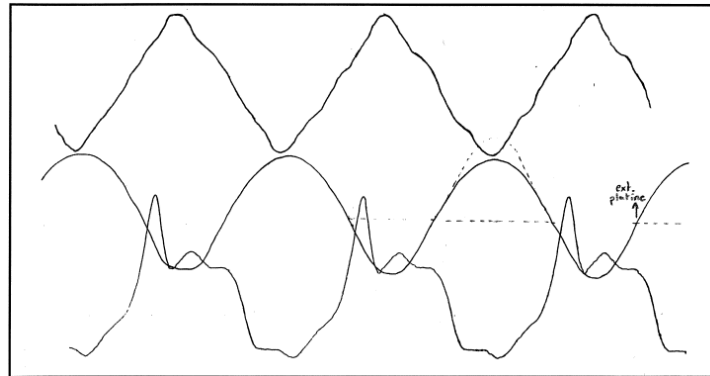


Figure 8. Modélisation de l'accordéon : données mesurées sur le dispositif expérimental (haut : pression interne - milieu : déplacement de l'anche - bas : pression externe)

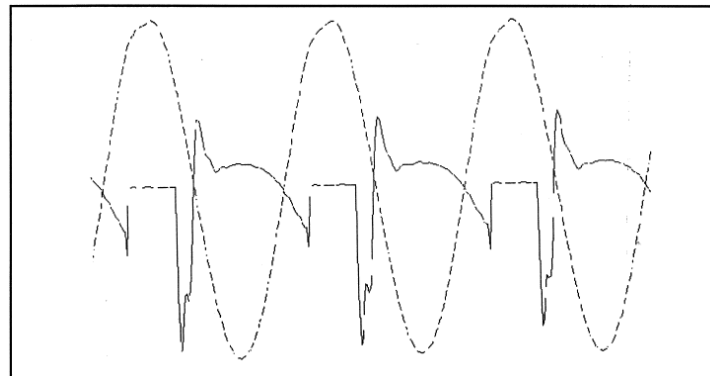


Figure 9. Modélisation de l'accordéon : résultats issus du modèle de synthèse (haut : déplacement de l'anche - bas : débit externe)

1.5.2.2 Construction et évaluation du modèle

Pour les raisons évoquées précédemment, l'objet accordéon tout entier est construit sur le même schéma que l'objet Modalys dont il est issu, la clarinette : l'anche est un système masse-ressort (objet générique '*bi-two-mass*', implémenté par défaut et constitué de 2 masses – M et m –, reliées par un ressort de raideur k) dont le rapport M/m est élevé pour pouvoir considérer que seule la petite masse (m) est en mouvement ; le résonateur est un tube cylindrique fermé-ouvert ('*closed-open-tube*') qui reçoit l'anche à son extrémité fermée.

Sur cette base, plusieurs modifications sont apportées :

- étant donné les caractéristiques de l'anche libre (voir plus haut), les paramètres relatifs à la petite masse (m) de l'objet '*bi-two-mass*' doivent être ajustés à la fréquence de jeu voulue ; cette ajustement se fait selon la relation :

$$2\pi \cdot f = \sqrt{k/m} \quad (\text{Eq. 9})$$

où f est la fréquence de jeu,
 m est la masse du système,
 k est la constante de raideur du ressort.

- d'après la configuration du résonateur réel de l'accordéon (voir § 1.5.1.3), il semble plus judicieux d'utiliser un tube acoustique fermé aux deux extrémités ('*closed-closed-tube*'). En outre, ce résonateur est accordé de manière à faire coïncider la

fréquence de jeu avec l'une de ses résonances, l'accord s'obtient en jouant sur la longueur du tube sachant que les résonances d'un tube "fermé-fermé" sont de la forme :

$$f_n = n \cdot \frac{c}{2L} \quad (\text{Eq. 10})$$

où f_n est la fréquence de la $n^{\text{ième}}$ résonance,
 c est la célérité du son dans l'air,
 L est la longueur du tube.

- par ailleurs, pour simuler le principe d'ouverture/fermeture de l'orifice situé sur la paroi de la cavité résonante (voir plus haut) et qui est en contact avec le mécanisme des touches, le résonateur est associé à un "trou acoustique" au moyen de la connexion '*hole*'. Le microphone d'écoute est alors placé à la sortie du trou – et non plus dans le tube – et, bien qu'aucun modèle de rayonnement ne soit présentement pris en compte, le rendu sonore s'avère de bonne qualité. Cette configuration permet également de jouer sur la dynamique de fermeture (ou d'ouverture) du trou, ce qui peut s'interpréter comme un contrôle de la vitesse d'actionnement de la touche. C'est une notion qui fait d'ailleurs partie des techniques de jeu réel dont l'accordéoniste dispose pour obtenir des effets de légères déviations de hauteur (*pitch bend*).
- conformément au paragraphe précédent (§ 1.5.2.1), l'interaction entre ces deux éléments s'opère cette fois par la connexion '*reed-free*' : sa syntaxe est la même que celle de la connexion '*reed*', seuls les arguments (enveloppe de pression, position initiale et surface active de l'anche) doivent être adaptés à la morphologie et aux conditions de jeu de l'accordéon. Le travail d'exploration de cette nouvelle connexion révèle que le contrôle de la pression est un élément très sensible, en premier lieu, pour l'émission d'une note mais aussi, par la suite, pour le timbre du son et le type d'attaque produits. En d'autres termes, ce paramètre de pression d'alimentation doit évoluer dans une plage de valeurs assez précise et selon un profil contrôlé, sous peine d'aboutir à un modèle d'accordéon inexploitable.

1.5.3 Conclusion de l'étude

Cette étude a permis de travailler sur un instrument de musique relativement méconnu, l'accordéon, et de s'intéresser plus généralement à une classe de la famille des instruments à vents : les instruments excités par une anche libre.

La confrontation avec l'instrument réel a permis, dans un premier temps, de se rendre compte de sa richesse musicale tant au niveau des possibilités de jeu qu'au niveau de la diversité des timbres disponibles ou de l'étendue de sa tessiture. Cette approche a également fait l'objet de mesures physiques sur un dispositif expérimental simplifié reproduisant le principe de base de la production du son dans l'accordéon, mesures qui constituent les premiers éléments de validation du modèle de synthèse qui sera par la suite effectuée au moyen d'expériences complémentaires.

A ce sujet, le développement de l'objet accordéon pour Modalys s'est effectué à deux niveaux : dans le noyau Modalys, une nouvelle connexion a été implémentée pour décrire spécifiquement les propriétés physiques de l'anche libre. Dans l'environnement Modalys, cette connexion, associée à d'autres objets existants, a permis d'aboutir à une synthèse sonore d'accordéon intéressante et supportant la comparaison avec le son de l'instrument réel. Par ailleurs, pour plus d'efficacité, ces développements se sont fortement inspirés d'un instrument déjà modélisé dans Modalys : la clarinette.

L'objectif initialement fixé pour cette étude a été atteint : développer un nouvel objet dans Modalys pour travailler autour des sonorités offertes par l'accordéon. Ceci étant, il semble évident que le modèle obtenu au final est loin d'être idéal et peut être largement amélioré. Mais, en tant que tel, il constitue un point de départ acceptable pour les travaux d'exploration des sons associés à cet instrument.

I.6 Etude de l'acoustique de l'anche libre

I.6.1 Préambule

L'étude initiale et préliminaire sur la physique du fonctionnement de l'anche d'accordéon s'est poursuivie dans le cadre du stage de DEA de Denis Ricot (Ecole Centrale Lyon) qui a revisité et amélioré principalement deux points : i/ le protocole expérimental, en développant un système plus précis, notamment en termes de visualisation de l'écoulement, grâce à l'utilisation de l'eau comme milieu de propagation ; ii/ la formalisation théorique, en modélisant plus finement les effets d'inertie du fluide présents en amont, responsables de la force d'excitation sur la lame. Ce travail a fait l'objet de communications scientifiques au 138^{ème} meeting de l'*Acoustical Society of America* (ASA) [Caussé et al., 1999a] et au V^{ème} Congrès Français d'Acoustique (CFA) [Misdariis et al., 2000] ; il a également abouti à une publication dans le *Journal of Acoustical Society of America* (JASA) [Ricot et al., 2005].

I.6.2 Publication [Misdariis et al., Congrès Français d'Acoustique, 2000]

Modélisation physique de la vibration d'une anche d'accordéon

Nicolas Misdariis, Denis Ricot, René Caussé
IRCAM, 1 place Igor Stravinsky 75004 Paris, France

I.6.2.1 Introduction

L'instrument est né vers 1829 en Allemagne (inventeur : C. Demian). A la même époque, en Angleterre, le scientifique C. Wheastone crée le concertina alors que le bandonéon apparaît un peu plus tard, vers 1850, en Allemagne (inventeur : H. Band).

Les anches utilisées dans l'accordéon sont des lames métalliques rivées sur une plaquette (platine) en aluminium percée d'une ouverture (rigole) dans laquelle la lame peut vibrer librement. C'est un écoulement d'air à travers l'anche, lié à la différence de pression de part et d'autre de la plaquette, qui est responsable de l'auto-oscillation. A chaque plaquette est associé un couple de lames identiques, montées en inverse, et qui permettent ainsi de jouer en poussé ou en tiré selon la direction du soufflet, une petite lamelle de cuir empêchant la lame non sélectionnée de vibrer.

Les anches sont de type +E et -I selon la classification de H.Bouasse et à la suite de l'analyse de H. von Helmholtz [Ricot, 1999] ; c'est à dire que la lame tend à fermer l'ouverture lorsque l'on souffle (+) de l'extérieur (E) ou que l'on aspire (-) de l'intérieur (I) de l'anche. Ces anches sont de plus libres (elles peuvent pénétrer à l'intérieur de la rigole et sont donc non battantes), et fortes (c'est à dire que la fréquence propre de la lame fixe la fréquence d'oscillation).

L'objectif de l'étude est de mettre au point un modèle physique de vibration de la lame pour des applications en synthèse sonore [Bouasse 1929 ; 1932].

1.6.2.2 Le son de l'accordéon : caractéristiques majeures

Ces caractéristiques sont déduites d'analyses, à la fois des sons de lames in vivo (montées sur un instrument) et in vitro (montées sur un réservoir et mises en vibration par une soufflerie artificielle). Les caractéristiques principales observables sur tout le registre et pour des conditions normales de jeu se traduisent par une forme d'onde de pression acoustique très "accidentée" (figure 10). Cette forme sur laquelle se détachent des zones de dépression (1) et de surpression localisées (2, 3 et 4), se retrouve pour le jeu poussé et tiré, à l'inversion près.

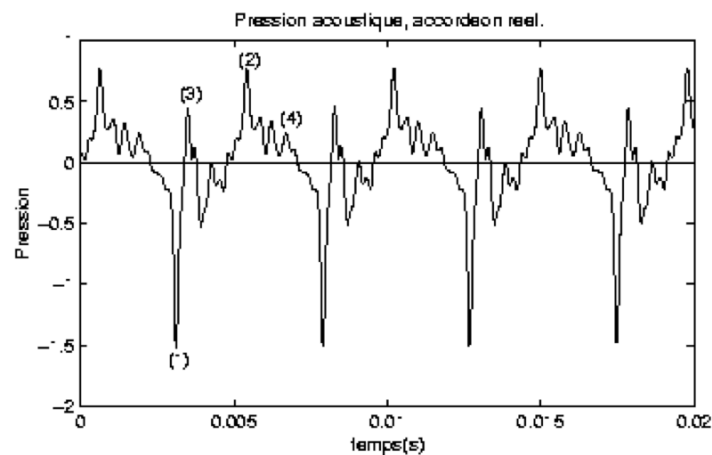


Figure 10. Accordéon– Evolution de la pression acoustique

Le spectre correspondant (figure 11) est large et riche en harmoniques. Pour les premiers partiels, les harmoniques de rang pair ont une amplitude beaucoup plus faible que les harmoniques impairs. Il apparaît de plus que les résonateurs (réservoir ou petites cavités propres à chaque lame), en amont ou en aval de l'anche, ont une influence négligeable sur le spectre.

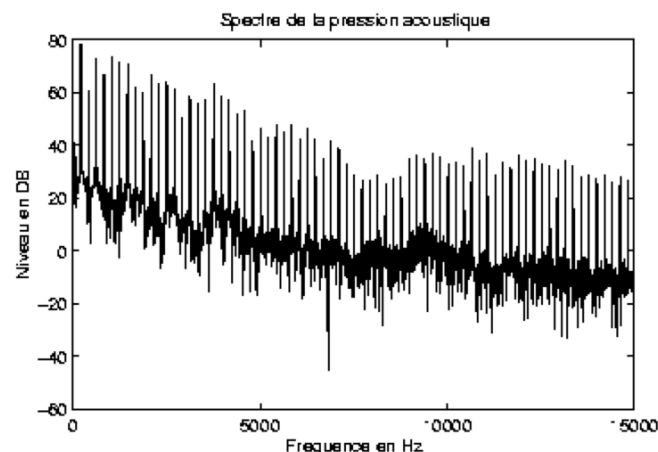


Figure 11. Accordéon – Spectre du son tenu (de l'accordéon)

La visualisation stroboscopique de la vibration de l'anche permet d'établir une correspondance entre ces caractéristiques et l'ouverture ou la fermeture de la plaquette par la lame. Celle-ci vibre sur son premier mode (lame encastree-libre) à une fréquence légèrement inférieure à sa fréquence propre [Koopman et al., 1996].

1.6.2.3 Caractérisation physique de l'instrument

L'objectif de l'approche expérimentale de l'étude se résume en plusieurs points : arriver

à une compréhension plus "fine" du fonctionnement du système, évaluer l'importance des différentes grandeurs physiques dans le fonctionnement, déceler "intuitivement" des relations entre les paramètres physiques importants et en dernier lieu valider le modèle. Pour cela, nous procédons à des mesures du déplacement relatif de la lame, de la pression hydrodynamique mesurée à quelques millimètres en aval de la lame (représentative de l'écoulement et du mouvement réciproque local du fluide) et de la pression acoustique en champ proche. Un enregistrement simultané de ces trois variables est reporté à la figure 12.

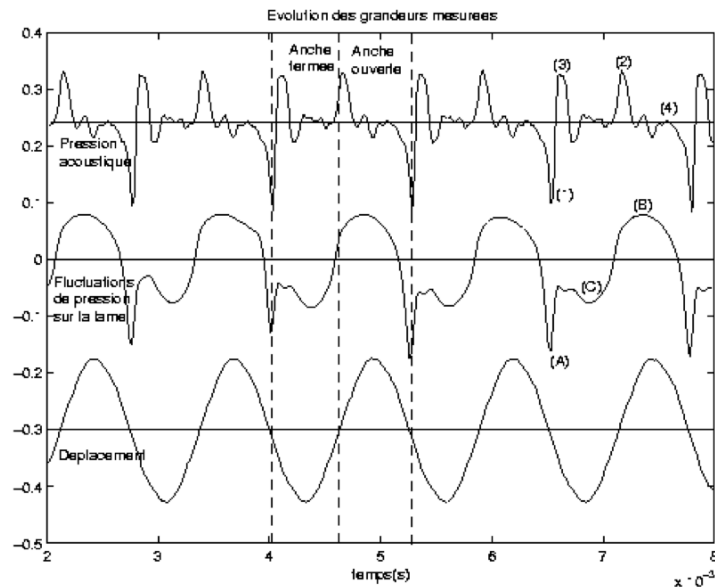


Figure 12. Accordéon – Evolution des signaux de pression et de déplacement (mesure)

On constate :

- que le déplacement est sinusoïdal sur toute la plage de variation de la pression d'alimentation P_0 utilisée (10 à 600 Pa). Précisons que la plage normale de jeu est située entre 10 et 300 Pa pour une lame du registre médium. Devant la non-linéarité liée à la faible dynamique de notre capteur de déplacement, une vérification assez fine du caractère sinusoïdal de la vibration a été effectuée avec un vibromètre laser. Un échantillonnage spatial de la lame avec le vibromètre conduit à la mesure de la déformée.
- qu'il existe deux régimes de fonctionnement aux formes d'onde différentes et dont le comportement en fréquence, en fonction de P_0 , est inversé. Pour le premier la fréquence décroît linéairement en dessous de la fréquence propre de la lame. Alors que pour le deuxième cette fréquence augmente rapidement, pouvant même passer au-dessus de la fréquence propre (voir figure 13). Ces deux régimes correspondent à deux positions d'équilibre de la lame par rapport à la plaquette, respectivement à l'extérieur et à l'intérieur. Le deuxième régime est observable pour certaines lames dans les conditions normales de jeu. On remarque également que le changement de régime est soumis à des effets d'hystérésis.
- que le tracé de la pression acoustique en fonction de la plage utile de P_0 , non reporté ici, nous renseigne directement sur la dynamique de jeu qui est de l'ordre d'une quarantaine de décibels. Pour connaître la force exercée par la pression hydrodynamique (pression excitatrice) sur la lame, il reste à décrire le plus précisément

possible l'écoulement. Pour cela, des visualisations ont été réalisées, dans l'eau après une tentative infructueuse dans l'air. La réalisation de traceurs, à l'aide de panaches de colorant, est plus facile à réaliser dans l'eau que dans l'air. Ces visualisations nous renseignent sur le caractère laminaire ou turbulent des jets qui se forment en aval de la lame.

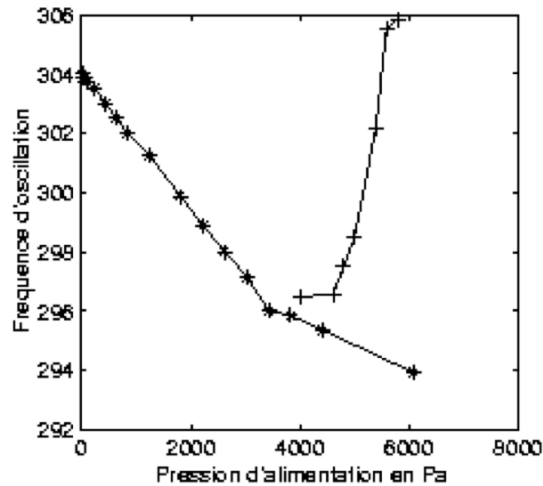


Figure 13. Accordéon – Fréquence d'oscillation en fonction de P0
(*) : régime 1 ; (+) : régime 2

1.6.2.4 Caractérisation de l'écoulement

Nous observons avec étonnement que les lames, prévues pour un fonctionnement dans l'air, oscillent également sur leur premier mode lorsqu'elles sont soumises à un écoulement d'eau.

Les vitesses d'écoulement sont d'environ 15 m/s dans l'air et de 2 m/s dans l'eau, ce qui donne des nombres de Reynolds du même ordre de grandeur et qui valide donc l'analogie entre l'eau et l'air pour l'étude de l'écoulement.

Les visualisations permettent de caractériser les écoulements :

- en amont, il est laminaire et de type "puits". Près de la lame, l'écoulement est bidimensionnel : le fluide envoyé dans un plan perpendiculaire à la grande dimension de la lame sera "aspiré" par la fente se trouvant dans le même plan.
- l'écoulement aval est de type "jet", unidirectionnel. Son étude pourra être faite dans une hypothèse de bidimensionnalité. Cet écoulement reste "collé" à la paroi de la rigole ; ce n'est donc pas tout à fait un jet libre mais plutôt un écoulement laminaire sur un plan.

L'eau est un fluide lourd. La fréquence d'oscillation des anches dans l'eau est bien plus basse que dans l'air, respectivement 142Hz et 330Hz pour l'une des lames. Cependant, malgré les différences de milieu le comportement de la fréquence en fonction de P0 reste semblable.

Les visualisations permettent de conclure, comme l'avait montré l'étude de St Hilaire [St-Hilaire et al., 1971], que les instabilités turbulentes qui peuvent se développer dans le jet ne sont pas à l'origine de l'auto-oscillation de la lame. D'autres arguments nous conduisent à la même conclusion. Parmi eux, le fait que ces instabilités ne soient pas très efficaces ou que la fréquence d'oscillation soit stable pour une plage de variation très importante de la vitesse de l'écoulement (pour les dimensions et la géométrie de l'ensemble lame/platine, le nombre de Strouhal correspondant à une telle configuration est de l'ordre de 0.15).

La force d'excitation est créée par les effets d'inertie du fluide présent en amont de la lame comme l'avait entrevu H. Bouasse [Ricot, 1999]. Les forces de pression s'appliquent seulement sur la face amont de la lame bordée par les deux jets. La face aval est une zone de fluide "mort". Les paramètres importants du système sont P_0 , la pression d'alimentation, et K_0 , la quantité de fluide disponible dans le réservoir et qui rentre en jeu dans les phénomènes d'inertie.

1.6.2.5 Modèle physique de la vibration de l'anche

A la différence de la clarinette [Hirschberg et al., 1994], l'auto-oscillation est le résultat du caractère instationnaire de l'écoulement. Ici le jet est "libre" et n'a pratiquement aucun contact avec la face aval de la lame.

Le modèle est basé sur une représentation simple de la lame de type masse/ressort/amortissement et sur le calcul d'un écoulement potentiel bidimensionnel instationnaire. Le fluide (air) est supposé parfait (non visqueux) et les caractéristiques de l'écoulement seront : potentiel, irrotationnel, incompressible, bidimensionnel et instationnaire. L'écoulement amont est de type "puits" et l'écoulement aval de type "jet". On calculera la pression qu'exerce l'écoulement sur une section de la lame en fonction de sa position. C'est cette dernière qui fixe la surface d'ouverture de l'anche.

Le modèle ne permet pas de résolution analytique et les équations différentielles sont fortement non-linéaires. Une résolution numérique du problème est donc mise en oeuvre. La complexité du modèle vient en partie du fait que la géométrie de l'écoulement varie au cours du temps : le "puits" a une longueur variable. Afin de simplifier on négligera ce point et on supposera que le "puits" est ponctuel et que les variations d'intensité de l'écoulement prédominent dans l'établissement des forces.

Sur la figure suivante (figure 14) sont reportés plusieurs résultats de la simulation obtenus à partir du modèle simplifié. Ainsi le déplacement, la pression hydrodynamique et de la pression acoustique pour un couple (P_0 , K_0) donné peuvent être comparés avec les mesures respectives de la figure 12.

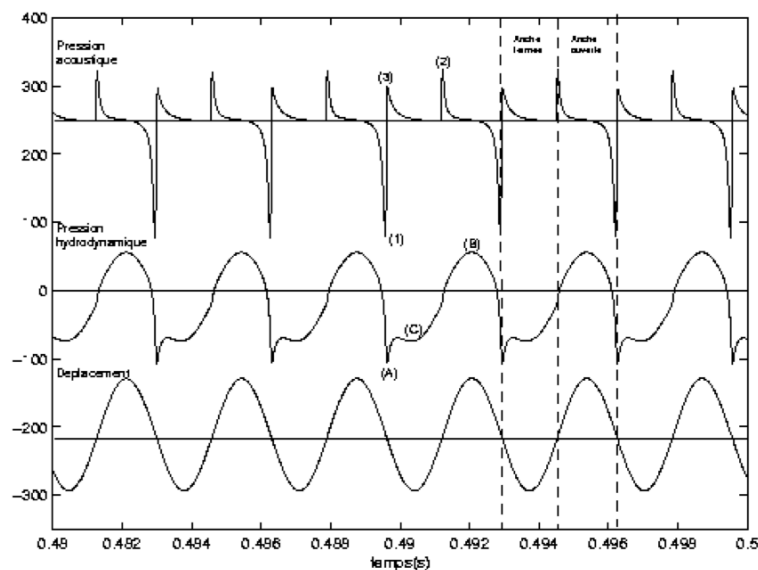


Figure 14. Accordéon – Evolution des signaux de pression et de déplacement (simulation)

La valeur de la pression acoustique est obtenue par la combinaison linéaire d'une

source monopolaire et d'une source dipolaire et du calcul du débit donné par le modèle d'écoulement précédent.

Malgré la solution simplifiée choisie pour la simulation, les résultats obtenus sont très encourageants tant au niveau du réalisme des sons synthétisés qu'au niveau de la similitude du comportement physique comme les courbes de la figure 14 le soulignent.

I.6.2.6 Conclusion – Perspectives

Nous avons étudié la lame d'accordéon à la fois du point de vue mécanique, aéroacoustique et acoustique. Suite à cette étude, nous proposons un modèle physique d'oscillation de la lame d'accordéon, non-linéaire et très complexe. Un modèle simplifié est dérivé de ce dernier qui donne des résultats très prometteurs, même si sa validation est incomplète. Cependant des améliorations seraient à apporter en particulier sur le modèle de génération de la pression acoustique.

Des questions fondamentales restent toujours ouvertes comme par exemple l'influence de la taille du réservoir sur les effets d'inertie et donc sur la vibration.

Pour la synthèse sonore, des mesures in vivo de l'évolution temporelle de P_0 au cours du jeu, du comportement des lames en fonction du registre ou de l'influence des petites cavités conduiraient certainement à des sons encore plus proches de la réalité.

I.7 Conclusion

Cette première partie relative à la thématique de la **synthèse du son** a donc permis de dresser un état de l'art des différentes pratiques de synthèse sonore numérique, de se focaliser sur l'approche dite par modélisation physique et, dans ce cadre, de s'intéresser plus particulièrement à la synthèse modale, à la fois d'un point de vue formel mais également par rapport à son implémentation dans l'environnement de synthèse Modalys.

Les connaissances acquises dans ce domaine – tant du point de vue de la physique que de l'informatique – m'ont permis d'apporter des contributions personnelles à ce champ de recherche sous la forme d'une étude expérimentale des mécanismes physiques régissant le fonctionnement de l'anche libre d'accordéon et de la conception d'un modèle développé pour le logiciel Modalys.

Nous verrons par la suite que l'expertise acquise vis-à-vis de cette approche de synthèse et de cet environnement de programmation me permettra également de formaliser et de mettre en oeuvre une synthèse sonore augmentée dite « Synthèse Additive Spatiale » (voir § II.5.1) qui combine deux grandes thématiques abordées dans le document, à savoir la synthèse et la reproduction des sons.

Deuxième partie

Reproduction

PARTIE II. Reproduction

II.1 Problématique

Par définition⁷, la reproduction est « le fait de produire des copies d'un original à l'aide d'une technique, d'un procédé particulier ». Appliquée au son, la reproduction sonore désigne, plus spécifiquement « le fait de restituer un son direct ou enregistré par un procédé technique permettant soit une écoute directe soit une écoute différée ».

Cette seconde définition rassemble donc les deux principaux cas de figure où l'on parle de reproduction sonore : l'augmentation (ou l'amplification) « en direct » d'une **source réelle** « directe », ou bien la révélation (ou restitution) « en différé » d'une **source virtuelle** « enregistrée ». Nous verrons d'ailleurs, un peu plus loin, que cette dualité source réelle / virtuelle constitue un questionnement important de la musique contemporaine vis-à-vis de la technologie, dans un cadre de composition dite "mixte" qui fait coexister, dans la partition, des parties instrumentales et électroniques.

Dans tous les cas, la notion de reproduction sonore est née avec les grands progrès de l'histoire des sciences et techniques de la seconde moitié du XIX^{ème} siècle et s'est prolongée au XX^{ème} siècle – voire au début XXI^{ème} – grâce à de nombreuses innovations technologiques. Elle est par ailleurs liée de manière indissociable à celle de l'enregistrement (voir [Watkinson, 1998] pour un développement complet du sujet).

Historiquement, c'est en effet une machine à enregistrer la vibration acoustique sur un support matériel, par un procédé purement mécanique, qui est d'abord inventée par Edouard-Leon Scott de Martinville (1857) : un stylet solidaire d'une membrane élastique en contact avec un cylindre recouvert de noir de fumée sur lequel il crée un sillon ; le dispositif produisant une analogie visuelle des ondes acoustiques mais ne permettant pas de les rejouer.

Vingt ans plus tard, Charles Cros (1877) améliore le procédé en effectuant la gravure dans un cylindre de métal (étain), mais surtout en considérant le système réversible : lorsque la pointe du stylet repasse dans le sillon, la membrane restitue le signal acoustique original. Il met ainsi au point un « procédé d'enregistrement et de reproduction des phénomènes perçus par l'ouïe » qu'il nomme paléophone (en rapport aux voix du passé) et qui préfigure le prototype du **phonographe**, dispositif purement mécanique de reproduction des sons. Sur le même principe et dans le même temps – mais juste un peu avant C. Cros – l'Américain T. Edison dépose le brevet du phonographe avec lequel il enregistre et reproduit, pour la première fois, la voix humaine (sa voix récitant la comptine "*Mary had a little lamb*").

Viennent ensuite diverses améliorations techniques dont le graphophone (Alexandre Graham Bell et associés) utilisant un cylindre de cire (pour une meilleure reproduction) et un mégaphone (pour une meilleure captation), puis le **gramophone** d'Emile Berliner qui réalise les enregistrements latéralement (et non plus verticalement), sur un disque de shellac (et non plus de cire) ; ce dernier invente également le principe de la matrice en développant un procédé de copie d'une gravure originale à partir d'un négatif.

⁷ <http://www.cnrtl.fr>

Par la suite, supportées par la maîtrise de l'électricité et sur la base du téléphone (que Bell invente en 1876), les techniques d'enregistrement et de reproduction se développent progressivement sur le principe de la transduction électro-acoustique. L'enregistrement est effectué au moyen d'un **microphone**, les techniques évoluent de mécanique à numérique en passant notamment par les supports magnétiques ou optiques. La restitution utilise une chaîne d'amplification du signal enregistré à base d'un amplificateur et d'un **haut-parleur**.

Le haut-parleur (HP) réalise donc une **transduction** électro-mécano-acoustique qui a pour résultat la transformation des modulations d'un signal électrique en onde acoustique audible. La structure du haut-parleur comporte deux principaux composants pour assurer ces différentes opérations :

- le **moteur** qui réalise la transduction **électro-mécanique** ; cette opération est basée sur deux principes physiques exprimant la transformation d'une énergie électrique en une énergie mécanique :
 - une force électromagnétique, dite force de Laplace, qui exprime la force exercée sur un élément conducteur, de longueur dl , parcouru par un courant électrique I et placé dans un champ magnétique B (voir figure 15) :

$$d\vec{F} = I \cdot d\vec{l} \wedge \vec{B} \quad (\text{Eq. 11})$$

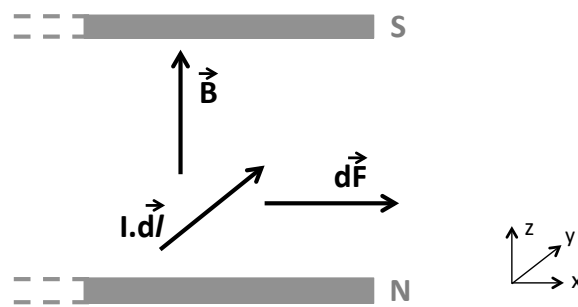


Figure 15. Illustration de la force électromagnétique (force de Laplace)

- une force électrostatique, résultat de la loi de Coulomb, qui exprime la force exercée sur une particule électrique q placée dans un champ électrostatique E (voir figure 16) :

$$\vec{F} = q \cdot \vec{E} \quad (\text{Eq. 12})$$

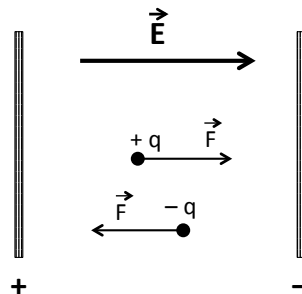


Figure 16. Illustration de la force électrostatique (loi de Coulomb)

- la **membrane**, quant à elle, réalise la transduction **mécano-acoustique** en mettant en mouvement le milieu avoisinant sa surface ; définissant ainsi deux types d'ondes :
 - l'onde avant couplée à des « charges avant » : généralement l'espace de rayonnement direct ou délimité par un pavillon ;

- l'onde arrière couplée à des « charges arrière » : entre autres, enceinte close, résonateur couplé, ou système bass-reflex.

Suivant les deux principes physiques énoncés précédemment, il existe donc deux grands types de haut-parleurs (voir [Loyez, 1992 ; Mercier, 2002] pour des éléments de théorie et de pratique sur cette question) :

- le **haut-parleur électrodynamique**, basé sur l'effet électromagnétique et composé principalement d'une bobine mobile, solidaire de la membrane et parcourue d'un courant électrique alternatif, ainsi que d'un aimant créant le champ magnétique au niveau de la bobine (voir figure 17).

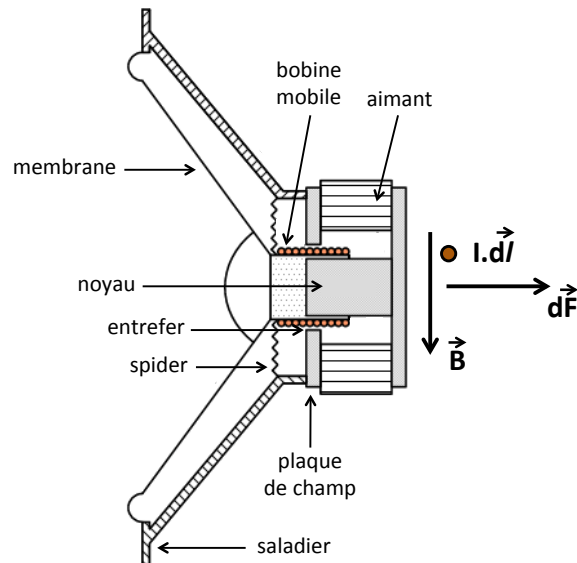


Figure 17. Détails d'un haut-parleur électrodynamique

A noter qu'il existe des variantes de ce principe, dont entre autres :

- le haut-parleur à ruban : bobine et membrane sont rassemblées dans un même élément ;
- le haut-parleur magnétostrictif : utilisant la propriété de certains matériaux ferromagnétiques (fer, nickel, etc.) de se contracter/dilater sous l'effet d'un champ magnétique – mais rarement utilisé en audio.
- Le **haut-parleur électrostatique** basé sur l'effet électrostatique et composé principalement d'une membrane polarisée (diaphragme) – sur laquelle sont stockées les charges électriques – et de deux électrodes dans lesquelles circulent le courant électrique modulé (voir figure 18) :

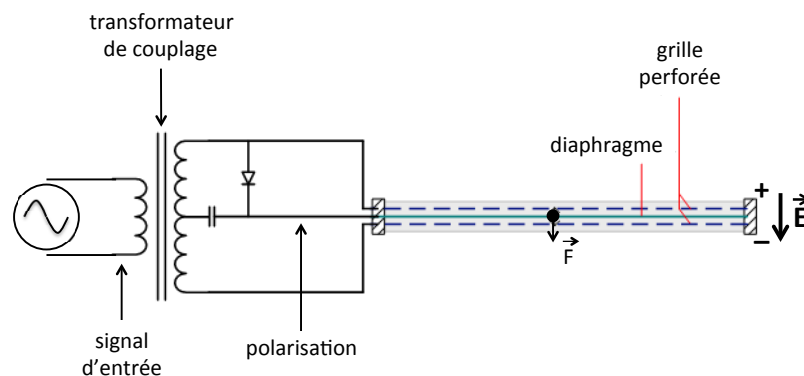


Figure 18. Détails d'un haut-parleur électrostatique

A noter qu'il existe également des variantes de ce principe, dont entre autres :

- le HP à électret : les charges électriques sont stockées en permanence dans une membrane isolante ([Sakamoto et al., 1975] – cité dans [Loyez, 1992]).
- le haut-parleur piézo-électrique : utilisant la propriété de certains matériaux (quartz, ferrites, polymères ou céramiques PZT – Titano-Zirconiate de Plomb) à se contracter/dilater sous l'effet d'un champ électrique (analogue au principe de magnétostriction – on parle d'ailleurs d'électrostriction).
- le buzzer piézo-électrique : source de son élémentaire (type alarme) dont la conception peut combiner à la fois l'électronique (oscillateur), le transducteur (céramique piézo-électrique) et la charge acoustique (résonateur de Helmholtz) ; voir § IV.2.2 pour une application de cette technologie.

A partir de cet élément de base qu'est le haut-parleur, l'objectif principal de la reproduction sonore va donc consister à créer – ou projeter – des **images sonores** (images de réalités acoustiques pré-existantes ou co-existantes à un instant donné) afin de recréer des scènes auditives complexes (*auditory scene*) [Bregman, 1994]. La qualité de l'image est alors jugée à la fois sur des critères spectraux (réponse en fréquence) ou temporels (dynamique, sensibilité), mais aussi – avec des dispositifs autres que monophoniques – sur des critères spatiaux : localisation, immersion. On arrive alors à la notion de **son spatialisé**, et par extension, "son 3D" [Blauert, 1997].

Il existe plusieurs manières de traiter la question du rendu sonore spatialisé qui dépendent avant tout de la dimensionnalité elle-même et qui correspondent à des approches différentes de description et de modélisation du champ sonore.

Un **modèle perceptif** qui s'appuie sur deux indices fondamentaux utilisés par le système auditif pour se représenter le champ sonore environnant : la différence de temps d'arrivée de l'onde sonore entre les 2 oreilles (ITD – *Inter-aural Time Difference*) et la différence de niveau d'arrivée de l'onde sonore entre les 2 oreilles (ILD – *Inter-aural Level Difference*) – voir figure 19.

Le principe de ce modèle a été posé par Rayleigh dans une théorie (théorie « Duplex ») qui postule la complémentarité de ces deux indices en les assignant, en 1^{ère} approximation, à des bandes fréquentielles du spectre audible distinctes : l'ITD en "basses" fréquences (< 1600 Hz) et l'ILD en "hautes" fréquences (> 1600 Hz).

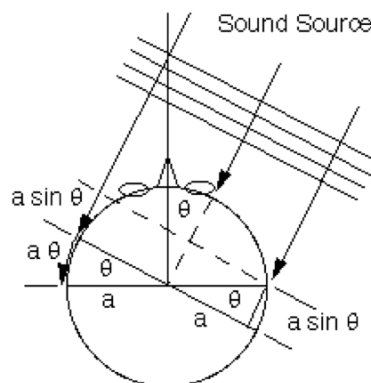


Figure 19. Illustration du principe de la théorie Duplex de Rayleigh. (ITD et ILD)

Les techniques de reproduction sonore exploitant ce principe sont essentiellement la **stéréophonie** (2 haut-parleurs contrôlés en amplitude et en phase) et toutes les extensions associées, aussi bien 2D (5.1, 7.1, 10.2, etc.) que 3D (p. ex., VBAP – *Vector-Based Amplitude Panning* [Pulkki, 1997]).

Un **modèle signal** qui s'appuie sur des fonctions de transfert caractérisant le trajet d'une onde acoustique entre un point source et le système récepteur humain (précisément, le tympan). Ces fonctions, appelées communément **HRTF** (*Head-Related Transfer Functions*), dépendent de la position de la source ainsi que des caractéristiques morphologiques du récepteur, c'est-à-dire formes du pavillon de l'oreille, de la tête, du buste, etc. ; elles sont donc, a priori, propres à chaque individu même si des invariants existent et permettent d'utiliser ce principe de manière relativement efficace. Les HRTFs constituent donc des filtres auditifs (voir figure 20) qui permettent d'encoder le champ sonore en 3D et restituer la position d'une ou plusieurs sources – voire d'une scène sonore complexe – dans l'espace par une opération de traitement du signal (convolution).

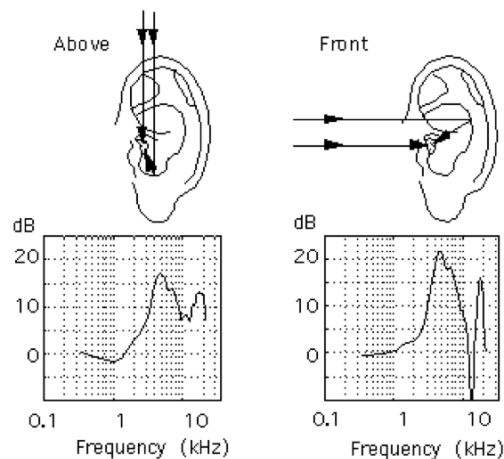


Figure 20. Illustration du principe des HRTFs (*Head-Related Transfer Functions*)

Les techniques de reproduction sonore exploitant ce principe sont essentiellement la restitution binaurale (sous casque), ou transaural (2 haut-parleurs et un traitement de signal complémentaire permettant d'annuler les chemins croisés vers les 2 oreilles).

Un **modèle physique** qui s'appuie sur la description physique du champ sonore selon deux approches différentes :

- le principe de Huyghens (1678) qui stipule que tout champ sonore résultant de l'émission et la propagation d'une onde issue d'une source (principale) peut être décomposé en une infinité de sources secondaires générant des ondes sphériques élémentaires et disposées continument le long du front d'onde ; ces sources secondaires jouant le rôle de relais capables de reproduire l'onde originale en l'absence de la source principale (voir figure 21).

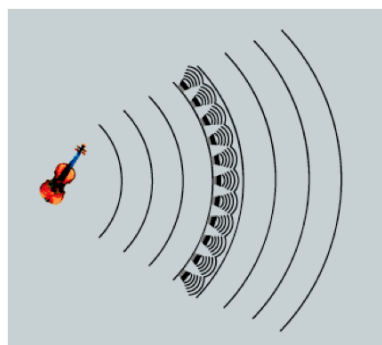


Figure 21. Illustration du théorème de Huyghens

Ce principe – moyennant des simplifications pour le rendre applicable – détermine un procédé de reproduction sonore dit "holophonique" (par analogie avec l'holographie en vision) [Berkhout et al., 1993 ; Nicol, 1999 ; Corteel, 2004 ; Gauthier, 2007]. Il réalise ainsi une **synthèse de fronts d'onde** (WFS – *Wave Field Synthesis*) qui, moyennant un dispositif massivement multi haut-parleurs et une importante puissance de calcul, permet en partie de résoudre l'éternel problème de la position d'écoute privilégiée (*sweet spot*) en étendant de manière significative la zone d'écoute efficace, c'est-à-dire la zone où le champ sonore est cohérent pour l'auditeur.

- l'approche modale qui considère que tout champ sonore peut être décomposé sur une base orthonormée, et a priori infinie, de fonctions spatiales, appelées "**harmoniques sphériques**" [Gerzon, 1985 ; Malham, 1995 ; Daniel, 2000] (voir figure 22 – et § II.2.2 pour les détails sur les harmoniques sphériques).

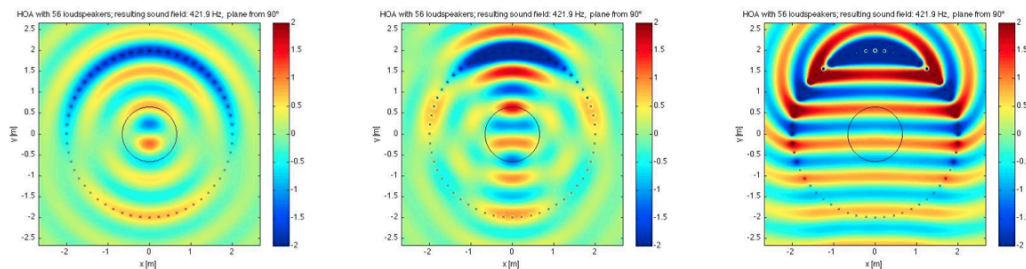


Figure 22. Illustration du principe de décomposition en harmoniques sphériques
Simulation – Reproduction d'une onde plane au degré 1 (gauche), 3 (centre) et 20 (droite)

Ce principe, limité à un ordre de décomposition donné, détermine la technique **Ambisonics** (ou HOA – *High Order Ambisonics* –, pour les degrés élevés) qui permet de recréer l'image spatiale et ses propriétés directionnelles à partir d'un nombre scalable de haut-parleurs (déterminant la précision de la restitution).

Dans cette approche sur la restitution de la scène sonore spatialisée, la localisation et l'immersion (enveloppement) sont des critères de premier ordre. Cependant, un autre facteur joue un rôle important dans la qualité de l'image projetée : la **directivité**. Ce paramètre rend compte du phénomène de rayonnement acoustique, c'est-à-dire de la manière dont l'énergie acoustique émise par une source est distribuée dans l'espace. En tant que tel, il rend compte de l'interaction de la source avec son milieu, notamment la salle de diffusion, et peut donc avoir une influence significative sur la nature et l'amplitude des réflexions qui ont lieu durant la propagation acoustique. En conférant des propriétés spatiales spécifiques à la source d'émission, la directivité peut ainsi avoir un impact non négligeable sur la manière dont la source "éclaire" l'espace de diffusion et, *in fine*, la manière dont elle perçue par l'auditeur.

II.2 Etude du rayonnement acoustique

L'étude du rayonnement acoustique des structures représente une activité importante dans le domaine de l'acoustique autant industrielle (p. ex., pour pouvoir réduire efficacement l'énergie émise par une machine) que musicale (p. ex., pour caractériser l'énergie émise par un instrument, et le cas échéant, déterminer la bonne configuration d'un dispositif d'enregistrement) (voir [Lesueur, 1988] pour un développement complet sur cette question générale).

Nous nous contenterons dans ce chapitre d'aborder uniquement deux points théoriques : la modélisation du rayonnement du haut-parleur et la décomposition du champ sonore sur la base des harmoniques sphériques.

II.2.1 Modélisation du rayonnement du haut-parleur : le piston plan

En première approche, il est d'usage de modéliser le rayonnement du haut-parleur en utilisant le modèle du piston plan circulaire encastré dans un écran infini (baffle). Bien que la formulation générale de ce problème ne puisse pas être résolue de manière analytique, on peut néanmoins en calculer une solution exacte dans l'axe du piston et une solution approchée en champ lointain., c'est-à-dire lorsque la distance du point d'observation est très grande devant la dimension (rayon) du piston.

Les données du problème sont les suivantes (voir figure 23) :

- un piston de rayon a , animé d'une vitesse normale V_0 uniforme à la surface du disque Σ ;
- un point courant M_0 à la surface du disque, définissant un vecteur \mathbf{r}_0 repéré par l'angle φ dans le plan (x,y) du disque;
- un point d'observation M dans le demi-espace avant délimité par l'écran infini, définissant un vecteur \mathbf{r} repéré par les angles θ et Φ , dans l'espace (x,y,z) .

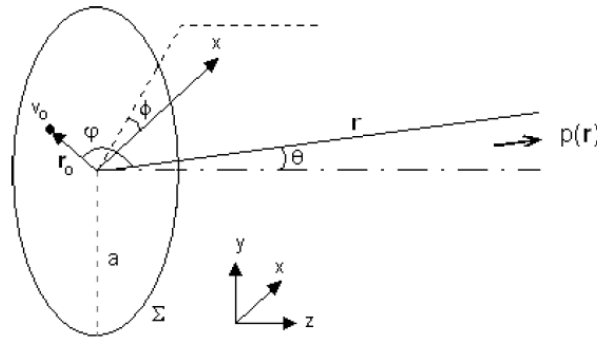


Figure 23. Rayonnement du piston plan – données du problème

La formulation du problème générale s'effectue au moyen de l'intégrale de Rayleigh qui permet de calculer le champ de pression acoustique rayonnée par une surface plane dont on connaît la vitesse normale [Chaigne et al., 2008]. Dans les conditions de champ lointain, c'est-à-dire lorsque $||\mathbf{r}|| \gg a$, et en faisant des approximations correspondant à des développements limités des grandeurs aux 1^{ers} ordres, l'intégrale de Rayleigh se simplifie, et peut alors s'exprimer à l'aide de la fonction de Bessel de 1^{ère} espèce d'ordre 0 $J_0(z)$ puis – grâce aux propriétés de cette famille de fonctions – se résoudre à l'aide de la fonction de Bessel du premier ordre, $J_1(z)$.

La pression acoustique du piston plan circulaire (encastré dans un baffle infini) peut alors s'écrire de la manière suivante (voir Annexe 1 pour les détails du calcul) :

$$p(r, \theta) = j\rho_0 c V_0 \cdot \frac{ka^2}{2} \cdot \frac{e^{-jkr}}{r} \cdot \left[\frac{2J_1(ka \sin \theta)}{ka \sin \theta} \right] \quad (\text{Eq. 13})$$

où :

- ρ_0 est la masse volumique du milieu,
- c est la célérité du son dans le milieu,
- V_0 est la vitesse normale à la surface du piston,

- $k = \omega/c$ est le nombre d'onde,
- a est le rayon du piston circulaire,
- r est la distance du point d'observation,
- J_1 est la fonction de Bessel de 1^{ère} espèce à l'ordre 1,
- θ est l'angle du point d'observation par rapport à l'axe du piston.

Cette formulation fait notamment apparaître l'expression d'une **fonction de directivité** $D(\theta)$ qui représente la dépendance du champ de pression en fonction de θ et de la grandeur caractéristique ka :

$$D(\theta) = \frac{2J_1(ka \sin \theta)}{ka \sin \theta} \quad (\text{Eq. 14})$$

Cette fonction D décrit la manière dont l'amplitude de la pression acoustique est modulée dans l'espace et est donc responsable des "lobes" de directivité que l'on peut observer en fonction des valeurs du paramètre ka (voir figure 24 et figure 1-3, Annexe 1 pour le tracé de la fonction $J_1(z)/z$) :

- lorsque $ka \leq 1$ – c'est-à-dire lorsque la dimension de la source a est petite devant la longueur d'onde λ –, elle devient proche de 1, et le rayonnement est alors quasi-monopolaire ;
- lorsque $ka \gg 1$ – c'est-à-dire lorsque la dimension de la source a est grande devant la longueur d'onde λ –, de nombreux minima apparaissent et correspondent à des zones de l'espace où le piston ne rayonne pas.

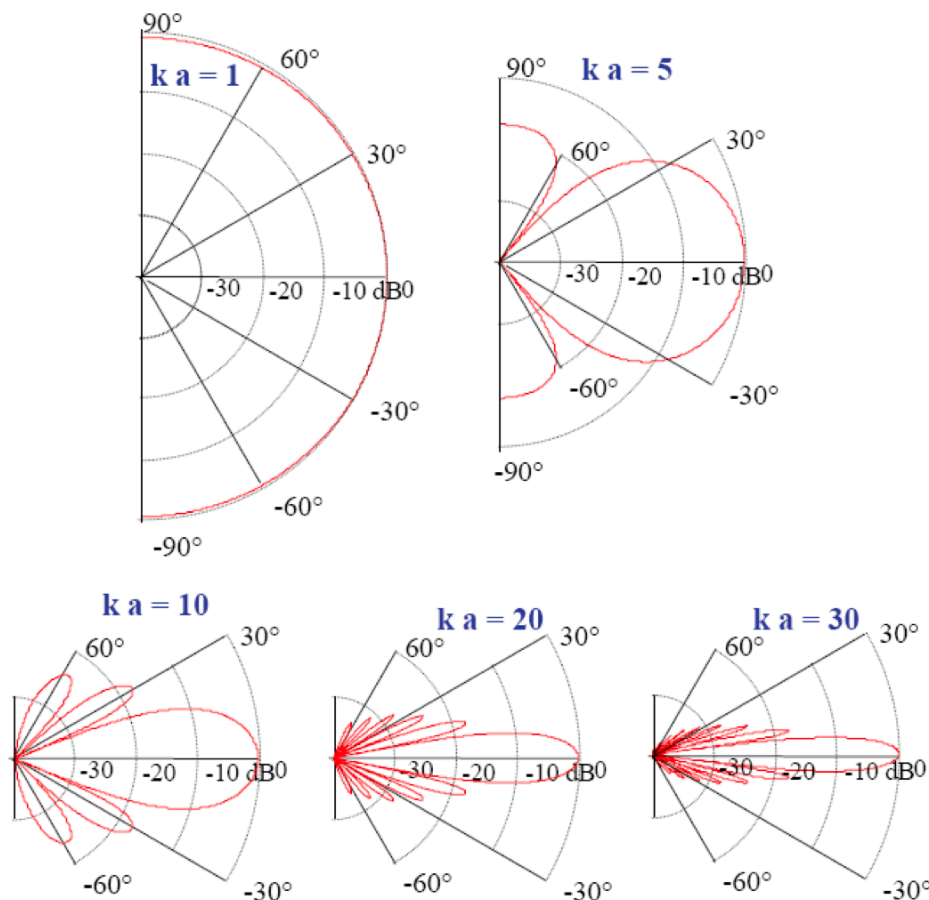


Figure 24. Diagrammes de directivité d'un piston plan en fonction de θ et ka .
(angle de la position d'observation, p/r à l'axe du piston) et ka (k , nombre d'onde ; a , rayon du piston)

Ce modèle constitue une première étape de la compréhension des phénomènes de directivité d'un haut-parleur et permet également d'élaborer des modèles plus complexes faisant intervenir plusieurs haut-parleurs en interaction mutuelle, comme dans le cas d'un réseau linéaire ou tri-dimensionnel.

II.2.2 Description du champ acoustique sur la base des harmoniques sphériques

Les harmoniques sphériques représentent un outil mathématique puissant permettant de décomposer un champ acoustique quelconque sur une base – a priori infinie – de fonctions spatiales normées et orthogonales. Elle permet ainsi de décrire et d'étudier le rayonnement acoustique d'une source à géométrie quelconque.

En effet, les harmoniques sphériques constituent une famille de solutions de l'équation d'onde, dans un repère sphérique (voir figure 25), lorsque les phénomènes d'expansion radiale ne sont pas pris en compte.

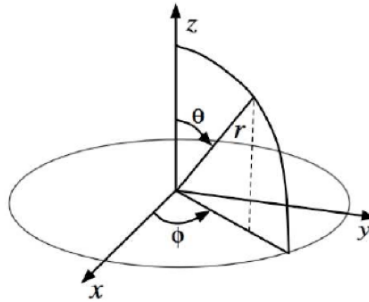


Figure 25. Système de coordonnées sphériques

L'expression générale des solutions de l'équation d'onde en coordonnées sphériques associée à une hypothèse de séparation des variables spatiales et temporelles, amène dans un premier temps à la résolution spatiale de cette équation. Sur la base d'une nouvelle hypothèse de séparation des variables, cette fois-ci, spatiales, les solutions peuvent être exprimées sous la forme d'une combinaison linéaire d'ondes spatiales élémentaires (série de Fourier-Bessel). Enfin, en se limitant à l'expression angulaire de ces solutions – c'est-à-dire en écartant leur composante radiale –, on aboutit à l'expression des harmoniques sphériques Y_n^m sous la forme suivante (voir Annexe 2 pour les détails du calcul) :

$$Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} \cdot P_n^m(\cos\theta) \cdot e^{jm\phi} \quad (\text{Eq. 15})$$

où les fonctions $P_n^m(\cos\theta)$ sont les fonctions associées de Legendre, issues des polynômes de Legendre.

Par ailleurs, les harmoniques sphériques possèdent deux propriétés remarquables :

- symétrie hermitienne (Y^* désignant la fonction complexe conjuguée) :

$$Y_n^{-m}(\theta, \phi) = (-1)^m \cdot Y_n^m(\theta, \phi)^* \quad (\text{Eq. 16})$$

- orthonormalité :

$$\int_0^{2\pi} d\phi \int_0^\pi Y_n^m(\theta, \phi) Y_{n'}^{m'}(\theta, \phi)^* \sin\theta d\theta = \delta_{nn'} \delta_{mm'} \quad (\text{Eq. 17})$$

La figure 26 représente les premiers degrés de la base des harmoniques sphériques réelles, dérivées des fonctions complexes, et qui permettent de décomposer une fonction de directivité réelle.

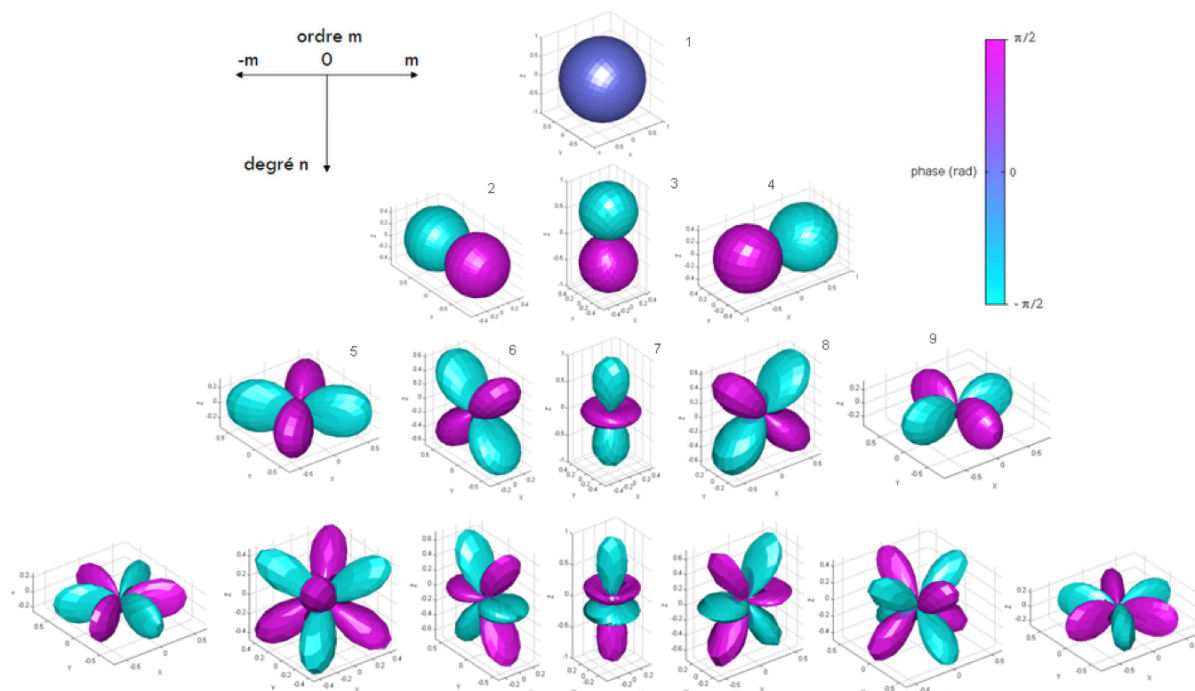


Figure 26. Partie réelle des harmoniques sphériques pour les 4 premiers degrés

II.3 Contrôle du rayonnement acoustique

Cette nouvelle thématique, relative à la diffusion sonore, est apparue dans les années 80. Elle a, d'une part, intéressé le champ de l'acoustique des salles dans l'objectif de concevoir des sources omnidirectionnelles pour effectuer des mesures de salles plus efficaces ; elle a, d'autre part, été étudiée dans le domaine de l'acoustique musicale afin de développer des méthodes capables de reproduire les propriétés spatiales d'un instrument de musique et d'améliorer les systèmes de reproduction sonore existants. Par ailleurs, cette question s'est également posée d'un point de vue musical, avec notamment les idées de Pierre Boulez (énoncées en 1983, lors d'un congrès international d'acoustique) qui développe le fait que « les instruments [...] sont passés au laminoir de l'amplification et perdent une part de leur individualité » et qui formule le constat que « le haut parleur anonymise les sources réelles » ; aboutissant à la conclusion que « [...] nous n'avons encore rien qui puisse vraiment remédier à cette caractéristique haut-parleur » [Boulez, 1983].

II.3.1 Etat de l'art

D'un point de vue acoustique, la question du contrôle de la directivité des systèmes de reproduction est étroitement liée à la mesure et la simulation du rayonnement des instruments de musique. Cette section présente une sélection de travaux réalisés dans le domaine afin d'illustrer la complexité du phénomène acoustique en lui-même mais aussi la complexité méthodologique associée, notamment en termes de mesure de la directivité d'un instrument.

Des expériences effectuées à l'Ircam dans les années 80 par René Caussé et

Gabriel Weinreich mettent en évidence la complexité du rayonnement acoustique de certains instruments et notamment sa nature dispersive (dépendant de la fréquence) en mesurant une fonction de transfert acoustique entre deux directions d'émission autour d'un instrument (voir figure 27).

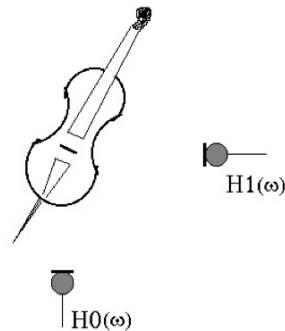


Figure 27. Mesure de la fonction de transfert entre 2 directions d'émission d'un instrument

Les mesures de ce type réalisées sur un trombone et un violon montrent des comportements très différents selon l'instrument :

- pour le trombone, un comportement plutôt "classique", c'est-à-dire omnidirectionnel en basse fréquence (< 800 Hz) puis de plus en plus directif dans l'axe du pavillon à mesure que la fréquence augmente (voir figure 28, gauche) ;
- pour le violon, un comportement relativement erratique et une grande variabilité du rapport rayonnement frontal / latéral à partir d'env. 800 Hz (voir figure 28, droite).

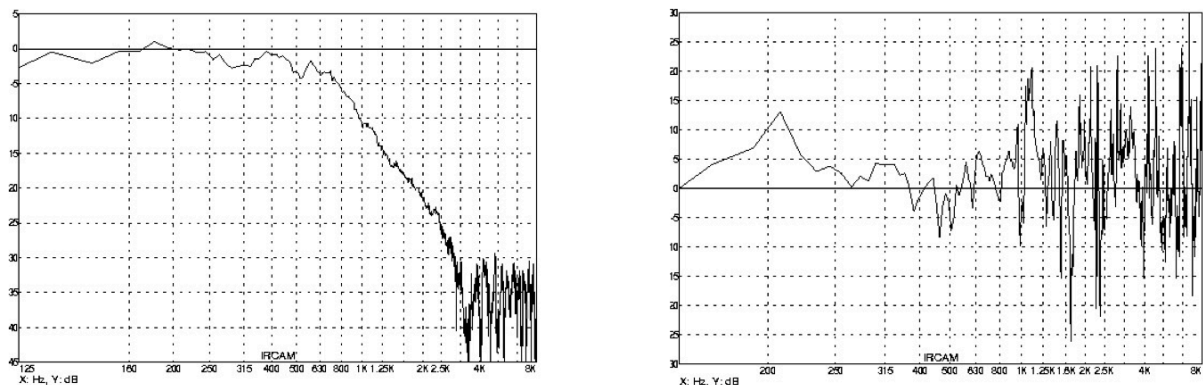


Figure 28. Mesure de la directivité du trombone (gauche) et du violon (droite).

Fonctions de transfert : pour le trombone entre 0° (pavillon) et -120° ; pour le violon, voir figure 27.

En 1983, Garcia [Garcia, 1983] mesure, en chambre anéchoïque, la directivité de plusieurs instruments (violon, trombone, guitare, sitar, cor, trompette, flutes, flute à bec et clavecin) au moyen d'une demi-sphère de microphones (voir figure 29) comportant 9 capsules en présence d'un instrumentiste interprétant une pièce représentative du répertoire de chacun d'eux. Les résultats, par exemple, sur la flute montrent un comportement encore différent des deux instruments précédents avec, entre autres, des différences de rayonnement avant/arrière (micros 1/3) et avant/côtés (micros 1/2 et 1/4) en fonction de la fréquence notamment dues au réseau de trous latéraux par lesquels une partie du rayonnement s'opère.

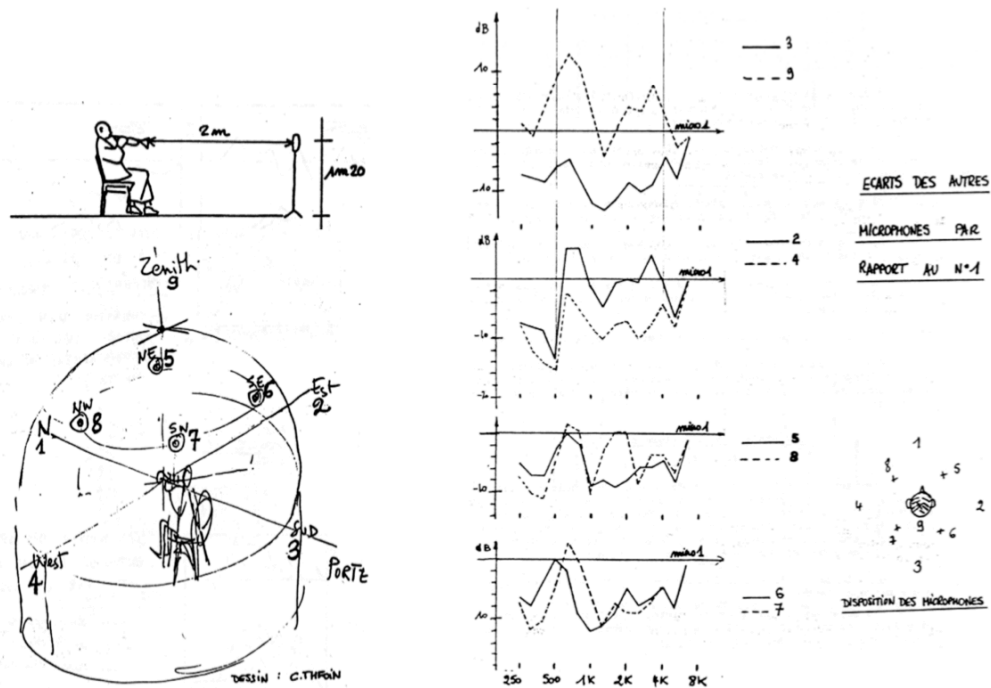


Figure 29. Mesure du rayonnement des instruments selon Garcia [Garcia, 1983]
Dispositif expérimental (gauche) – Fonctions de transfert en fonction du micro 1 pour la flute (droite)

Quelques années plus tard, Weinreich dans une étude de référence mesure le rayonnement du violon par une méthode indirecte exploitant le principe de réciprocité acoustique qui relie le champ acoustique rayonné par une force unitaire exercée sur le chevalet du violon (radiativité) au mouvement de ce même chevalet soumis à un champ acoustique contrôlé [Weinreich, 1997a]. Autrement dit, au lieu de mesurer la pression angulaire rayonnée par l'instrument excité au chevalet, Weinreich soumet le violon à un champ acoustique (ondes planes) provenant d'une direction donnée et mesure la force reçue par le chevalet ; pour des raisons expérimentales, il effectue cette mesure (sur 4 violons différents) suivant quelques directions spécifiques de l'instrument : frontalement par rapport à la table d'harmonie et latéralement par rapport au manche du violon (*neck*) (voir figure 30).



Figure 30. Mesure du rayonnement acoustique par une méthode indirecte [Weinreich, 1997]
Expérience sur le violon (seul le haut-parleur en position frontale par rapport à la table est visible).

Les résultats corroborent ceux présentés plus haut, à savoir un comportement globalement omnidirectionnel jusqu'à 1 kHz environ, puis très fortement dépendant de la fréquence et donc de la note jouée par l'instrument – la sensibilité à ce changement de directivité pouvant être de l'ordre du demi-ton (voir figure 31). Cette caractéristique est appelée par Weinreich « couleur tonale directionnelle » (*directional tone color*) et procure, selon lui, l'illusion qu'au-dessus de ce seuil de 1 kHz, chaque note de l'instrument provient d'une direction différente – ou vu sous un autre angle, « éclaire » la salle de manière différente.

Ce phénomène confère au violon des qualités perceptives propres, notamment dans des situations de jeu comme le vibrato et peut expliquer certaines de ses propriétés comme la « présence » du violon soliste devant l'orchestre au complet (et particulièrement le pupitre des violons) ou bien l'attribut de « projection » généralement utilisé par les violonistes pour évoquer la qualité d'un instrument.

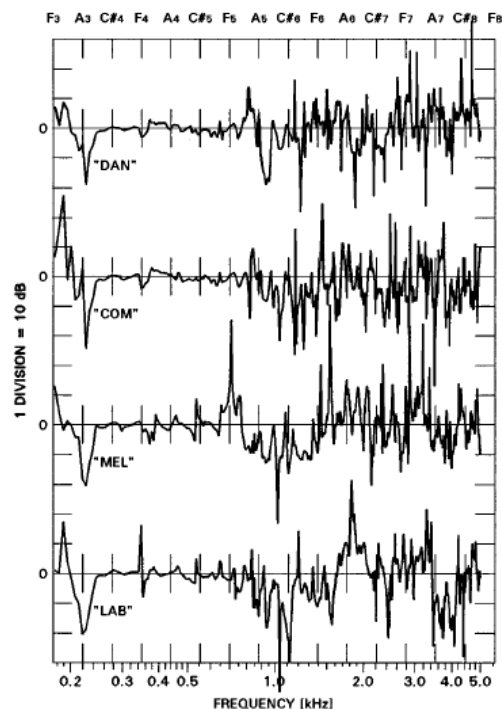


Figure 31. Mesure du rayonnement du violon par la méthode indirecte (données).
Rapport radiativité "latérale"/"frontale", sur 4 violons différents (de haut en bas) [Weinreich, 1997]

Ce travail théorique et expérimental a donné lieu à l'un des premiers prototypes de système de diffusion à directivité contrôlable – précédé néanmoins des expérimentations pionnières effectuées par Mas et Leipp sur la construction d'une enceinte rayonnante utilisant des méthodes semblables à celles utilisées pour la construction d'instrument réel (rosaces, ouïes, pavillons ...) [Mas et al., 1977].

L'application que Weinreich développe en termes de « reproduction électronique », à partir de ses études sur le rayonnement du violon, propose une enceinte constituée d'une caisse en bois (à l'intérieur de laquelle est placé le haut-parleur) reliée à quatre résonateurs verticaux en plastique de longueurs différentes (voir figure 32). Les phases de ces sources interférant entre elles rapidement selon la fréquence, le système produit des sons dont la directivité varie, selon un comportement proche de celui du violon [Weinreich, 1997b].



Figure 32. *DTC loudspeaker* – Dispositif expérimental élaboré par G. Weinreich pour restituer le rayonnement caractéristique d'un signal de violon [Weinreich, 1997b]

En parallèle, Bresciani entreprend des mesures de directivité d'instruments à la fois en chambre anéchoïque et en auditorium (Espace de Projection de l'Ircam) afin de recueillir également la puissance rayonnée par les instruments considérés (cuivres, bois et cordes) [Bresciani, 1996]. Le dispositif expérimental comprend une source mécanique d'excitation (pot vibrant ou moteur de haut-parleur) et une sphère de 20 points de mesure déployés selon un standard – norme ISO 3745⁸ – (voir figure 33).

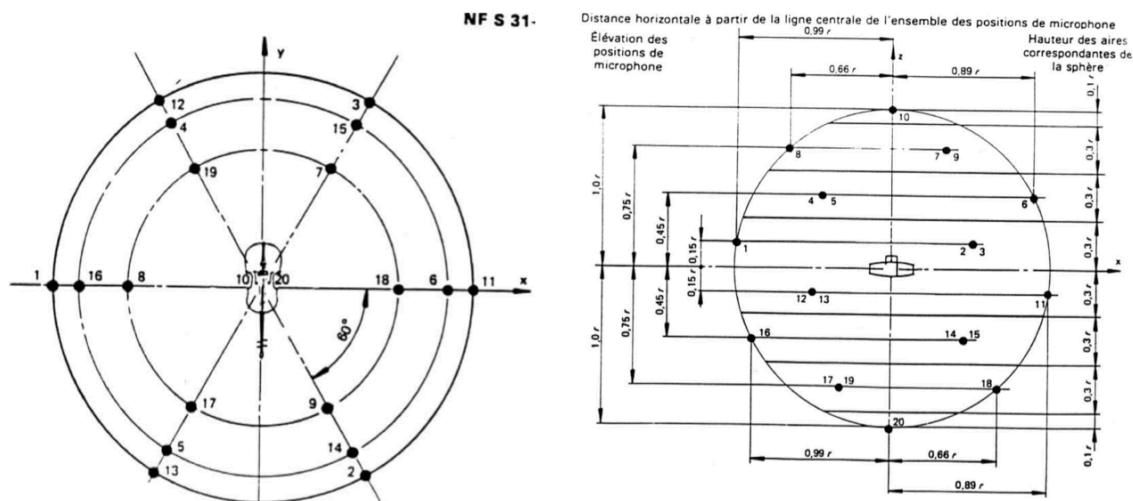


Figure 33. Dispositif de mesure du rayonnement selon Bresciani [Bresciani, 1996]

Les données sont converties, en fonction de la fréquence (1/3 d'octave), en rapport de la pression mesurée (module au carré) sur la puissance totale rayonnée par l'instrument. La figure 34 montre l'expression du ratio pression/puissance pour la clarinette – avec un doigté de type "tous trous fermés" – dans l'hémisphère supérieure (figure 34, gauche) et l'hémisphère inférieure (figure 34, droite). On observe un effet passe-bas global ainsi

⁸ http://www.iso.org/iso/fr/home/store/catalogue_tc/catalogue_detail.htm?csnumber=21750

que l'émergence d'une direction d'émission privilégiée au fur et à mesure que la fréquence augmente, qui correspond certainement à l'axe du pavillon de l'instrument (micro #18)

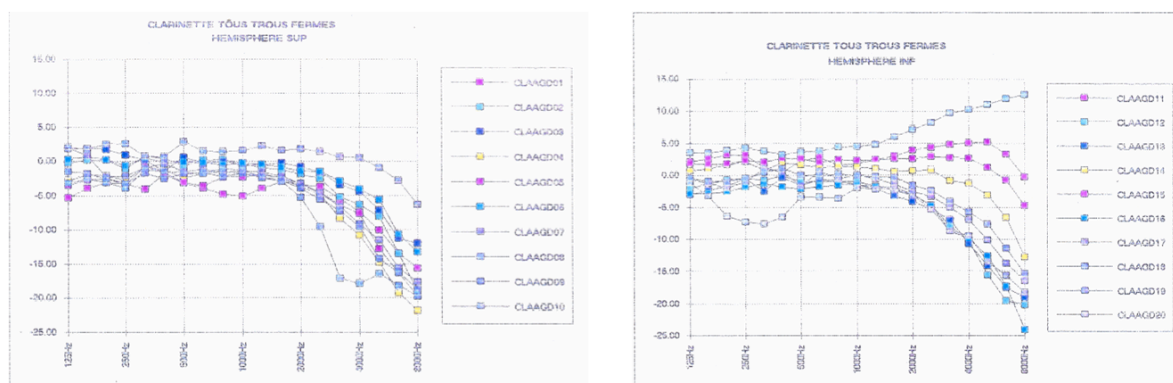


Figure 34. Mesure de la directivité de la clarinette pour le doigté "tous trous fermés" [Bresciani, 1996]
Rapport du module au carré de la pression par la puissance totale rayonnée (1/3 octave)
Droite : micros de l'hémisphère sup. (1 à 10) - Gauche : micros de l'hémisphère inf. (11 à 20)

A partir de ces mesures, Bresciani et coll. instancient une première approche prototype de reproduction sonore à directivité contrôlée, au moyen d'un système dodécaédrique multi haut-parleurs dont les 12 voies sont indépendantes (voir figure 35). Cette base est ensuite reprise par Derogis dans le cadre de son étude sur le rayonnement de la table d'harmonie de piano et la conception d'un système de reproduction du champ acoustique [Derogis, 1997] (voir § II.3.3.2.1 pour la présentation du principe). La poursuite de cette axe de recherche fera l'objet du développement du projet Timée par Misdariis, Warusfel et Caussé, et la conception de nouveaux dispositifs [Misdariis et al., 2002] (voir § II.3.3, pour les explications de ce projet).

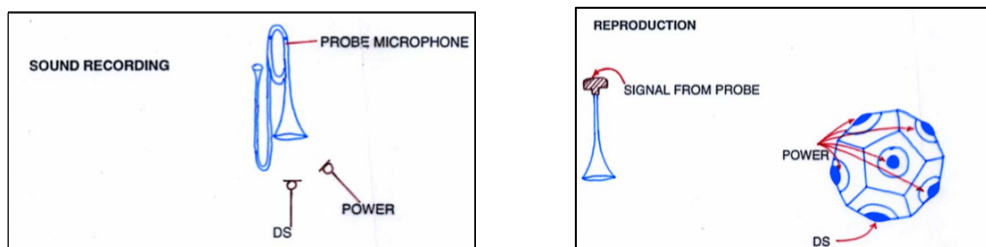


Figure 35. Expérience de Bresciani – Reproduction du trombone par un haut-parleur dodécaédrique à directivité contrôlée [Bresciani, 1996]

En termes de mesure de rayonnement, Pätynen et Lokki ont entrepris une approche expérimentale à grande échelle de la directivité des instruments [Pätynen et al., 2010]. Ils mesurent indépendamment 14 instruments courants de l'orchestre symphonique (+ musicien) appartenant aux familles des bois, cuivres, cordes et percussions – ainsi qu'une voix chantée –, à différentes nuances de jeu (*piano*, *forte* et *fortissimo*). Le dispositif technique est un système tri-dimensionnel à 22 micros placé en chambre anéchoïque, le musicien jouant une pièce représentative du répertoire de l'instrument considéré.

Les conclusions générales de l'étude sont que, comme prévu, la directivité des bois et des cordes est beaucoup plus complexe que celle des cuivres qui rayonnent globalement en direction du pavillon. En particulierisant par famille d'instruments :

- le hautbois et la clarinette présentent des directivités relativement semblables alors que le basson rayonne de manière beaucoup plus désordonnée ;
- les cuivres (tuba, trompette, trombone et cor) présentent des directivités quasi-communes allant de "omnidirectionnel" en basse fréquences à "directif" en hautes fréquences dans la direction de leur pavillon respectif ;
- les cordes rayonnent de manière encore plus complexe que les instruments à vent et présentent globalement une ou deux bandes de fréquences propres (selon les instruments) en termes de directivité ; par exemple, le violon possède une fréquence de coupure basse (500 Hz) en-deçà de laquelle sa directivité est globalement omnidirectionnelle et au-delà de laquelle celle-ci évolue rapidement en fonction de la fréquence.

Enfin, de manière générale, si la dynamique influence de manière significative la plupart des timbres de ces instruments (surtout dans le cas des cuivres), elle ne semble pas affecter notablement les propriétés de rayonnement acoustique.

Il est, par ailleurs, important de noter que la connaissance exhaustive du rayonnement des instruments telle que l'ont développée Pätynen et Lokki dans leur étude peut avoir des retombées intéressantes dans le domaine de la simulation de l'acoustique des salles et les techniques d'auralisation associées à des moteurs de rendu d'acoustique virtuelle comme Odeon⁹ [Otondo, 2005] ou le Spatialisateur^{©10} de l'Ircam [Jot et al., 1995].

Dans la même idée et en y adjoignant un développement en termes de contrôle de la directivité, Belher et Pollow (ITA, Aachen) travaillent sur les haut-parleurs sphériques pour la conception d'une source permettant d'évaluer les qualités acoustiques d'une salle, en étudiant plusieurs méthodes de synthèse d'une directivité cible et d'évaluation des erreurs associées [Behler, 2007] puis proposent une méthode d'optimisation pour la décomposition sur la base des harmoniques sphériques [Pollow et al., 2009] ; ils élaborent ensuite une base de données de directivité d'instruments utilisable pour améliorer les paradigmes de simulation acoustique et les mesures de réponses impulsionnelles associées [Pollow et al., 2010].

Dans une approche plus applicative et moins contrainte en termes de reproduction du rayonnement, Wessel et coll. (CNMAT) ont imaginé un système de haut-parleurs localisé possédant une fonction de rayonnement « programmable » ; ce dispositif devenant, selon Wessel, une source sonore associée à l'interprète de musique informatique [Wessel, 1991]. Afin de formaliser l'idée sur une structure multi-faces, Kassakian et Wessel développent une méthode générale de calcul de filtres permettant de tenir compte des contraintes de chaque face, de la géométrie arbitraire du système ainsi que des caractéristiques mesurées de chaque haut-parleur [Kassaskian et al., 2003] , Par la suite, ils caractérisent de manière systématique la famille des réseaux sphériques de haut-parleurs (*spherical loudspeaker arrays*) conçus sur la base des polyèdres réguliers (cube, dodécaèdre, icosaèdre et polyèdre à 60 faces). Au moyen d'une simulation, ils déterminent l'erreur de reproduction des premières harmoniques sphériques (jusqu'au 6^{ème} degré) en fonction de paramètres structuraux du réseau sphérique (type / nombre de haut-parleurs, diamètre et distance inter-haut-parleurs). Ils valident enfin leurs estimations sur un dispositif dodécaédrique construit pour l'occasion [Kassakian et al., 2004]. Plus récemment, Avizienis, Freed, Kassakian et Wessel ont mis au point le design d'une source à 120 haut-parleurs sur une structure

⁹ <http://www.odeon.dk>

¹⁰ <http://support.ircam.fr/docs/spat/3.0/spat-3-intro/co/overview.html>

icosaédrique en tentant de répondre à la question suivante : comment maximiser le nombre de haut-parleurs (notamment pour améliorer les performances de la source en hautes fréquences) tout en gardant une dimension "correcte" du dispositif ? L'étude décrit les méthodes de multiplexage, filtrage et amplification appliquées à un grand nombre de voies [Avizienis et al., 2006]. Enfin, sur la base de ces différents dispositifs, Schmeder explore les possibilités offertes en termes d'expérience-utilisateur et d'interactivité en détaillant des cas d'étude potentiels dans des domaines tels que les Interfaces Homme-Machine (pointage/sélection audio), le contrôle haptique (couplage avec des capteurs) ou encore la performance scénique (installation sonore, concert) [Schmeder, 2009].

Dans la même idée que les récents travaux du CNMAT, et partant des travaux théoriques de Zotter et Noisternig (IEM, Graz) sur la simulation et la modélisation de figures de directivité par un réseau de haut-parleurs sphérique [Zotter et al., 2007a], Pomberger développe une étude autour du contrôle angulaire et radiale de la directivité appliquée à une source icosaédrique multivoies [Pomberger, 2008]. Dans le même temps, Zotter et coll. effectuent des travaux connexes concernant notamment la description analytique du système en modélisant chaque haut-parleur par une capsule sphérique vibrante [Zotter et al., 2007b], l'extension du modèle en tenant compte des couplages acoustiques internes et externes entre haut-parleurs [Zotter et al., 2007c], la description analytique de filtres entrées/sorties multiples (MIMO – *Multiple Input Multiple Output*) permettant de minimiser le couplage entre haut-parleurs partageant un même volume [Zotter et al., 2008a] et la validation du calcul des filtres MIMO à une source icosaédrique en utilisant une antenne microphonique pour mesurer la directivité des haut-parleurs [Zotter et al., 2008b].

D'autres approches du problème sont proposées, notamment par Pasqual et Arruda (Universidade Estadual de Campinas, Brazil) associés à Herzog (LMA, Marseille) qui s'intéressent à l'effet du couplage interne de l'enceinte d'un système de haut-parleurs sphérique effectuant une comparaison analytique de sources avec et sans volume partagé (et avec / sans matériaux absorbants) [Pasqual et al., 2008a]. Par ailleurs, en parallèle d'une étude comparative des 5 premiers polyèdres réguliers (« solides platoniciens ») sur la capacité de reproduction des harmoniques sphériques [Pasqual et al., 2010a], ils développent une méthode de calcul de gains associés à chaque haut-parleur à partir d'une minimisation itérative en négligeant la phase du rayonnement cible et comparent l'erreur de modélisation ainsi produite avec les résultats obtenus par des méthodes classiques [Pasqual et al., 2008b]. Ils modélisent aussi le comportement d'un haut-parleur sphérique par une sphère rigide sur laquelle est monté un ensemble de capsules vibrantes et proposent une méthode alternative (aux harmoniques sphériques) pour décrire la directivité du système basée sur les modes de rayonnement de cette « sphère discrète » (*discrete sphere*) qui permet, entre autres, de calculer l'efficacité de rayonnement de chaque haut-parleur et d'optimiser ainsi le nombre de voies en fonction de la fréquence émise (certains modes ne rayonnant pas d'énergie dans certaines plages de fréquences) [Pasqual et al., 2010b]. Cette alternative aux harmoniques sphériques est également explorée par Khaykin et Rafaely en considérant une base de description hybride associant les harmoniques sphériques à un ensemble de fonctions relatives aux signaux d'alimentation de chaque haut-parleur au moyen d'un procédé mathématique de décomposition en valeurs singulières (SVD – *Singular-Value Decomposition*) [Khaykin et al., 2010].

Dans une perspective à nouveau plus applicative et pragmatique, Trueman et Cook (Princeton Univ.) proposent une « voix alternative pour les sons électroniques » sous la forme d'un haut-parleur sphérique associé à différents capteurs de données permettant de piloter des propriétés de diffusion de la source. Plusieurs versions du haut-parleur sphérique – dont l'un hémisphérique – sont utilisées dans diverses applications interactives impliquant une performance de musique électronique ou mixte où, à l'image de l'idée initiale de Wessel, le dispositif de reproduction devient une prolongation de l'instrument et se contrôle de manière musicale. [Trueman et al., 2000]. Le paradigme de reproduction du rayonnement se limite ici, le cas échéant, à une mesure de l'espace acoustique d'un instrument selon un schéma d'échantillonnage spatial (localisation des micros) en relation directe avec la structure discrète du système de reproduction (sphère de mesure répartie de la même façon que les points sources du haut-parleur sphérique) ; le traitement – et le contrôle – se faisant en assignant un filtre à chaque voie de mesure/reproduction [Cook et al., 1998a]. Suivant cette méthode, une base de données de plusieurs instruments est établie et les filtres correspondants (pour un système donné) sont calculés pour être exploités en situation de jeu [Cook et al., 1998b].

A noter que, dans cette approche plutôt centrée sur les nouvelles interfaces musicales, d'autres dispositifs de projection de son existent comme le « Hemisphere » utilisé par Cook et Trueman au sein d'« orchestres d'ordinateurs portables (*laptop*) » (Plork¹¹ /Slork¹² – *Princeton / Stanford Laptop Orchestra*) ou l'« Orbophone » développé par Lock et Schiemer [Lock et al., 2006].

II.3.2 Etat des lieux

II.3.2.1 Système multi haut-parleurs pour simulation de source omnidirectionnelle

D'une manière générale, les deux principaux champs d'utilisation des sources omnidirectionnelles sont, d'une part, le calcul du temps de réverbération des salles pour leur caractérisation acoustique, et d'autre part, l'évaluation de l'isolation acoustique garantie par les éléments d'un bâtiment séparant deux pièces adjacentes.

La mesure du temps de réverbération correspond au temps de décroissance de la pression acoustique de 60 dB par rapport à une valeur maximum initiale. Pour effectuer cette mesure, il est nécessaire de disposer d'une source capable de rayonner de manière homogène avec suffisamment de puissance. La plupart des sources omnidirectionnelles proposées par les professionnels sont des sources sphériques (ou bâties sur un polyèdre régulier) contenant plusieurs haut-parleurs alimentés en phase ; elles sont par ailleurs soumises à la norme ISO 3382 [Leishman et al., 2006].

Si les différentes sources se différencient essentiellement par leur puissance de sortie et la qualité de leur propriété omnidirectionnelle, certains constructeurs donnent la possibilité supplémentaire de séparer les canaux diffusion afin de reproduire des directivités typiques (p. ex., dipôle par opposition de phase). Parmi eux Bruel & Kjaer [Tarnow, 1974], Euphonia qui développe la source « Polyfusor » ou Outline qui développe la « *Global Source Radiator* » (GSR).

II.3.2.2 Système multi haut-parleurs pour contrôle et reproduction du rayonnement

Sur la base des nombreux travaux entrepris dans le domaine et détaillés dans la

¹¹ <http://plork.cs.princeton.edu>

¹² <http://slork.stanford.edu>

section "Etat de l'art" (§ II.3.1), un certain nombre de laboratoires de recherche ont développé des systèmes prototypes de sources multi haut-parleurs à des fins d'expérimentation mais aussi, pour certains, de performances musicales ou sonores. Parmi ces réalisations, on trouve notamment :

- la source icosaédrique du **CNMAT** (Wessel, Freed, Avizienis, Schmeder et coll.) possédant 120 haut-parleurs isolés acoustiquement et alimentés/amplifiés indépendamment ; elle a été fabriquée par la société Meyer Sound (figure 36).

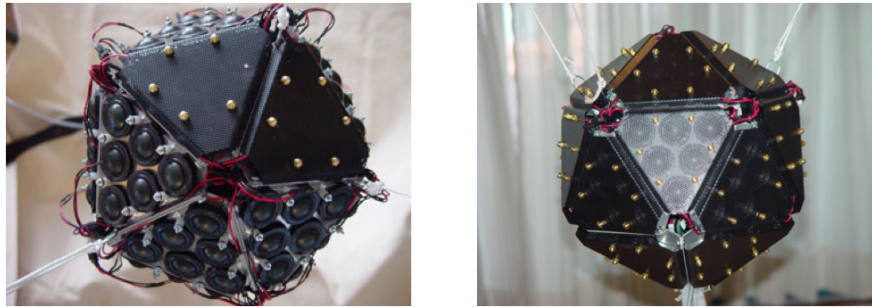


Figure 36. Source icosaédrique à 120 voies développée par le CNMAT

- la source icosaédrique de l'**IEM** (Zotter, Noisternig, Pomberger et coll.) possédant 1 caisson basse fréquence + 20 x2 voies déployées respectivement sur les fréquences médiums et aigues ; chaque groupe de 20 haut-parleurs partage le même volume d'enceinte (figure 37).
- la source dodécaédrique de l'**ITA** (Behler, Pollow et coll.) possédant 1 caisson basse fréquence + 12 x2 voies déployées respectivement sur les fréquences médiums et aigues ; chaque groupe de 12 haut-parleurs partage le même volume d'enceinte (figure 38).



Figure 37. Source icosaédrique à 20 x2 voies (+ 1 caisson BF) développée par l'IEM



Figure 38. Source dodécaédrique à 12 x2 voies (+ 1 caisson BF) développée par l'ITA

- les sources cubiques de l'**Ircam** (Bresciani, Derogis, Warusfel, Caussé, Misdariis et coll.) possédant respectivement 6 voies large-bande et 1 caisson de basse + 6 x2 voies déployées respectivement sur les fréquences médiums et aigues ; chaque groupe de 6 haut-parleurs de chaque cube partage le même volume d'enceinte. La dernière version du cube multi-voies a été fabriquée par la société Focal-JMLab (figure 39).



Figure 39. Source cubique à 6 voies – gauche – et cubique à 12 x2 voies (+ 1 caisson BF) – droite – développée par l'Ircam

II.3.3 Système multi haut-parleurs à directivité contrôlable : projet *Timée*

II.3.3.1 Préambule

A partir d'études précédentes relatives à la reproduction du son par un système multi haut-parleurs [Bresciani, 1996 ; Derogis, 1997] (voir section § II.3.1), une nouvelle méthode a été développée ; elle s'appuie sur le principe de décomposition du champ acoustique sur la base des harmoniques sphériques et sur un réseau tri-dimensionnel de haut-parleurs de type "sphérique" pour constituer un ensemble de directivités de base. Cette architecture permet de contrôler la nature et la direction du rayonnement et, grâce à des optimisations, permet une gestion en temps réel et simultanée de plusieurs sources. Une maîtrise indépendante des caractéristiques spectrales et spatiales offre, en outre, des possibilités intéressantes pour la reproduction des instruments de musique et permet d'entrevoir des perspectives d'applications dans le domaine de la synthèse sonore. Enfin, cette approche – impliquant d'une manière nouvelle le système de diffusion sonore dans l'œuvre musicale – traduit une volonté de repenser le rapport entre espace et musique et de renouveler le dialogue entre instruments acoustiques et partie électronique qui sont souvent amenés à "combattre" dans les productions musicales contemporaines [Nicolas, 2000].

Cette étude est également le fruit d'une étroite collaboration avec le compositeur François Nicolas¹³, inscrite dans le cadre d'une position de compositeur en recherche et qui a abouti à la création d'une œuvre pour ensemble instrumental et électronique exploitant les résultats théoriques et techniques obtenus (« Duelle », œuvre pour piano, voix, violon et électronique diffusée par la *Timée* – François Nicolas, assisté de Eric Daubresse, création Festival Agora 2001, Ircam)¹⁴. Ces résultats ont également été utilisés dans le cadre de plusieurs installations artistiques sonores réalisées par le compositeur/designer sonore Louis Dandrel.

¹³ <http://brahms.ircam.fr/francois-nicolas>

¹⁴ <http://brahms.ircam.fr/works/work/27682/>

Elle a, par ailleurs, fait l'objet de diverses communications scientifiques aux 140^{ème} et 150^{ème} meeting de l'*Acoustical Society of America* (ASA) [Warusfel et al., 2000 ; 2004a], au 4^{ème} congrès *Digital Audio Effects* (DAFx'01) [Warusfel et al., 2001], au *International Symposium on Musical Acoustics 2001* (ISMA) [Misdariis et al., 2001a], au 27th *International Computer Music Conference* (ICMC) [Misdariis et al., 2001b], aux 9^{ème} Journées d'Informatique Musicale (JIM) [Misdariis et al., 2002], à la 116^{ème} convention de l'*Audio Engineering Society* (AES) [Warusfel et al., 2004b] ainsi qu'au 18th *International Congress on Acoustics* (ICA) [Warusfel et al., 2004c].

II.3.3.2 Publication [Misdariis et al., Journées d'Informatique Musicale, 2002]

Le contrôle de la directivité par un système multi haut-parleurs – Applications

Misdariis Nicolas, Warusfel Olivier, Caussé René, Nicolas François
IRCAM – 1, place Igor Stravinsky - 75004 Paris - France

II.3.3.2.1 Introduction

Le contrôle de la directivité représente un nouvel enjeu dans le processus de diffusion des sons. En effet, si les technologies actuelles permettent un ajustement précis des caractéristiques de timbre et de niveau sonore – le cas échéant, au moyen de modules de traitement de signal numérique –, elles n'offrent que très peu de latitude en terme de contrôle des données spatiales (directivité) et ne permettent donc pas d'imposer un rayonnement adapté au signal diffusé par le système. A fortiori, dans un espace clos tel qu'une salle de concert, ces propriétés de rayonnement jouent un rôle fondamental puisqu'elles déterminent la distribution temporelle et spectrale de l'onde acoustique arrivant aux oreilles de l'auditeur : c'est, en partie, ce qui permet de mieux distinguer un piano réel et un son de piano diffusé par une enceinte, mis à part les problèmes de distorsion et de bande-passante inhérents au système de haut-parleurs.

Le contrôle de la directivité implique donc une nouvelle architecture, basée sur le principe des réseaux multidimensionnels de transducteurs. Si plusieurs voies ont déjà été explorées à ce sujet, elles ont plutôt eu pour objectif commun de s'affranchir de l'effet de salle en focalisant l'information sonore en un point de l'espace, soit par des systèmes hyper-directifs [Jacob et al., 1990] [Pompei, 1999], soit par des systèmes holophoniques [De Vries, 1996]. La présente étude propose une démarche différente dans la mesure où elle vise justement à restituer, ou du moins contrôler, la manière dont les sources sonores interagissent avec la salle.

II.3.3.2.2 Méthodes de reproduction

II.3.3.2.2.1 Approche "statique"

Le cadre théorique générale de cette étude a fait l'objet d'un travail de thèse [Derogis, 1997] [Warusfel et al., 1997] et peut se résumer de la manière suivante.

Considérons une source acoustique T , délimitée par un domaine D , et rayonnant une pression $T(\vec{r}, \omega)$ dans un domaine V extérieur à D . Les lois de la propagation acoustique nous permettent d'affirmer que le rayonnement de la source T est entièrement déterminé par la connaissance de la pression (amplitude et phase) sur une surface arbitraire S entourant D (figure 40).

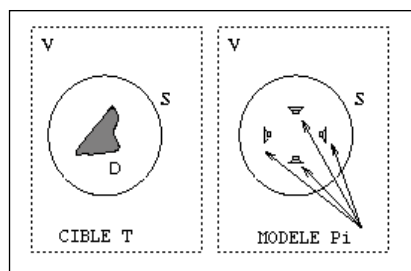


Figure 40. Reproduction du rayonnement (approche "statique")

De fait, si l'on remplace la source T par une source P générant exactement le même champ de pression sur la surface S , on est alors assuré que le champ rayonné par les deux sources seront identiques dans tout l'espace extérieur V . Si de plus, la source P est une source composite constituée de N sources élémentaires produisant chacune le champ de pression P_i , la condition précédente peut se traduire par une décomposition de la pression T sur l'ensemble des $\{P_i\}$ suivant la relation :

$$\sum_{i=1}^N a_i(\omega) \cdot P_i(\vec{r}_0, \omega) = T(\vec{r}_0, \omega) \quad , \forall \vec{r}_0 \in S \quad (\text{Eq. 18})$$

où les $\{a_i\}$ sont des valeurs complexes dépendant de la fréquence ω .

Ceci étant, puisque les $\{P_i\}$ ne constituent pas, a priori, une base du problème, l'existence et l'unicité des $\{a_i\}$ ne sont pas assurées et la solution de l'équation (1) ne peut être que le résultat d'une approximation. L'une des méthodes d'optimisation consiste à interpréter cette équation comme une distance algébrique entre la source-modèle P et la source-cible T , les coefficients $\{a_i\}$ étant alors obtenus numériquement, par minimisation de cette grandeur.

Il en résulte un jeu de coefficients complexes dépendant de la fréquence – c'est-à-dire des filtres – qui, appliqués respectivement à chaque source P_i , reproduisent au mieux le champ de pression visé. Par nature même de la méthode, les relations d'amplitude et de phase entre les filtres doivent être scrupuleusement respectées, c'est pourquoi leur implémentation comporte un module commun d'égalisation (IIR) et un partie résiduelle de filtrage (FIR) capable de modéliser ces différences mutuelles (figure 41).

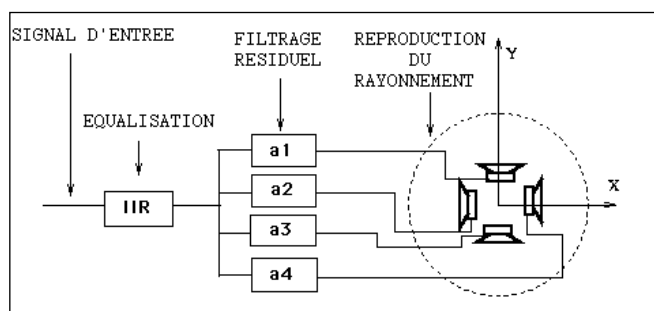


Figure 41. Implémentation du traitement du signal (approche statique)

II.3.3.2.2 Approche "modale"

Une approche dérivée consiste à relâcher les contraintes sur la fidélité de reproduction au profit d'une plus grande pertinence musicale (facilité de contrôle, interactivité, continuité, ...), induisant du même coup des problèmes d'interpolation de figures de

directivité, et donc de filtres. Pour ce faire, nous proposons une nouvelle méthode d'implémentation basée sur une simplification du principe précédent.

Au lieu de considérer des directivités indépendamment les unes des autres, on constitue un ensemble de directivités élémentaires qui, combinées entre elles, permettront de construire une fonction de rayonnement, avec un degré d'approximation donné. Ces "entités de rayonnement" élémentaires sont choisies dans la base des harmoniques sphériques ce qui confère à la méthode son aspect "modale" puisque tout champ de pression est une combinaison linéaire infinie des vecteurs de cette base et que, dans ce cas, précis, la décomposition se limite aux premiers "modes" de propagation.

En pratique, les directivités de base sont le monopole (H0) ainsi que le dipôle (H1) et l'une des harmoniques sphériques d'ordre 2 (H2), dans les trois directions de l'espace (figure 42, gauche). La combinaison des ces figurent fournissent un contrôle du rayonnement selon deux critères : la nature et la direction d'émission. Par ailleurs, les problèmes d'interpolation de filtres sont en partie résolus par le fait que toutes les directivités élémentaires sont en permanence générées par le système, le contrôle ne s'opérant uniquement qu'au moyen de gains pondérateurs. Cette description synthétique possède, en outre, l'avantage de pouvoir traiter en parallèle plusieurs signaux d'entrée sans multiplier le coût de calcul : chaque directivité étant assignée par un jeu de gains différent, avant mixage de tous les signaux (figure 42, droite).

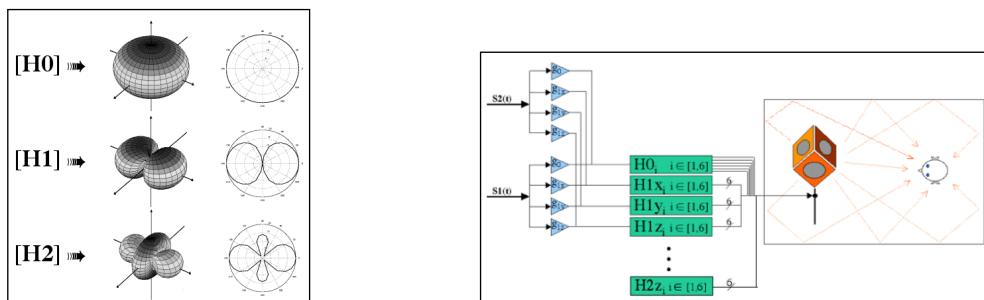


Figure 42. Approche "modale" : figures de directivité (gauche) et nouvelle implémentation (droite)

Afin de valider la méthode, plusieurs prototypes ont été conçus sur la base d'une structure cubique équipée d'un haut-parleur indépendant sur chacune des faces. De plus, pour tenter de minimiser le phénomène de repliement spatial inhérent aux réseaux de transducteurs, un prototype multi-voies a également été élaboré et constitue le développement le plus récent du projet. Dans ce cas précis, la structure cubique est dupliquée trois fois dans les dimensions relatives aux basses, moyennes et hautes fréquences (figure 43), ce qui permet à chacune des sous-structures de travailler quasiment toujours en-dessous de la fréquence de coupure au-delà de laquelle les lobes de directivité parasites apparaissent.

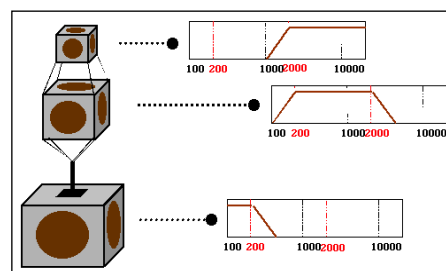


Figure 43. Approche "modale" : prototype multi-voies

II.3.3.2.2.3 Résultats.

Les performances des prototypes sont examinées en terme d'égalisation, de fidélité de reconstruction du champ sonore et de caractérisation objective du point d'écoute.

Des premières mesures en champ diffus atteste d'une égalisation en puissance satisfaisante pour le système multi haut-parleurs (figure 44). Ce résultat a pour principale conséquence de considérer la source spectralement "neutre" vis-à-vis du signal d'entrée et donc d'assurer une cohérence de timbre quelque soit le rayonnement imposé : dès lors, les modifications spectrales perçues entre deux directivités différentes (par exemple, le monopole et le dipôle), sont uniquement dues aux propriétés acoustiques du rayonnement.

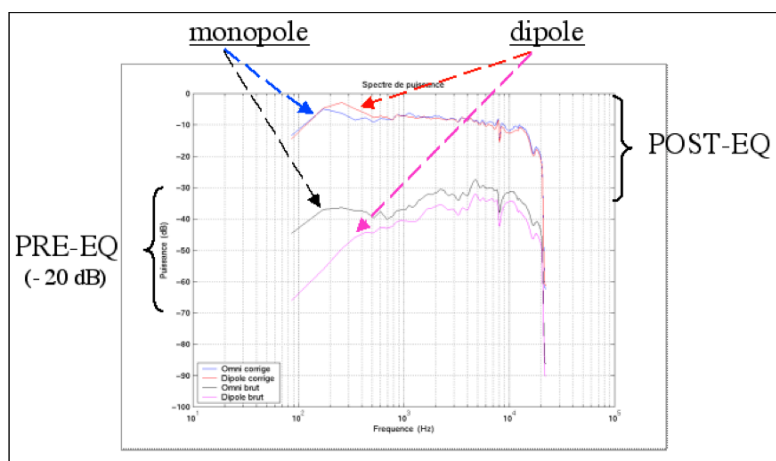


Figure 44. Prototype multi-voies : égalisation en puissance de H0 et H1

Des mesures plus complètes en chambre sourde, nous permettent ensuite de mesurer la qualité de reproduction des figures élémentaires (figure 45 : monopole ; figure 46 : dipôle) sur une grande partie de la bande passante du système. Les résultats sont également satisfaisants et la comparaison entre les différents prototypes fait apparaître le gain obtenu grâce à l'approche multi-voies qui repousse effectivement les limites du repliement – et donc de la dégradation de la reproduction – en dehors de la bande passante utile de chacune des sous-structures.

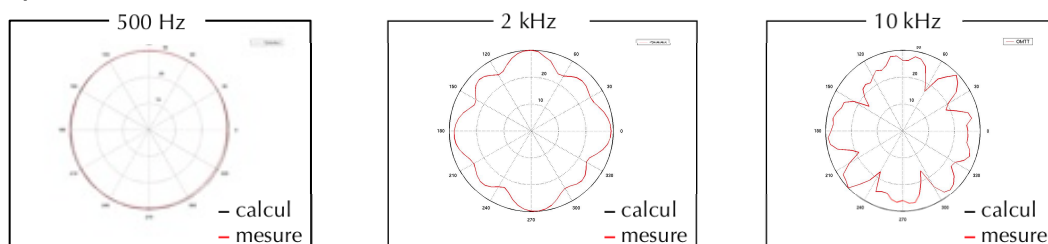


Figure 45. Reproduction de H0 à 500 Hz (gauche), 2000Hz (centre) et 10000 Hz (droite)

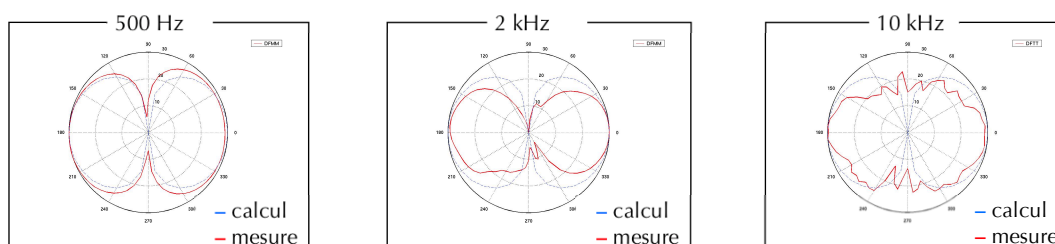


Figure 46. Reproduction de H1 à 500 Hz (gauche), 2000Hz (centre) & 10000 Hz (droite)

Enfin, une caractérisation objective au point d'écoute a été tentée afin d'interpréter physiquement les effets spatiaux produits et de vérifier l'efficacité des figures de directivité élémentaires. Pour ce faire, les réponses impulsionnelles de chacune d'entre elles sont mesurées et analysées en terme de distribution spatio-temporelle de l'énergie acoustique.

La figure 47 (gauche) représente l'évolution temporelle de l'énergie frontale mesurée avec un microphone cardioïde pointant la source : cette première analyse met nettement en évidence la carence tardive de l'énergie frontale pour le dipôle vertical et latéral. La figure 47 (droite), quant à elle, montre le rapport de l'énergie frontale à l'énergie latérale, celle-ci étant mesurée au moyen d'un microphone bi-directionnel. De manière prévisible, le dipôle frontal donne la meilleure localisation de la source, supporté en cela par les premières réflexions. Le dipôle latéral présente, au début de la réponse, un rapport faible entre énergies frontale et latérale ce qui a pour conséquence de produire une sensation de source élargie et d'enveloppement de la salle. Quant au dipôle vertical, en dépit d'un faible son direct, il fournit la sensation de source étroite à cause d'un équilibre tardif entre l'énergie frontale et l'énergie latérale. Notons, par ailleurs, que toutes ces propriétés sont aisément détectables à l'écoute, robustes à la reproduction et faiblement dépendant de la salle : ils permettent donc d'assigner des identités perceptives pertinentes aux sons produits ou diffusés par le dispositif.

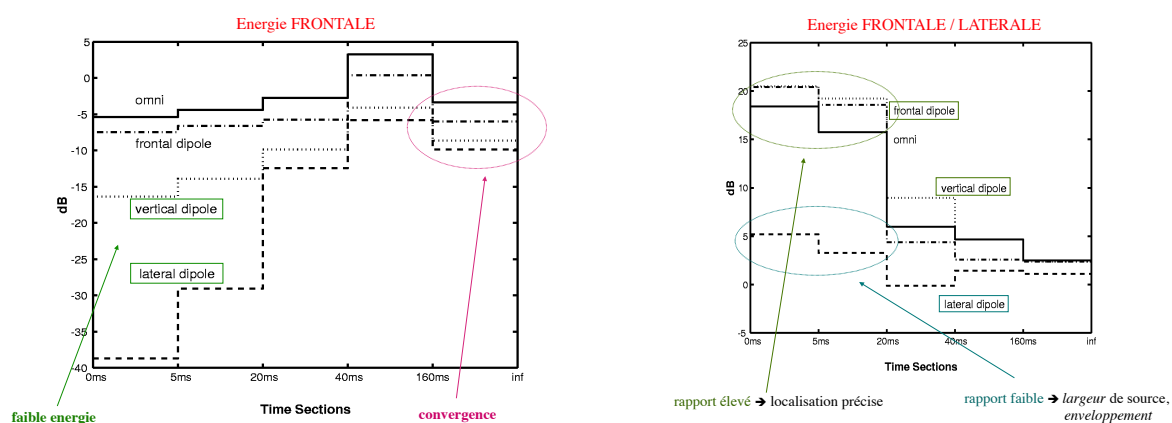


Figure 47. Prototype multi-voies : énergie frontale (gauche), énergies frontale/latérale (droite)

II.3.3.2.3 Une nouvelle instrumentalisation de la musique mixte

II.3.3.2.3.1 Problématique.

Dans le contexte actuel de la musique, une grande partie des œuvres tendent à faire co-exister des sons acoustiques provenant d'instruments de musiques *live* et des sons électroniques synthétiques ou transformés qui sont diffusés par un système électro-acoustique. De plus, dans la plupart des cas, les instruments sur scène – y compris les voix – doivent être eux-mêmes amplifiés pour rivaliser en puissance avec la partie électronique. Enfin, le système de diffusion est généralement constitué d'une ceinture de haut-parleurs entourant l'auditoire, imposant une acoustique propre, en s'affranchissant totalement de la salle existante.

Pour toutes ces raisons, la cohésion entre les parties acoustiques et électroniques d'une même œuvre musicale peut apparaître difficile, voire illusoire. C'est pourquoi, l'approche proposée dans la présente étude visant à considérer un système de diffusion sonore localisé, jouant avec l'espace architectural plutôt que le contraignant, et qui, à l'image d'un instrument de musique, possède des propriétés de rayonnement

spécifiques – et dans le cas présent, modulables – peut sembler une alternative à cette problématique de la musique mixte.

Possibilités sonores et musicales du dispositif.

Dans cette optique, une collaboration étroite avec le compositeur François Nicolas a permis d'explorer les potentialités sonores et musicales de la méthode ; l'un des points-clés du travail étant de vérifier si les attributs spatiaux liés au rayonnement des sources sonores pouvaient constituer un vocabulaire musical à part entière.

Pour cela des outils spécifiques (enveloppes temporelles, fonctions aléatoires, contraintes, ...) ont été développés afin de tester les différentes possibilités – aussi bien statiques que dynamiques – d'utiliser le système. Ainsi, de nombreux exemples ont été créés dans le but de formaliser, et dans la mesure du possible notifier, les effets auditifs obtenus dont voici trois illustrations : le morphing de directivité qui, au cours du jeu, transforme la directivité d'un instrument de musique ou d'un évènement sonore (figure 48, gauche) ; le contrepoint spatial qui assigne des directivités à des flux rythmiques ou mélodiques d'une partition (figure 48, droite) ; le portrait cubiste qui, à partir d'éléments échantillonnés et grâce à des traitements de directivité, donne à percevoir un instrument de musique sous un angle de vue particulier (figure 49).

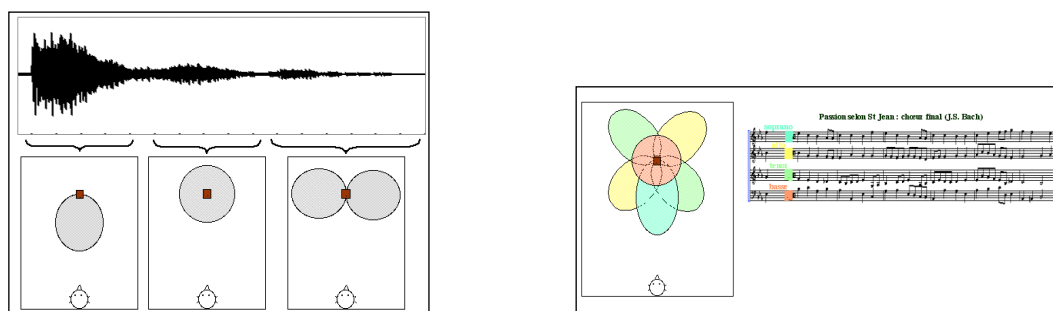


Figure 48. Prototype multi-voies : morphing de directivité (gauche), contrepoint spatial (droite)

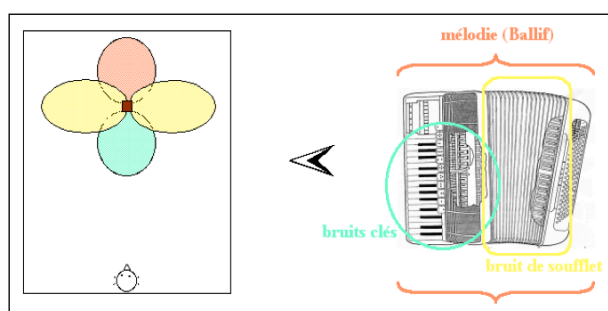


Figure 49. Prototype multi-voies : portrait cubiste

II.3.3.2.3.2 Premières performances

Cette collaboration s'est finalement concrétisée par une création de François Nicolas : *Duelle*, oeuvre pour piano, violon, voix et électronique dans laquelle la diffusion de la partie électronique est assurée un dispositif multi haut-parleurs à directivité contrôlable, nommé pour l'occasion Timée (en rapport au Timée de Platon). Le matériau sonore utilisé pour cette pièce est constitué de séquences échantillonnées correspondant à différents instruments : ceux présents sur scène, mais aussi des voix parlées, la flûte, le clavecin et le grand orgue. S'il peut paraître paradoxal de postuler une approche alternative de la musique mixte par une oeuvre n'utilisant que des instruments (*live* ou

enregistrés), c'est qu'en fait le caractère 'mixte' dans *Duelle* n'est pas – d'après les propos mêmes du compositeur – un mixte de sonorités (instrumentales / synthétiques), ni à un mixte d'instruments (acoustiques / électroniques) mais à un mixte quant aux manières de concevoir l'instrumentalité musicale : un mixte d'instrumentalisation.

Suite à cette première expérience, une seconde œuvre impliquant la Timée est en cours d'écriture : "Five Imaginary Spaces" (Oliver Schneller), oeuvre pour piano et électronique. Du fait de l'instrumentarium choisi et de la nature du matériau sonore développé dans la partie électronique, la Timée est ici employé dans l'optique d'un prolongement / extension de l'instrument plutôt que dans celle précédente qui l'assimile davantage à un "générateur" d'images instrumentales.

II.3.3.2.4 Une nouvelle simulation de l'instrument de musique

II.3.3.2.4.1 Contrôle indépendant du timbre et de la directivité

Dans le cas précis de la simulation d'un instrument de musique, il apparaît également intéressant de pouvoir contrôler indépendamment les contributions spatiales et spectrales de la diffusion. Pour cela, outre le fait de doter le dispositif d'une pondération des directivités élémentaires en fonction de la fréquence, nous nous sommes intéressés plus particulièrement à la reproduction du spectre de puissance rayonnée par un instrument de musique. C'est en effet cette grandeur moyenne, prenant en compte globalement toutes les directions de l'espace, qui détermine, pour la plus grande part de l'auditoire, la perception du timbre des sources placées sur scène.

Pour ce faire, nous utilisons une technique traditionnelle d'analyse de la puissance acoustique (en milieu réverbérant) adaptée à une salle de concert quelconque [Jot et al., 1997]. Une gamme chromatique jouée par l'instrument est utilisée comme signal de référence et est enregistrée en champ lointain par un microphone bi-directionnel placé de telle manière à réduire au maximum la captation du champ direct. L'estimation du temps de réverbération et la prise en compte d'une fonction d'égalisation du microphone permettent de calculer une estimation de la puissance rayonnée par l'instrument. A titre d'exemple significatif, la figure 50 (gauche) montre les spectres de puissance pour différents jeux d'un clavecin.

Ceci étant, une prise de son en champ lointain ne pourrait en aucun cas constituer le signal d'entrée de notre dispositif puisqu'elle contient également toute l'information de la salle. Nous devons donc calculer une correction entre champ proche et champ lointain afin d'utiliser une prise de son rapprochée. Ce filtre correcteur est directement issu de la fonction de transfert entre le spectre de puissance estimé précédemment et le spectre émis par l'instrument en situation 'anéchoïque' ; la figure 50 (droite) donne un exemple de cette correction dans le cas de la flûte. Le signal capté en champ proche peut alors venir 'alimenter' le dispositif de reproduction à directivité contrôlable qui, puisqu'il a été préalablement égalisé en puissance (voir précédemment), restituera au mieux les propriétés spectrales "moyennes" de l'instrument étudié tout en gardant toute sa latitude quant aux caractéristiques spatiales (rayonnement) de l'ensemble.

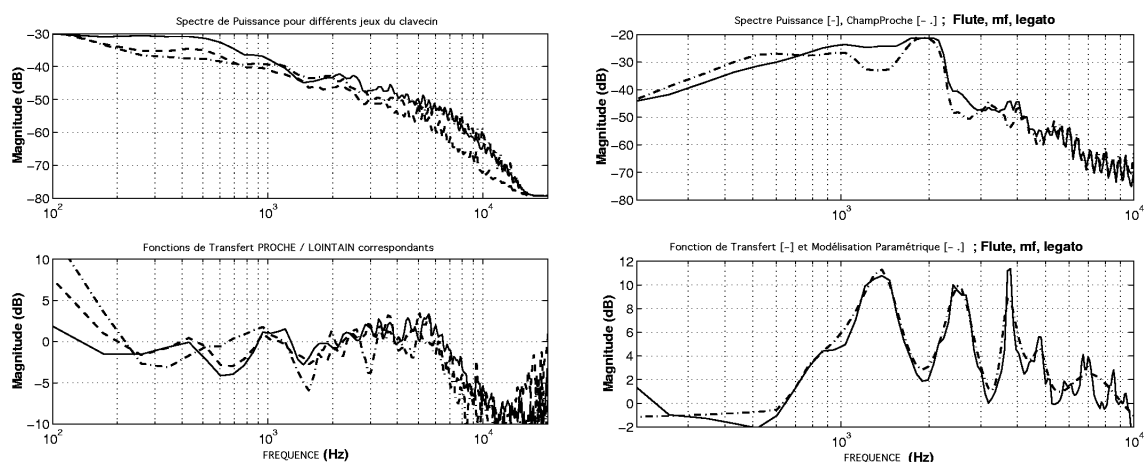


Figure 50. Mesures et modèles de spectres de puissance d'instruments : clavecin (gauche), flûte (droite)
 Gauche : spectre de puissance de différents jeux de clavecin (haut) ;
 fonction de transfert champ lointain / champ proche (bas).
 Droite : spectre de puissance [—] et spectre en champ proche [- -] de la flûte ;
 fonction de transfert [—] et sa modélisation paramétrique [- -].

II.3.3.2.4.2 Vers une synthèse sonore “augmentée”

Ce complément à la méthode de contrôle du rayonnement développée précédemment a démontré ses capacités sur des instruments comme le clavecin ou la flûte. Dans la cadre d'une utilisation généralisée à la reproduction sonore, il nous paraît intéressant d'étudier son efficacité sur des instruments plus 'complexes' comme le piano qui pose, encore aujourd'hui, des problèmes de rendu sonore réaliste.

Par ailleurs, en projetant une perspective d'application plus large, il semble également pertinent d'associer cette approche à une application dédiée à la synthèse sonore. En effet, selon la technique utilisée, un moteur de synthèse sonore va soit reconstituer les caractéristiques spectro-temporelles du signal (synthèse de type signal), soit reproduire des échantillons pré-enregistrés (synthèse par échantillonnage), soit calculer les variations de grandeurs physiques liées à l'instrument (synthèse par modélisation physique). La plupart du temps, les modes de rayonnement ne sont pas pris en compte et sont finalement délégués aux systèmes de diffusion traditionnels. Le dispositif de contrôle du rayonnement tel qu'il est développé dans la présente étude pourrait ainsi constituer un prolongement 'naturel' et fonctionnel à de nombreux synthétiseurs qui se verraient, de fait, “enrichis” d'une dimension spatiale supplémentaire. En ce qui concerne le piano, des études ont été ébauchées dans deux directions différentes. D'une part, coupler un piano numérique (échantillonneur) au système de reproduction en développant des interfaces de contrôle en temps-réel spécifiques afin, dans un premier temps de restituer l'image spatiale de l'instrument, puis, dans un second de temps, de proposer des extensions pouvant constituer de nouveaux modes de jeu de l'instrument ; un peu à l'image de l'organiste qui, en modifiant ses jeux, change également la position et la nature des sources sonores actives (tuyaux) et, du même coup, le rayonnement global de l'orgue. Et d'autre part, travailler dans l'environnement de synthèse sonore par modélisation physique Modalys qui semble propice à une approche de ce genre. En effet, conceptuellement, le logiciel est d'ores et déjà capable de modéliser un piano, c'est-à-dire, une marteau qui frappe une corde reliée à une table d'harmonie via un chevalet. Cependant, ce modèle implique pour l'instant de nombreuses simplifications, tant au niveau du marteau que

du type d'interaction – frappée – ou bien encore de la prise en compte du doublet ou du triplet de cordes pour certains registres de l'instrument. Ceci étant, et quelque soit la précision du modèle, la 'sortie' de Modalys délivre un signal relatif à la vibration des résonateurs (vitesse) – comme si l'instrument était écouté via un microphone de contact posé sur la table d'harmonie – sans que la propagation de cette grandeur jusqu'à un point d'écoute donné ne soit modélisée. On imagine alors aisément le rôle que pourrait jouer le dispositif de contrôle du rayonnement dans la modélisation de ce couplage fluide-structure. Selon l'avancement des travaux, les résultats obtenus dans ces deux axes de recherche seront détaillés et, dans la mesure du possible, "joués" lors de la présentation

II.3.3.2.5 Conclusion

Une précédente étude a posé les bases théoriques et expérimentales d'une procédure visant à reproduire une figure de directivité donnée au moyen d'un dispositif multi haut-parleurs. A partir de ces résultats, une étude dérivée a été explorée dans le but de proposer une approche plus musicale du contrôle du rayonnement. Pour ce faire, une nouvelle méthode a été développée et validée sur différents prototypes : elle repose sur un réseau de transducteurs tri-dimensionnel et sur la constitution d'un ensemble de directivités élémentaires issues de la base des harmoniques sphériques. Cette architecture "originale" donne la maîtrise de deux paramètres d'espace : la nature et la direction d'émission de l'onde sonore. En outre, grâce à une implémentation optimisée, le dispositif autorise également un contrôle en temps-réel ainsi que la gestion simultanée de plusieurs signaux d'entrée. Il offre enfin la possibilité de restituer le spectre de puissance rayonné indépendamment des caractéristiques spatiales, ce qui constitue un atout supplémentaire pour la simulation d'un instrument de musique.

D'un point de vue musical, le système tente de répondre à une nouvelle approche de la musique mixte – mêlant sons acoustiques et synthétiques – en concevant un système de diffusion sonore sur le modèle de l'instrument de musique : localisé et possédant une fonction de rayonnement propre. Cette démarche tente finalement de replacer sur un même plan les instruments *live* et la partie électronique diffusée qui se livrent le plus souvent un 'combat' inégal.

Dans le cadre plus restreint de la reproduction d'un instrument de musique, le dispositif de contrôle du rayonnement, grâce à ses capacités techniques relatives au spectre de puissance, permet d'entrevoir des associations intéressantes avec différentes techniques de synthèse sonore : des plus traditionnelles (échantillonnage), aux plus avancées (modélisation physique). Dans tous les cas, le système apparaît comme un prolongement possible des synthétiseurs qui, à l'heure actuelle, ne prennent en compte que très rarement la dimension spatiale du son.

II.4 Etude du rayonnement appliqué à la reproduction d'un instrument

Dans la continuité des travaux effectués sur le contrôle de la directivité des sources, cette étude sur la mesure et la modélisation de la directivité des instruments de musique s'inscrit à la fois dans un objectif scientifique, mais également dans le cadre d'une collaboration artistique, en réponse à une demande formulée par le compositeur

Marco Stroppa¹⁵. En effet, à la faveur d'une commande du festival de Donaueschingen, (2010), M. Stroppa a eu pour projet d'écrire un concerto pour cor de basset et orchestre. L'instrument soliste étant amplifié car loin du public, le compositeur a souhaité exploiter cet élément électroacoustique de son intrumentarium et utiliser les possibilités offertes par le contrôle de la directivité pour reproduire au mieux l'instrument et le faire co-exister avec l'orchestre.

II.4.1 Description de l'étude

L'étude s'est inscrite dans le cadre d'un stage de Master (Eric Boyer, ENSIM) et a fédéré une collaboration interne entre les différentes équipes de l'Ircam concernées par la thématique : acoustique instrumentale (René Caussé) et acoustique des salles (Markus Noisternig) [Misdariis et al., 2010a].

L'objectif du projet est donc d'évaluer la faisabilité de la reproduction du rayonnement d'un instrument de musique à vent à trous latéraux à l'aide d'une source multi-haut-parleurs à directivité contrôlable. Les différents éléments de l'étude sont les suivants :

- la **source** : Global Source Radiator (GSR) – prêtée par le Laboratoire d'Acoustique Musicale (Hugues Genevois, LAM/Université Pierre et Marie Curie). La GSR est une source omnidirectionnelle munie de 12 haut-parleurs répartis uniformément à la surface d'une sphère (voir figure 51 et § II.3.2.1). Elle est initialement conçue pour la mesure de salles (normes ISO 3382, ISO 354, ISO 140-3 et DIN 5221 relatives à l'isotropie) ; cela étant, ses 12 canaux peuvent aussi être alimentés, et donc contrôlés, indépendamment. Pour le besoin de l'étude, les 12 canaux sont donc équipés chacun d'une voie de conversion et d'amplification identique.

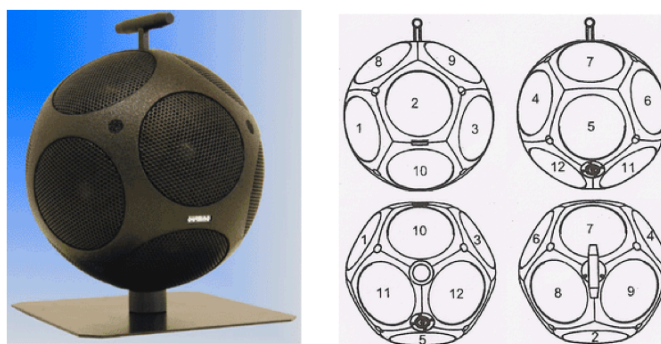


Figure 51. Global Source Radiator (GSR) et nomenclature des haut-parleurs

- la **cible** : le projet du compositeur M. Stroppa s'intéresse au cor de basset, instrument de la famille des clarinettes (anche simple, réseau de trous latéraux, pavillon, perce cylindrique). A des fins de comparaison – mais aussi pour des raisons pratiques de disponibilité des instruments –, l'étude a débuté, dans un premier temps, sur un instrument relativement proche : le saxophone ténor (anche simple, réseau de trous latéraux, pavillon, mais perce conique) (voir figure 52). Le champ acoustique rayonné par ce type d'instrument se révèle être complexe, dépendant de la note jouée et du doigté utilisé par le musicien (plusieurs doigtés pour une même note) [Bresciani, 1996].

¹⁵ <http://brahms.ircam.fr/marco-stroppa>



Figure 52. Cor de Basset (gauche) – Saxophone tenor (droite)

- la **caractérisation** : les champs acoustiques rayonnés par la source et les cibles sont mesurés en chambre anéchoïque (condition de champ libre). Les dispositifs (source ou cibles) sont placés au centre d'une sphère de mesure constituée d'une table tournante (azimuth) et d'un bras élévateur équipé d'un arc de microphones (élévation) ; l'ensemble étant automatisé tant du point de vue de la manipulation que de l'acquisition (voir figure 53). La source GSR est excitée successivement par chacune de ses voies ; les cibles sont excitées par un moteur de haut-parleur placé au niveau du bec (voir § II.4.2.3).
- la **modélisation** : ces données de mesure sont ensuite analysées, et en l'occurrence décomposées sur la base des harmoniques sphériques. L'analyse au moyen de ce formalisme permet : i/ pour la source, de tenir compte des nombreux couplages internes entre haut-parleurs [Behler et al., 2006 ; Leishman et al., 2006] et de modéliser les filtres *ad hoc* pour lui permettre de reproduire au mieux les figures de directivité élémentaires ; ii/ pour les cibles, de quantifier la projection de leur rayonnement sur la base des harmoniques sphériques et calculer ainsi le coefficient de pondération pour chacun des filtres de base.

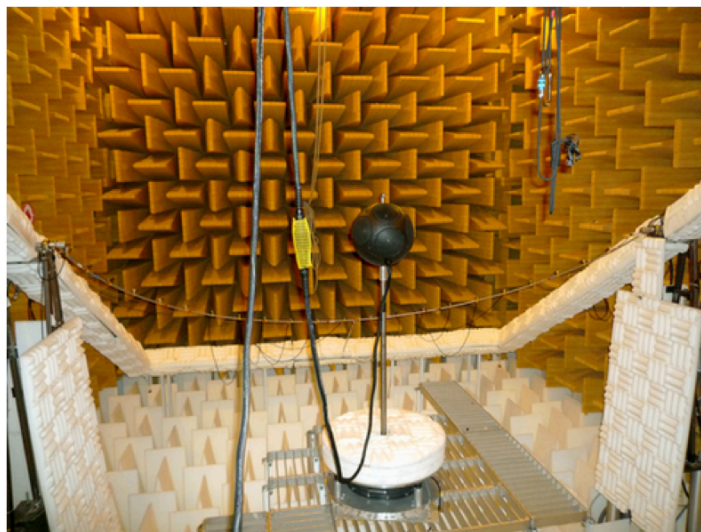


Figure 53. Dispositif de mesure chambre anéchoïque

- **l'évaluation** : la reproduction de la directivité des cibles (saxophone, cor de basset) par la source (GSR) est ensuite évaluée par une nouvelle série de mesures de la source filtrée. Les signaux d'alimentation de la source sont précalculés ; cette première approche "temps différé" du problème permettant de réaliser un filtrage précis et optimisé, à la différence d'une implémentation temps-réel (non abordée dans cette étude) qui demande des approximations en termes de design de filtres. La caractérisation objective de la source "contrôlée" – c'est-à-dire, filtrée – permet de valider la reproduction du champ acoustique des instruments cibles ; elle ouvre également la voie à des évaluations subjectives par des tests perceptifs en situation réelle (également non abordée dans cette étude).

II.4.2 Caractérisation des données

II.4.2.1 Dispositif expérimental

Les mesures effectuées consistent en un relevé de la pression acoustique sur un ensemble de points situés sur une sphère de mesure centrée sur le système étudié. Du fait de la nature du signal d'excitation (voir plus bas), ce relevé permet ensuite de calculer la réponse impulsionnelle du système puis d'en déduire sa réponse fréquentielle en fonction des directions de mesure (figures de directivité).

Les mesures sont effectuées dans la chambre anéchoïque de l'Ircam dans des conditions de *quasi*- champ libre (la fréquence de coupure de la chambre se situe aux alentours de 70 Hz). Cet espace est équipé d'une table tournante Brüel&Kjaer commandée en rotation autour de l'axe vertical. De plus, un bras robotisé muni d'un réseau de 24 microphones en demi-cercle est commandé en élévation autour de l'axe transversal. L'ensemble permet de "scanner" une source sur une demi-sphère (supérieure) de mesure, et par retournement, sur la sphère toute entière. La source est centrée sur le dispositif à l'aide de 3 lasers coïncidents.

Les microphones (capsules à électret mce2000 Monacor) sont répartis tous les 7.8° et situés, à 146.5 ± 1 cm du centre de la sphère de mesure. Ils sont préalablement calibrés en fréquence afin de s'assurer de la neutralité de leur réponse. Un échantillonnage tous les 10° d'élévation est choisi, ce qui donne 828 points de mesure et 36 références (un microphone, fixe par rapport à la source permet une référence tous les 10°). Ce choix résulte d'un compromis temps/résolution spatiale, ayant aussi des conséquences sur le temps de traitement des données.

Un programme conçu sous Matlab et communiquant avec Max/MSP *via* le protocole UDP¹⁶ (User Datagram Protocol) permet de contrôler les différents équipements de la chambre, les signaux d'excitation et l'enregistrement des mesures (voir figure 54).

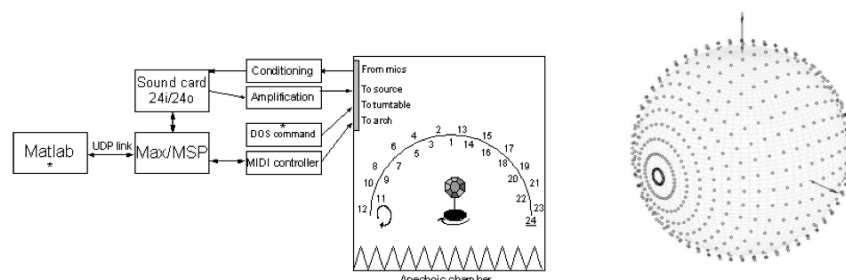


Figure 54. Dispositif de mesure du rayonnement en chambre sourde
Schéma du système de mesure (gauche) et de l'échantillonnage spatial (droite)

¹⁶ http://fr.wikipedia.org/wiki/User_Datagram_Protocol

En outre, une attention particulière est portée à l'homogénéité de la distribution des points de mesure sur la sphère, condition nécessaire à la bonne décomposition du champ mesuré sur la base des harmoniques sphériques (la matrice des coefficients issus de la décomposition pouvant, dans certains cas, présenter un mauvais conditionnement). Avec des résolutions spatiales élevées, il est donc nécessaire d'employer une technique de pondération des microphones. Les 828 points du système se densifiant significativement au voisinage de l'axe de rotation transversal, une correction est appliquée à chaque microphone selon sa position sur le maillage – en fait, selon sa position sur l'arc de mesure (voir figure 55).

Une technique couramment employée est la pondération de Voronoï dont l'algorithme détermine les surfaces contenant l'ensemble des points les plus proches d'un point de référence. Dans le cas présent, les surfaces sont calculées le long d'une sphère et les pondérations sont définies en termes d'angle solide intervenant dans le calcul des coefficients de la décomposition en harmoniques sphériques (voir Eq. 20).

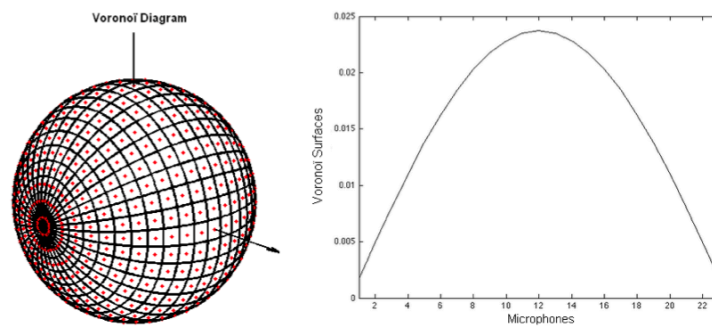


Figure 55. Surface de Voronoï sur la sphère et pondération selon une ligne de microphones

Le signal d'excitation utilisé est de type balayage fréquentiel logarithmique. Ce type d'excitation est bien adapté à un haut-parleur et recommandé pour une mesure en chambre anéchoïque où elle permet la quasi-absence de distorsion harmonique et un bon rapport signal sur bruit [Stan et al., 2002]. On utilise ici plus particulièrement la méthode décrite par Farina [Farina, 2000] : elle consiste à exciter le système avec un sinus glissant dont la fréquence augmente de manière logarithmique dans le temps. Le signal temporel de pression relevé par les microphones est ensuite déconvolué par le sinus glissant afin d'obtenir la réponse impulsionnelle du système étudié.

Par ailleurs, d'un point de vue fréquentiel, la bande passante est coupée dans les basses fréquences à 90 Hz, notamment pour protéger les haut-parleurs de la GSR. D'autre part, les instruments cibles possèdent des tessitures qui débutent au Fa2 (87 Hz) pour le cor de basset et au Fa#2 (92 Hz) pour le saxophone ténor.

II.4.2.2 Mesure de la source

La mesure de la source GSR a d'abord nécessité la mise en place d'un important cadre technique afin de disposer de 12 voies d'amplification audio (cartes-son, convertisseurs, amplificateurs) rigoureusement identiques. Grâce à une procédure automatisée, 14 configurations sont successivement mesurées pour chaque pas de mesure (1 position de la table tournante en azimuth et une position du bras en élévation) :

- 'HPS' : les 12 haut-parleurs seuls ;
- 'OMNI' : les 12 haut-parleurs ensemble en phase pour simuler le monopole ;
- 'DIPO' : 2 haut-parleurs diamétralement opposés sur la source (n° 2 et 5), alimentés en opposition de phase pour simuler une configuration de type dipôle.

Les figures de directivité obtenues pour les configurations 'HPs' permettent de retrouver le comportement classique du rayonnement d'un haut-parleur électrodynamique : isotropie en basses fréquences (ici, env. 500 Hz), puis augmentation progressive de la directivité en fonction de la fréquence (voir figure 56).

On remarque néanmoins quelques particularités certainement dues au couplage entre haut-parleurs (voir figure 57). On peut en effet imaginer qu'à certaines fréquences, un couplage fluide de type "push-pull" s'opère entre deux haut-parleurs opposés. Plus largement, les fortes symétries de la source, la présence de systèmes résonants en contact mais aussi le volume d'air commun à tous les haut-parleurs peuvent potentiellement expliquer ces phénomènes. Dans leur étude de 2008, Zotter et coll. montrent d'ailleurs qu'un volume d'air individuel et non partagé minimiserait les couplages entre les haut-parleurs d'une source sphérique [Zotter et al., 2008a].

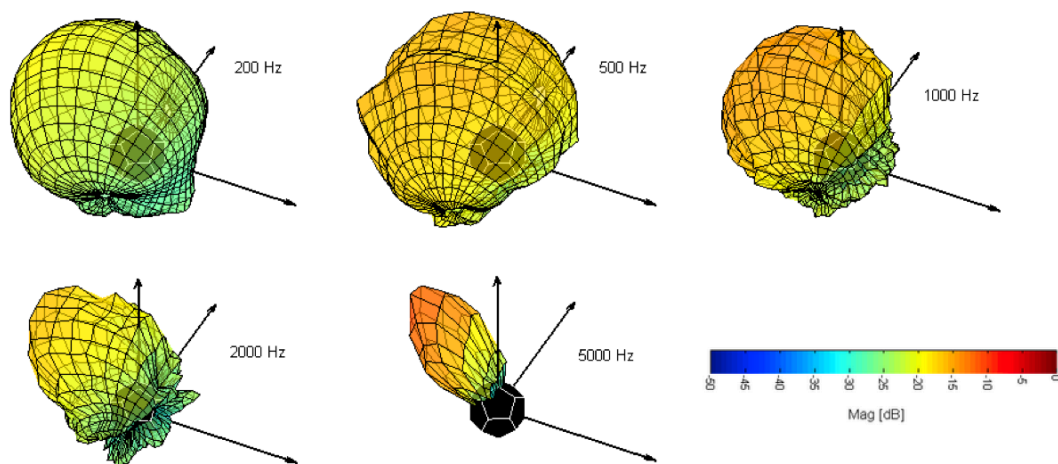


Figure 56. Champ acoustique mesuré pour le HP n° 7 du GSR

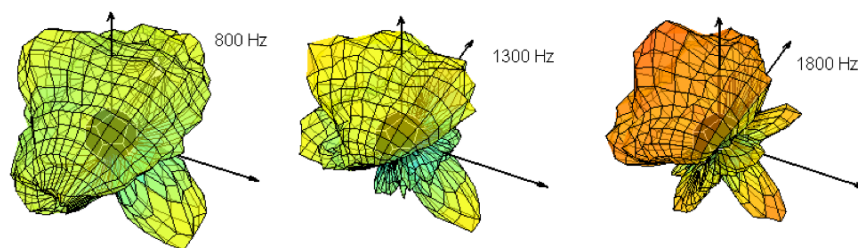


Figure 57. Singularités de rayonnement du HP n° 7 du GSR

Les deux autres configurations de mesure, 'OMNI' et 'DIPO', ont respectivement permis de connaître les limites de fonctionnement de la source en termes d'isotropie – relativement en accord avec la limite de 5 kHz donnée par le constructeur – (voir figures 3-1, 3-2 – Annexe 3), et de vérifier les notions de phases relatives dans les champs de pression acoustique, élément important dans le processus de reproduction du rayonnement (voir figures 3-3 – Annexe 3). La configuration 'DIPO' fait, par ailleurs, apparaître une figure intéressante présentant une réjection de presque 35 dB perpendiculairement à l'axe du dipôle.

A noter que l'architecture de la source elle-même détermine une fréquence de coupure au-delà de laquelle apparaissent des phénomènes de repliement spatial (*spatial aliasing*). En effet, la source étant assimilable à un réseau de transducteurs, en

l'occurrence tri-dimensionnel, avec un pas d'échantillonnage donné, l'équivalent du théorème de Shannon dans le domaine spatial délimite une longueur d'onde λ , en fonction de la distance inter haut-parleurs d , au-delà de laquelle le phénomène d'*aliasing* apparaît en théorie : $\lambda_c = 2d$. Dans le cas présent, la distance d valant 13.5 cm., la fréquence de coupure se situe aux alentours de 1.26 kHz.

II.4.2.3 Mesure des cibles

La mesure sur les instruments a d'abord nécessité le développement d'un système d'excitation *ad hoc* : un moteur de haut-parleur, placé à l'embouchure de l'instrument dans une enceinte close afin d'éviter toute perturbation des mesures due à son rayonnement propre (voir figure 58). Cela étant, des problèmes propres à la nature des sources instrumentales sont venus s'ajouter à la difficulté de mise en place du protocole expérimental :

- La notion de centre acoustique. On trouve dans la littérature plusieurs définitions du concept de centre acoustique d'une source principalement considéré en champ lointain [Jacobsen et al., 2004]. En champ proche (notre cas de figure expérimental), le problème est plus complexe car il consiste à remplacer une source étendue par un point unique, qui plus est, variant en fonction de la fréquence et du doigté. De manière pragmatique, un point arbitraire est choisi ici comme centre de la source à partir duquel l'instrument est positionné dans le système de mesure et autour duquel il pivote lors des retournements successifs. Pour le saxophone, ce point est situé entre le pavillon et le conduit, à mi-chemin du pavillon. Pour le cor de basset, ce point est situé au milieu de l'axe du tube.
- La question des doigtés et des modes de jeu. Idéalement, la mesure de rayonnement de l'instrument devrait s'effectuer pour chaque note, donc pour chaque doigté, sachant en outre que – notamment pour le saxophone – une note peut aussi être obtenue avec plusieurs doigtés différents. A nouveau pour des raisons pratiques, quatre doigtés ont été choisis arbitrairement sur la base d'une représentation variée des situations de l'instrument :
 - doigté "ouvert" (instrument au repos, on ne touche à aucune clé – le Si médium pour le saxophone, 246 Hz) ;
 - doigté "fermé" (on joue la note qui ferme le maximum de trous – le La bémol grave pour le saxophone, 104 Hz) ;
 - doigté "semi-fermé" (la main gauche actionne ses clés, les trous de la moitié supérieure du conduit sont alors fermés – le Fa médium pour le saxophone, 175 Hz) ;
 - doigté "aigu" (on ouvre quelques trous très proche de l'entrée du tube – le Mi bémol aigu pour le saxophone, 622 Hz)

Les doigtés sont obtenus en bloquant les clés et les mécanismes essentiellement au moyen d'élastiques.



Figure 58. Dispositif pour la mesure du saxophone (gauche) et du cor de basset (droite)

Pour le saxophone, dans la configuration "ouverte", les figures de directivité obtenues sur les instruments sont cohérentes avec des comportements classiques : globalement monopolaire en basse fréquence (ici jusqu'à env. 400 Hz) puis variant notablement en fonction du doigté – et donc de la fréquence – de la note jouée avec une majorité de l'énergie rayonnée par le pavillon au fur et à mesure que la fréquence augmente. En outre, deux figures caractéristiques des instruments à vent sont observables : vers 600 Hz, l'apparition d'une couronne dans le plan perpendiculaire à l'axe principal de l'instrument ; vers 1000 Hz, un dédoublement de cette couronne, phénomène identique à celui modélisé sur le hautbois par Rousseau [Rousseau, 1996] (voir figure 59).

Les autres configurations ("fermée", "semi-fermée" et "aigue") sont globalement régies par les mêmes principes, avec néanmoins des rayonnements plus variables pour les doigtés "semi-fermé" et "aigu" – avec par exemple, apparition de dipôles (voir figures 4-1 à 4-3 – Annexe 4).

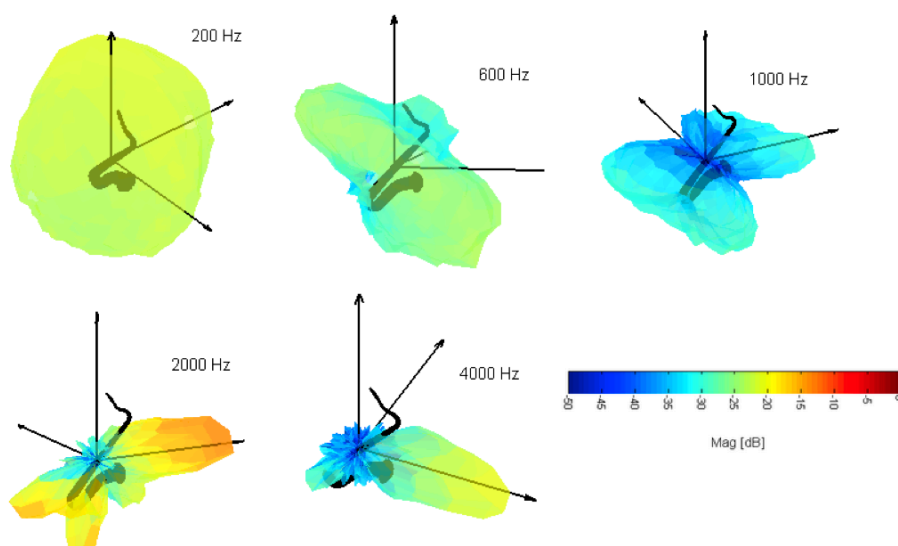


Figure 59. Directivité du saxophone mesurée pour un doigté "ouvert" (Si medium, 246 Hz)

Pour le cor de basset, on retrouve globalement les mêmes comportements de rayonnement qu'avec le saxophone, avec néanmoins quelques particularités. Le caractère omnidirectionnel du champ n'est réellement présent que pour le doigté "fermé". Pour les autres configurations, le champ présente un maximum dans l'axe du pavillon mais aussi vers le bas, dans l'axe du corps de l'instrument. Il est intéressant d'observer également qu'à partir du moment où quelques trous sont ouverts, le champ se comporte de façon assez similaire quelque soit le doigté. De plus, à partir de 1.5 kHz la directivité présente de nombreuses zones d'atténuation et varie beaucoup avec la fréquence, ce qui fait que la directivité du cor de basset semble globalement plus complexe que celle du saxophone (voir figure 5-1 à 5-4 – Annexe 5).

II.4.3 Modélisation des données

Que ce soit pour la cible (GSR) ou les sources (saxophone, cor de basset), les données ont été mesurées avec le même dispositif et ont donc produit un ensemble de fonctions $f(\theta, \phi)$ donnant la pression acoustique en fonction de la position spatiale (θ, ϕ) sur la sphère de mesure.

Par ailleurs, et par définition, la base des harmoniques sphériques $\{Y_n^m(\theta, \phi)\}$ constituent un ensemble de fonctions mathématiques sur laquelle toute fonction $f(\theta, \phi)$ (de carré sommable sur une sphère) peut se décomposer, sous la forme suivante :

$$f(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} A_{nm} Y_n^m(\theta, \phi) \quad (\text{Eq. 19})$$

où les $\{A_{nm}\}$ sont les coefficients complexes issus de la décomposition caractérisant le champ f au niveau de la sphère d'analyse. En utilisant les propriétés de symétrie hermitienne et d'orthonormalité (voir § II.2.2), ces coefficients peuvent être exprimés de la manière suivante :

$$A_{nm} = \int Y_n^m(\theta, \phi)^* f(\theta, \phi) \cdot d\Omega \quad (\text{Eq. 20})$$

le terme $d\Omega$ étant le terme d'intégration correspondant aux angles solides calculés par l'algorithme de Voronoï (voir § II.4.2.1).

Pour la cible (GSR), la décomposition des fonctions décrivant le champ des $L = 12$ haut-parleurs au degré N aboutit ainsi à une série de $(N+1)^2$ coefficients complexes en fonction de la fréquence pouvant s'écrire sous la forme matricielle suivante :

$$\mathbf{C}(\omega) = \begin{bmatrix} {}^1A_0^0(\omega) & {}^2A_0^0(\omega) & \dots & {}^LA_0^0(\omega) \\ {}^1A_1^{-1}(\omega) & {}^2A_1^{-1}(\omega) & \dots & {}^LA_1^{-1}(\omega) \\ \dots & \dots & \dots & \dots \\ {}^1A_N^{+N}(\omega) & {}^2A_N^{+N}(\omega) & \dots & {}^LA_N^{+N}(\omega) \end{bmatrix} \quad (\text{Eq. 21})$$

Par ailleurs, le degré de la décomposition en harmoniques sphériques étant limité par le nombre de degrés de liberté de la source ($L=12$), la matrice \mathbf{C} est développée uniquement jusqu'au deuxième degré, définissant ainsi une matrice \mathbf{C}_r qui décrit les 9 premières fonctions de base déployées sur 3 séries complètes d'harmoniques sphériques. Cette matrice peut être vue comme un système MIMO (Multiple Input Multiple Output) à 12 entrées et 9 sorties reliant le comportement propre de chaque haut-parleur de la source aux figures de directivité de la base tronquée.

Les filtres permettant la reproduction des 9 premières harmoniques sphériques de la base, sont ensuite obtenus en inversant cette matrice \mathbf{C}_r ; la matrice \mathbf{C}_r^{-1} ainsi obtenue présentant 9 entrées – la commande des figures de directivité – et 12 sorties – les canaux de la source GSR. Chaque élément complexe (i,j) de cette matrice correspond donc littéralement au filtre (amplitude et phase) à appliquer au $i^{\text{ème}}$ haut-parleur de la GSR pour obtenir la $j^{\text{ème}}$ harmonique sphérique de la base (voir figure 60).

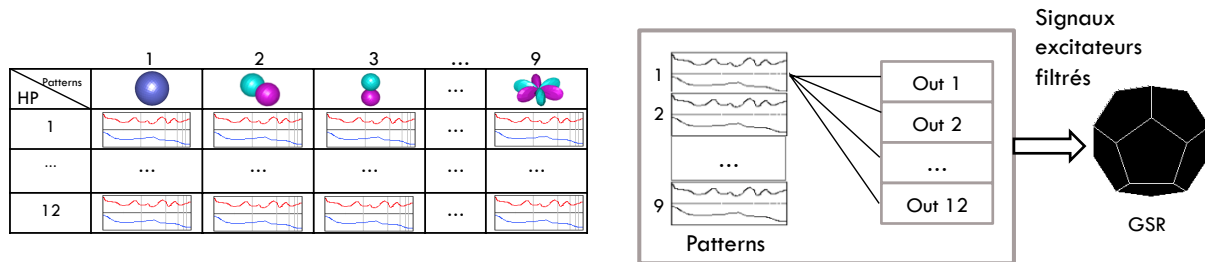


Figure 60. Inversion du système MIMO pour reproduction des harmoniques sphériques par la source GSR

D'un point de vue numérique, la matrice C_r n'étant pas carrée, l'opération d'inversion n'est pas triviale et correspond plutôt à une pseudo-inversion. Pour cela, la fonction `pinv`¹⁷ de Matlab utilisant la méthode de Moore-Penrose est choisie. De plus, la bande passante des coefficients est réduite (170 – 14600 Hz) afin d'améliorer la stabilité de cette opération d'inversion ; en effet, notamment en basses fréquences, la faible énergie de la source est compensée par des forts coefficients (processus d'égalisation) qui altèrent le conditionnement de la matrice.

II.4.4 Résultats

Dans un premier temps, la qualité de reproduction des harmoniques sphériques par la source GSR est évaluée. Pour cela, le dispositif expérimental est repris à l'identique (voir § II.4.2.2) ; le signal d'excitation (sinus glissant) est quant à lui filtré par la matrice C_r^{-1} précédemment calculée selon l'opération exprimée dans l'Eq. 22, le contrôle de la source s'effectuant en sélectionnant la fonction Y_n^m souhaitée :

$$\mathbf{u}(\omega) = C_r^{-1}(\omega) \cdot \begin{bmatrix} Y_0^0 \\ Y_1^{-1} \\ \dots \\ Y_2^{+2} \end{bmatrix} \quad (\text{Eq. 22})$$

Les 9 configurations de directivité sont ainsi reproduites par la source et mesurées par la chaîne de mesure ; le traitement des données – identique à celui opéré lors de la caractérisation initiale de la source – permettent de visualiser et d'analyser les résultats obtenus.

Globalement, les degrés 0 et 1 semblent correctement reproduits sur une plage de fréquences allant de 150 Hz à 1.5 kHz environ. La reproduction du degré 2 n'est quant à elle correcte qu'à partir d'une fréquence plus élevée (env. 400 Hz). La figure 61 montre les 9 directivités reproduites à une fréquence optimale du système, à savoir 1 kHz ; elle peut être comparée aux fonctions théoriques de la figure 26 (§ II.2.2).

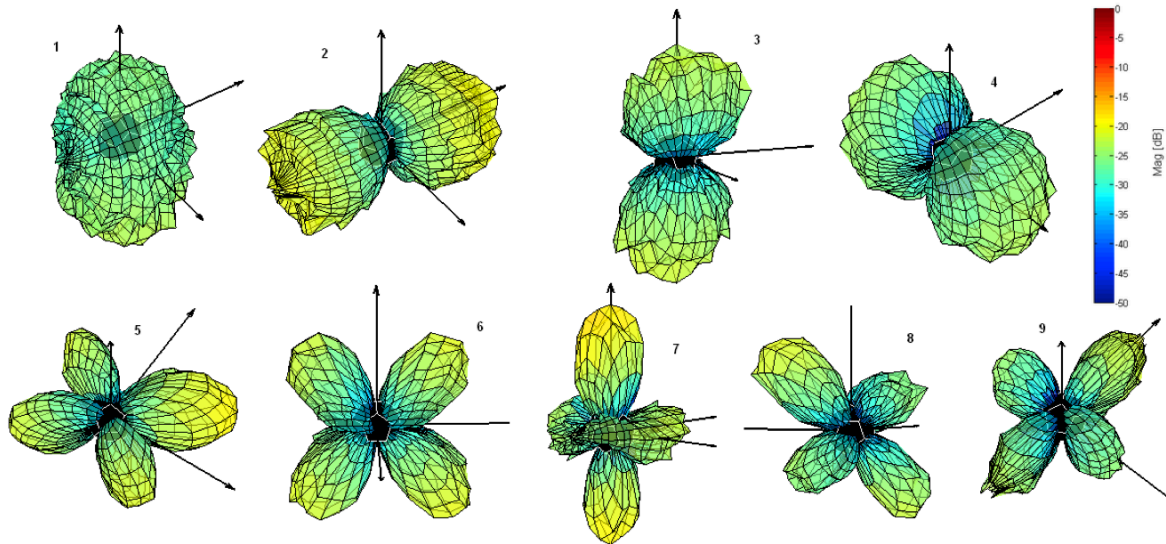


Figure 61. Mesures de directivité de la source GSR reproduisant les harmoniques sphériques à 1 kHz

¹⁷ <http://www.mathworks.com/help/matlab/ref/pinv.html>

On observe également que l'algorithme d'encodage permet une relative égalisation en énergie de l'ensemble des fonctions. Dans la même idée, on peut également comparer les spectres obtenus en configuration brute 'OMNI' et lors de la reproduction de la fonction Y_0^0 (monopole), en un même point de la sphère de mesure – en l'occurrence le point $\theta=0, \phi=0$. On constate que le processus tend bien à égaliser le comportement de la source même si quelques artefacts subsistent néanmoins le long de la bande-passante (voir figure 62).

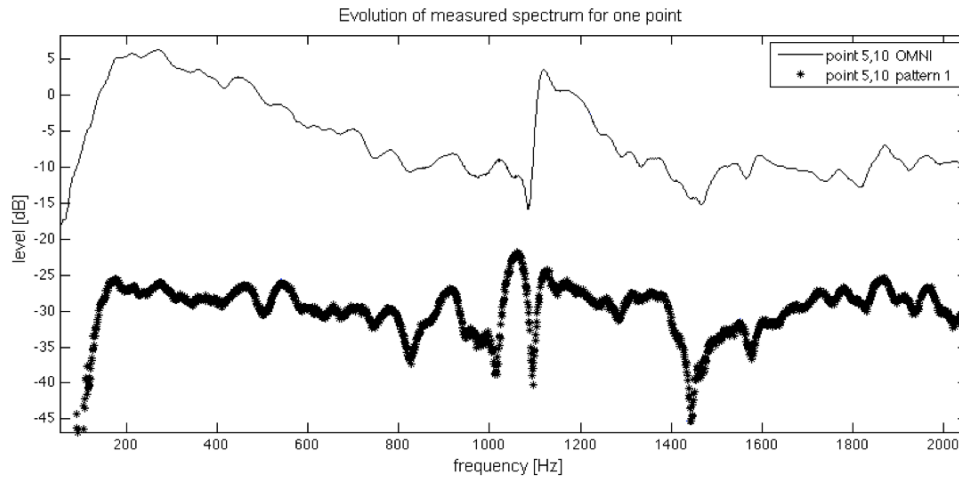


Figure 62. Influence du contrôle de directivité sur le spectre omnidirectionnel de la source GSR
Spectres mesurés au point ($\theta=0, \phi=0$) pour la configurations 'OMNI' (trait normal)
et la reproduction de l'harmonique de degré 0 par la source GSR contrôlée (trait **gras**)

Par ailleurs, du point de vue de la phase, on remarque une bonne capacité du système à respecter les déphasages au sein d'une même fonction ainsi que par rapport aux fonctions entre elles (voir figure 63). Ce paramètre est très important pour le contrôle de la directivité, notamment pour la reproduction de figures complexes qui combinent des harmoniques sphériques élémentaires.

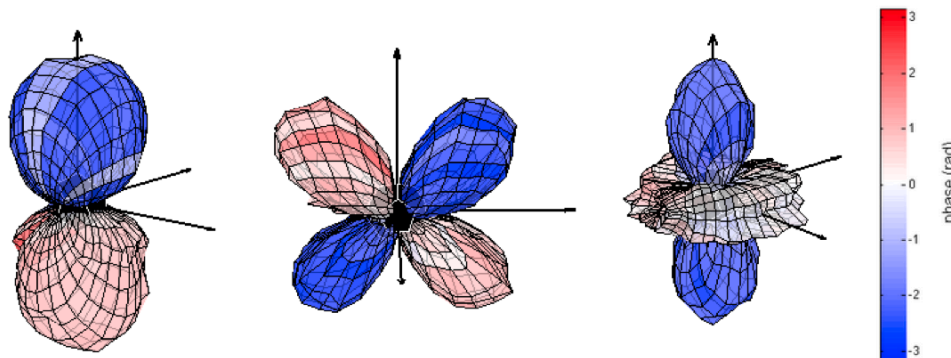


Figure 63. Répartition des phases pour la reproduction des harmoniques sphériques avec la GSR.
Fonctions n°3 (Y_1^0) à 400 Hz, n°5 (Y_2^2) à 650 Hz et n°7 (Y_2^0) à 650 Hz

Car, précisément, le rayonnement des instruments cibles (saxophone, cor de basset) présente des directivités complexes (voir § II.4.2.3) que le formalisme des harmoniques sphériques permet de décomposer en fonctions élémentaires ; ces mêmes fonctions élémentaires étant reproduites par la source GSR moyennant des troncatures et des limitations fréquentielles (cf. ci-dessus). La reproduction d'une directivité complexe s'effectue donc "simplement" en pondérant le jeu de filtres élémentaires de la source par les coefficients issus de la décomposition.

II.4.5 Conclusion

L'objectif visant à reproduire le rayonnement d'un instrument de musique par un système multi haut-parleurs contrôlé en directivité a – presque – été atteint.

En effet, un protocole pour caractériser de manière uniforme les différents éléments de l'étude (source et cibles) a été développé et validé expérimentalement. De nombreuses mesures fiables ont été produites notamment sur les deux instruments à vent considérés : saxophone et cor de basset.

Sur la base des harmoniques sphériques, les champs de pression mesurés spatialement ont été modélisés sous forme de coefficients de contribution complexes et une technique d'inversion du système MIMO correspondant a permis de calculer et d'implémenter les filtres *ad hoc* à chacun des 12 canaux de la source pour qu'elle reproduise au mieux les 9 premières harmoniques sphériques de la base. Les résultats ont été validés sur la source, par analyse comparative avec les données théoriques. A ce stade, une démarche de validation perceptive devrait néanmoins être intéressante à mener pour dégager les grandeurs relatives à la directivité effectivement perçue (voir § II.5.2) et, le cas échéant, pour permettre de confirmer certaines approximations ou, à l'inverse, assouplir certaines contraintes.

Dans le temps imparti pour l'étude, la deuxième phase d'implémentation et de validation concernant la reproduction des directivités complexes des instruments n'a pu être uniquement réalisée que sur des essais informels et préliminaires. Cette étape devrait permettre, d'une part, de concrétiser la capacité de la source à rayonner efficacement sous contraintes (dans la limite des approximations inhérentes au système), et d'autre part, d'aborder la question du temps réel en couplant la détection du jeu du musicien à la connaissance acquise du comportement de l'instrument afin de proposer une amplification fidèle et réaliste du son instrumental.

II.5 Extensions du projet

Le projet portant sur l'étude du rayonnement acoustique des sources et sa reproduction à l'aide de dispositifs électroacoustiques associés à du traitement du signal a suscité des extensions de recherche dans deux axes principaux : d'une part, le couplage de cette approche avec la synthèse sonore par modélisation physique, et notamment la synthèse modale, dans une démarche appelée **Synthèse Additive Spatiale**, et d'autre part, l'**évaluation perceptive** du contrôle de la directivité, et en particulier, l'évaluation de cette grandeur acoustique par rapport à la perception de la source et de son interaction avec l'espace de diffusion.

II.5.1 La Synthèse Additive Spatiale

La démarche de Synthèse Additive Spatiale (SAS) part du principe que les qualités de simulation d'un instrument de musique dépendent à la fois du moteur de synthèse mais aussi du moteur de rendu sonore. L'idée est donc d'envisager le couplage d'un moteur de synthèse sonore par modélisation physique capable de modéliser les différents modes vibratoires d'une source acoustique (typiquement, le logiciel *Modalys*), avec un moteur de rendu sonore (typiquement, *La Timée*, ou équivalent) capable d'assurer le rayonnement acoustique de ces mêmes modes suivant des fonctions spatiales pertinentes d'un point de vue physique, perceptif ou musical. Le "mélange" des différentes contributions spectrales (les modes) se faisant dès lors, non plus dans le domaine du signal – comme c'est le cas pour une synthèse additive classique – mais directement dans l'espace de propagation. La SAS repose donc sur un principe de

synthèse sonore "augmentée" de propriétés spatiales permettant ainsi de produire un signal autant sur le plan spectral que spatial et proposant de nouveaux développements pour la synthèse sonore, notamment instrumentale.

En la matière, l'expérience originale menée par McAdams et Reynolds en 1983 (cité dans [Bregman, 1994]¹⁸) peut permettre d'illustrer en partie le procédé. Elle consiste à créer une illusion perceptive en contrôlant les paramètres de diffusion d'un dispositif stéréophonique, c'est-à-dire en jouant sur la contribution spectrale d'une ligne mélodique interprétée par un hautbois, dans chacun des deux haut-parleurs. L'expérience sépare, au préalable, les harmoniques paires et impaires du signal et assigne chacun de ces ensembles à un haut parleur. Puis en jouant sur un paramètre de fluctuation entre les deux voies, le dispositif est capable soit de reconstituer le timbre du hautbois (fluctuations corrélées), soit de déstructurer l'image auditive en donnant l'illusion d'un son de clarinette d'un côté et de soprano de l'autre (fluctuations décorréliées). Selon un schéma très simple (voir figure 64), cette expérience réalise donc une recomposition – ou une décomposition – de la source sonore dans le milieu même de propagation.

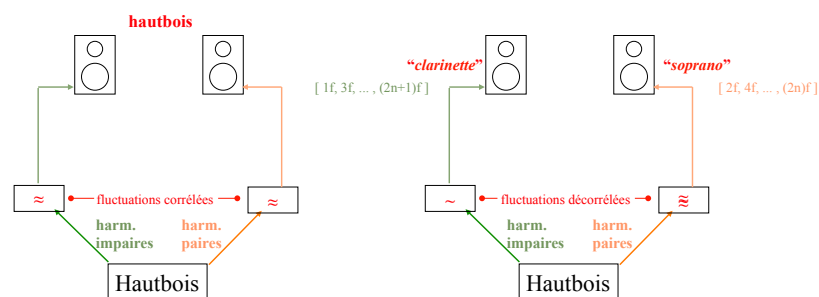


Figure 64. Expérience de séparation d'un son de hautbois (McAdams et Reynolds, 1983)

Plus récemment, une expérience similaire de synthèse dite « Spatio-Operationnelle Spectrale » (*SOS – Spatio-Operational Spectral synthesis*) a été proposée [Topper et al., 2002]. Cette approche consiste également en une technique de recombinaison spatiale et temporelle de composantes spectrales sur le principe suivant : un signal d'entrée est décomposé suivant une technique d'analyse / synthèse additive, chaque composante étant ensuite restituée par l'un des haut-parleurs d'un dispositif multicanal réparti (voir figure 65). La manipulation des paramètres inter-composantes comme, par exemple, le délai, permet alors de placer l'auditeur dans une situation perceptive ambiguë entre l'identification d'un signal unique et la perception de composantes individuelles.

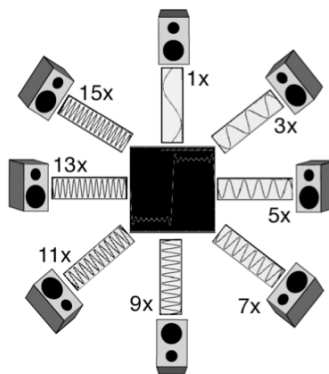


Figure 65. Dispositif multicanal pour la synthèse Spatio-Operationnelle Spectrale (SOS)

¹⁸ L'expérience n'a pas donné lieu à publication. ; ces résultats ont néanmoins été utilisés pour l'enregistrement de *Archipelago*, oeuvre musicale composée par Roger Reynolds en 1983.

A travers ces deux exemples proches du principe de Synthèse Additive Spatiale, il est donc intéressant de constater que la manipulation de composantes spectrales dans le milieu de propagation du signal acoustique produit des effets perceptibles et, pour certains, perceptivement pertinents.

II.5.1.1 Analyse et reproduction des propriétés spectrales et spatiales du violon

II.5.1.1.1 Préambule

L'étude et la mise en œuvre de cette approche SAS se sont focalisées sur un instrument qui présente une figure de rayonnement relativement complexe et fortement dépendant de la fréquence : le violon. Cette première étude a abordé la question expérimentale de la directivité de l'instrument au moyen d'une méthode indirecte, proposée par Weinreich reposant sur le principe de réciprocité acoustique (voir § II.3.1) et a tenté une première implémentation de synthèse augmentée *via* le logiciel Modalys. Ces travaux se sont déroulés dans le cadre du stage de Master de Rik Vos (TU Delft) et on fait l'objet d'une communication scientifique au 150^{ème} meeting de l'*Acoustical Society of America* (ASA) [Warusfel et al., 2004a] ainsi que d'une publication dans le journal de la société danoise d'acoustique (*NAG Journal*) [Vos et al., 2003].

II.5.1.1.2 Publication [Vos et al., Journ. Acoust. Soc. Nether., 2003]

Analysis and reproduction of the frequency spectrum and directivity of a violin

H.J. Vos^(a), O. Warusfel^(b), N. Misdariis^(b) and D. de Vries^(a)

^(a) Laboratory of Acoustical Imaging and Sound Control, Faculty of Applied Sciences
Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

^(b) Institut de Recherche et Coordination Acoustique/Musique (IRCAM)
1, place Igor Stravinsky, 75004 Paris, France

II.5.1.1.2.1 Introduction

The violin has been the topic of a lot of research over the past few hundred years [Hutchins et al., 1997]. Its characteristic sound does not have an omnidirectional radiation, but a directivity pattern depending on the frequency. This sound 'lights' the acoustical properties of the surroundings, for example the room or the concert hall, giving an extra dimension to the live performance.

The directivity pattern of a music instrument has been synthesised on for example a Wave Field Synthesis system [Baalman, 2003], but not yet on the Spherical Harmonic system of IRCAM. The goal of this research is thus to link the reproduction system to the characteristics of a natural instrument. In more details, the study will consist of (a) the characterization of an acoustical radiation transfer function, (b) the identification of the modes [frequencies, losses, shapes and their radiation pattern associated], (c) the exportation to a sound synthesis engine based on physical modeling, and (d) the reproduction of the radiation by a directivity controlled diffusion device. We have to state here, that no extensive search has been done in the literature, so that we emphasised on the feasibility of the proposed method and also that this first study is a preliminary exploration that should give perspectives for following works.

This article consist of four parts. The first part describes the necessary theory, the second parts describes the directivity measurements on a real violin. The third and fourth part deal with the analysis and reproduction of the measured characteristics, respectively.

II.5.1.1.2.2 Theory

II.5.1.1.2.2.1 Measurements

The measurements were not being applied the direct way, where the bridge or string is excited and the sound radiation is measured, but the inverse way, where an external source excites the violin from a certain direction. This method is based on the reciprocity assumption, which states that if the body vibrates strongly due to a sound wave with given frequency and direction, it will radiate sound waves with the same power and direction as this excitation. Weinreich has exhaustively described this method in 1980 [Weinreich et al., 1980]. By doing this, the main uncertainty in the assumption is the fact, that a radiated field from the body has the body as its acoustical centre, but the external field has the source as its acoustical centre. We neglected the influence of this shift in sound centre.

From the literature, it is known that the violin has standing wave solutions of the vibrations of its physical structure, which are called the modes of the violin [Weinreich et al., 1980 ; Weinreich, 1985]. The analysis consisted of finding the peaks in the measured power spectrum. We stress here, that we assumed the peaks to appear at the modal frequencies of the body. This makes a precise analysis of the violin modal shapes superfluous, but will not violate the overall reproduction of the violin sound.

II.5.1.1.2.2.2 Reproduction

The reproduction is based on the Huyghens theorem. If we have two sources of which the sound fields are equal at a surrounding surface S , then the fields outside this surface at observer point P are equal. See figure 66. Reproducing with a loudspeaker array, the goal is to find functions that minimise the difference between the two fields [Misdariis, 2001a] which finally constitute the set of filters to apply to the loudspeaker array.

This principle is, in formula,

$$\sum \alpha_i \cdot P_i(x, y, z, t) = T(x, y, z, t) + \varepsilon, \quad \text{at surface } S, \quad (\text{Eq. 23})$$

with P the acoustic field of the different loudspeakers on S , α_i the complex frequency-dependent filter coefficients, T the original field. The reproduction error is given by ε .

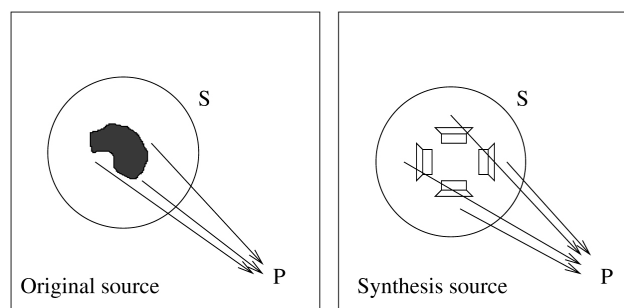


Figure 66. Reproduction of an original field with loudspeakers.

The used reproduction unit has been designed for reproducing the zeroth and first order spherical harmonics. These spherical harmonics are the mathematical solutions of the 3D Laplace equation in spherical co-ordinates.

This method is applied from a pragmatic point of view and results in four sets of filters for the whole system, which is shown in figure 67. These four sets of filters are permanently processed and the combination of the elementary directivities with a set of

gains g_i on figure 67 obtains the control. See [Misdariis et al., 2001a] for more details about this method.

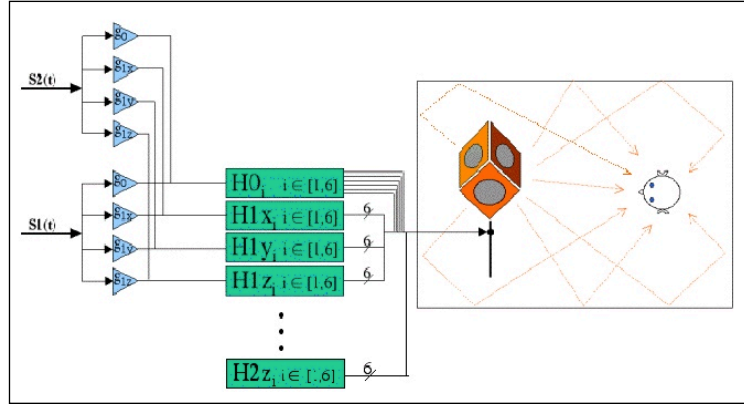


Figure 67. The reproduction signal paths of two different signals S_1 and S_2 . The gains g_i set the monopole and dipole contents of the reproduced field.

Now, the set-up with these filters has one great benefit. If the 3D spatial Fourier decomposition of the sound field is taken, which is nothing more than calculating the resemblance between the field and the spherical harmonics, then the coefficients are directly related to the gains. The formulas that describe this connection are

$$\begin{aligned}
 W &= a_{00}, \\
 X &= \frac{1}{\sqrt{2}} (a_{11} + a_{1,-1}) \\
 Y &= \frac{1}{\sqrt{2}i} (a_{11} - a_{1,-1}) \\
 Z &= a_{10}
 \end{aligned}
 \tag{Eq. 24}$$

with W the gain of the monopole, and X, Y , and Z the gains of the 3 dipoles. The a 's are the complex spatial Fourier coefficients calculated from the original violin directivity field at a certain frequency.

II.5.1.1.2.3 Measurement

One part of this research was the actual measurement of the violin body radiation. The violin was a 70 year old revised study violin.

II.5.1.1.2.3.1 Set-up

To measure the vibrations of the body, we used a piezo-electric crystal glued between the violin bridge foot and head. We have chosen a commercially available piezo-electric 'pill' of 1.3 mm thickness. This pill was cut into a small slice as wide as the bridge. A bridge was cut into two pieces at exactly the smallest width between the two notches. The head and the foot were then filed to achieve the same height of the bridge as an untouched one when the crystal is glued between the two parts. See figure 68.

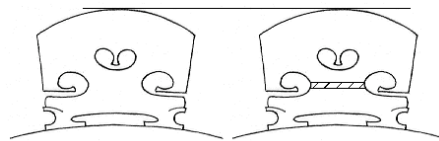


Figure 68. Schematic picture of the placement of the piezo-electric bar.

At this step, we clearly state the fact of using a single sensor, instead of two sensors put at the bottom of each foot of the bridge, like Weinreich [Weinreich, 1997] did in his experiment. In fact, on this preliminary approach, we are mainly interested in the reproduction procedure of the violin sound so that the actual goal is to pick up a relevant signal of the instrument and play it back with a transfer function that takes into account its spatial behaviour: in this case, we can justify our choice by the consideration that using two different signals would lead to a more complex situation in terms of calibration, mixing, ... etc. But, on the other hand, this pragmatic approach of the problem can also be valid within the physical modeling context – the large framework where we finally place the study – if one considers the input of the model in accordance with our experimental output. In practice, we neglect, in a first order approximation, the internal movements of the bridge by integrating it in the global movement of the violin top plate.

A B&K 2635 charge amplifier amplified the electric signal. The assumption is that the amplitude of the electric signal will be linear to the force acting between the body and the strings, and we assume this force to be linear with the amplitude of the body vibrations.

To measure the characteristics of the sensor itself, we sandwiched the sensor between a so-called vibrating pot and a contra mass. The characteristics of the set-up itself were measured with a thin-film piezo-electric which was assumed to have a flat spectrum over the audio frequencies.

The external source was a 2-way coaxial loudspeaker box, a Tannoy System 600. To reduce the influence of the characteristics of the loudspeaker, we took a measurement microphone (B&K 2619 with membrane B&K 4149) that captured its specifications. Afterwards, we corrected the measured signals with this reference by division of the complex frequency spectra.

The violin is mounted on a metal pipe-and-clamps construction, such that the it can both vibrate as freely as possible in the frequency range of interest, and is rigidly mounted. Between the clamps and the violin, small pieces of felt were put to reduce the influence of vibrations of the construction. See figure 69. Suspension will always alter the behaviour of the violin. Marshall, however, has shown that this happens anyhow with different players [Marshall, 1985b], thus this effect is neglected.

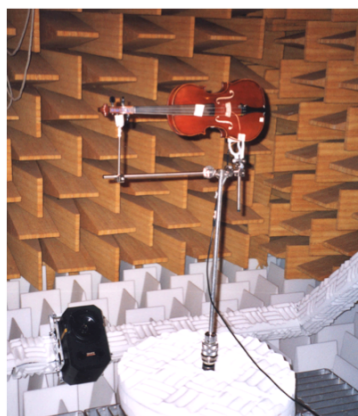


Figure 69. The violin in the anechoic room.
The vertical pipe at the bottom is connected to a B&K turning table.

Furthermore, to reduce the influence of the construction, the reference microphone (that captured the loudspeaker specifications) was mounted on the same construction. This setup was rebuilt twice to check its repeatability, which was quite good. The construction was attached to a B&K turning table. The loudspeaker was mounted on an elevation arm, which elevated from -40° to 90° . To cover the whole sphere of $180^\circ \times 360^\circ$, the violin was flipped over its long axis by hand. The distance between the loudspeaker and the sensor was 2.0 m.

The capture software was an IRCAM analysis program, called AMS. This software returned the impulse response of the system in the time domain. The electric cross-talk between the loudspeaker and sensor was eliminated by deleting the first microseconds of the captured signal; in this time interval, the acoustic wave had not reached the violin yet, but only contained the electric cross-talk.

The sampling frequency was 48kHz, with a sample size of 8192 points, giving a frequency resolution of approximately 6Hz. Weinreich has predicted that the directivity field of a violin can change every 44 Hz, due to the resolution of resonance frequencies of the top plate and lower plate of the violin body [Weinreich, 1985]. Thus, the frequency resolution appeared to be sufficient. The angular resolution was 10° , which was thought to be sufficient.

II.5.1.1.2.3.2 Spectrum

The spectrum of the violin is depicted in figure 70. As we had an irregularly spaced spatial grid, the spectrum is a weighted sum over all the directional spectra.

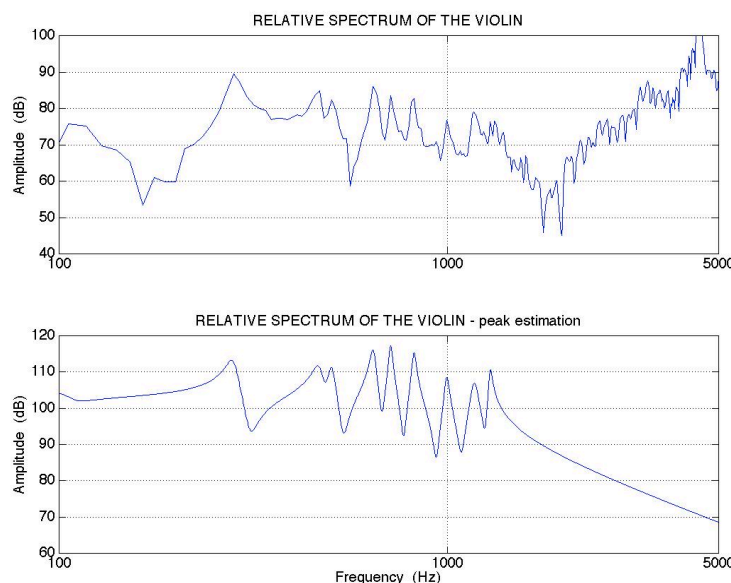


Figure 70. The relative spectrum of the violin.

The bottom spectrum is the relative spectrum of the 10 first estimated peaks (see section Analysis).

The amplitude unit is arbitrary, as the only concern was the relative amplitudes between the subsequent modes. The spectrum was compared with the results from both Weinreich [Weinreich, 1997a] and Marshall [Marshall, 1985] for frequencies up to 1 kHz, but as they conclude that each violin has its own frequency response, no exhaustive conclusions can be drawn from the comparison. Moreover, as we used a

single sensor experimental set-up we can't find the same results as theirs, because we are not observing exactly the same excitation point of the structure (a single point in the middle of the bridge body vs. two correlated points at the bridge feet), so that we can't detect, for instance, rolling movements of the bridge where each foot acts in contra-phase.

Some directivities are depicted in figure 71. The amplitude of the pressure is constant at the shown surface.

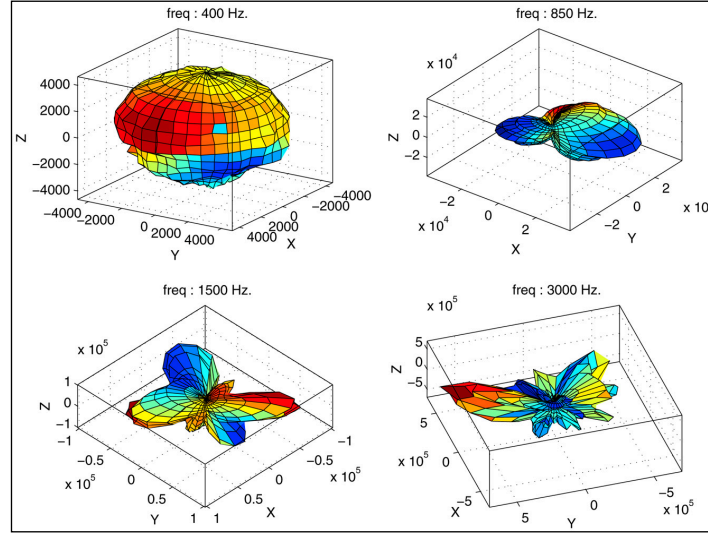


Figure 71. Some shapes of the directivity of the violin.
The f-holes point towards the positive x-axis, the fingerboard lies along the positive y-axis.

From figure 70, it shows that the frequency response is stronger for frequencies above 3kHz. We are not sure about the cause, but suspect that the piezo-electric sensor had a very good acoustic coupling with the higher-frequency body vibrations that was absent in the calibration setup. Thus, an overcompensation of those frequencies might have occurred.

II.5.1.1.2.4 Analysis

The analysis is done in two steps. The first step is to find the values of the modal parameters which are the relative amplitude, the frequency and the damping. The second step is to analyse the directivity of the modes. Both these modal characteristics are needed by the reproduction software.

II.5.1.1.2.4.1 Frequency Analysis

A simple peak-searching algorithm found the resonance frequencies, and the height of the peak was the relative amplitude. As we were only interested in frequencies up to 2kHz, we neglected the higher frequencies. This limit was forced by the reproduction system, which could not deal with frequencies higher than 2kHz.

The width of a peak defined the damping. From the theory of damped oscillations, [Fletcher et al., 1998], the quality factor Q is given by:

$$Q = \frac{f_n}{2|f_1 - f_2|} \quad (\text{Eq. 25})$$

where f_n is the unforced resonance frequency, and f_1 and f_2 are the two -3dB frequencies. This formula assumes $Q^2 \gg 4$, which holds for all peaks. The loss factor was calculated as $2/Q$ [Fletcher et al., 1998].

To check the ability of the suggested method to reproduce the power spectrum, the found values of the resonances are inserted in the forced resonance equation [Weinreich, 1985], and added. This yields the overall modeled frequency response. For convenience, this response is shown in figure 70 at the bottom, for the ten first peaks detected. The high frequencies are left out in the synthesised spectrum.

II.5.1.1.2.4.2 Directivity analysis

The directivity of the violin in the total frequency range has been found by the spatial Fourier decomposition.

Power involved in the zeroth and first order is a sum over the squared coefficients a_{ij} ,

$$P = \sum_{i=0}^1 \|a_{ij}\|^2, \quad ||j|| \leq i. \quad (\text{Eq. 26})$$

The applicability of the proposed method can be checked by comparison of the power contents of the monopole and dipoles to the overall power contents. If the ratio was found to be too low, we would have to conclude that the monopole and dipoles could not describe the radiation field. Then, the reproduction should include the quadripole and higher orders as well. The results are shown in figures 72 and 73.

From figure 72, it is clearly seen that up to 700 Hz the monopole dominates the power spectrum, and from 700 Hz up to 1.5 kHz the monopole and dipole powers are at similar amplitudes. In figure 73, higher frequencies show the diminishing influence of both these harmonics. From this we conclude that above 1.8 kHz the directivity can only be described by harmonic orders higher than zero and one.

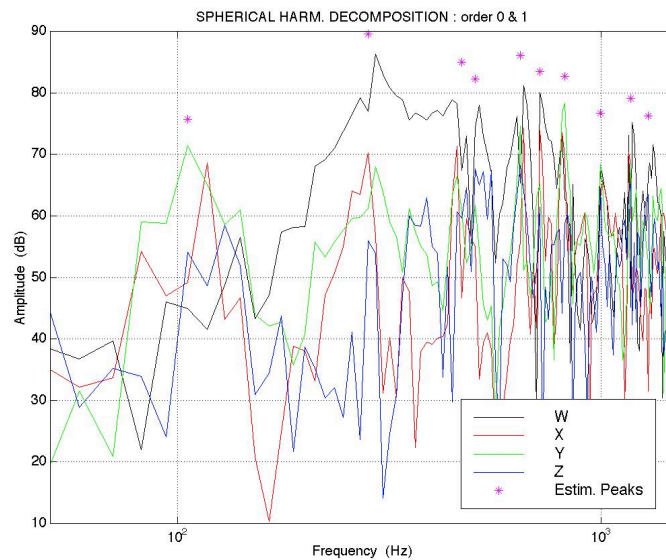


Figure 72. Spectrum of the coefficients of the violin showing its omnidirectional behaviour for frequencies up to 650Hz.

The solid line shows the monopole coefficients, the other three lines show the 3 dipole coefficients. The vertical scale is in dB with an arbitrary reference.

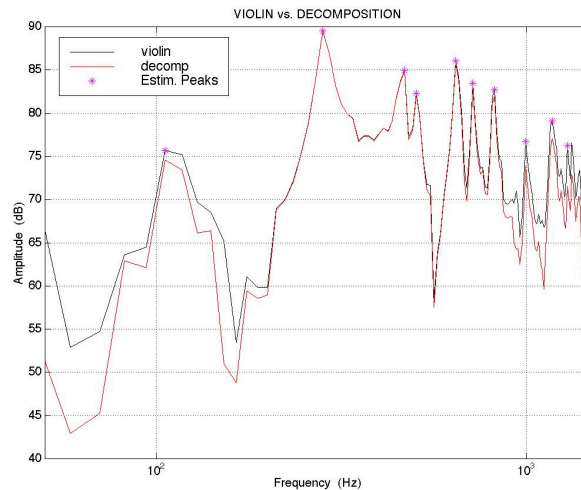


Figure 73. The violin spectrum (black , upper line) and the power of the decomposition in only zeroth and first order harmonics (red, lower line).

II.5.1.1.2.5 Reproduction

The reproduction system contains six loudspeakers in a cubic configuration. Due to spatial aliasing, the maximum applicable frequency of the used system is 2 kHz. As we have seen, the decomposition of the radiation patterns can be used directly as multiplication factors in the field reproduction. In other words, if the system is fed with four tracks, every track will produce one of the four spherical harmonics. Together they build the 3D sound field.

The sound signal of the violin, which is to be fed into the reproduction system, is made by a modal-based synthesis program called *Modalys*, which decomposes complex structures into simple structures having vibrating modes [Adrien, 1991].

Briefly, *Modalys* is based on the decomposition of a complex musical (or any other) instrument into simple structures like plates, strings and tubes. The vibration modes are calculated analytically, for which the parameters and formulas are stored in a database. The structures are connected through physical interactions, like a "pluck", "bow", "hit" or "glue" connection. An actuator is moved, bringing the energy for excitation. The output is the velocity at a user-defined point of the structure. This program is extended with a directivity processing function for the simple structures by the first author. The violin is modeled as having four strings, connected through a violin bridge glued onto a vibrating plate, which acts as a sound emitting body. Occasionally, a bow strikes the strings, with a "bow" connection (stick-and-loose algorithm). Modeled fingers shorten the strings by pressing the strings to a fingerboard and consequently playing the tune.

Inside this environment the connection to the found parameter values was made. The description of this connection is split into two parts:

- First, the values for the resonances of the vibrating plate are set equal to the values that were found from the measurements: the centre frequency per mode and the damping. We took 10 modes, as this value covered roughly all the peaks in the spectrum. *Modalys* also needs a shape definition of all the single modes, which are not compatible with the 3D modal shapes of a real violin. Therefore, we used the standard square thin plate modes that are available in the *Modalys* database

- Secondly, the directivity information from the measurements had to be included inside *Modalys*, see figure 74. As we have seen, every mode has its own directivity that is caught into 4 parameters: the contribution of the sound to the monopole W and 3 dipoles X, Y and Z (see equation 26). Thus, the sound signal of every mode of the vibrating plate was multiplied with the four W, X, Y and Z factors, yielding four tracks. These tracks are fed into *La Timée* in the adequate format and played.

A more formal description of the method is the following: if we look at one of the four tracks, then this signal is the weighted sum over the subsequent modes. The four tracks have different weighting coefficients per mode, the coefficients depending on the spherical harmonical contents of that mode.

See figure 74 for the overall scheme; the matrix multiplication is the implementation of the weighted sums.

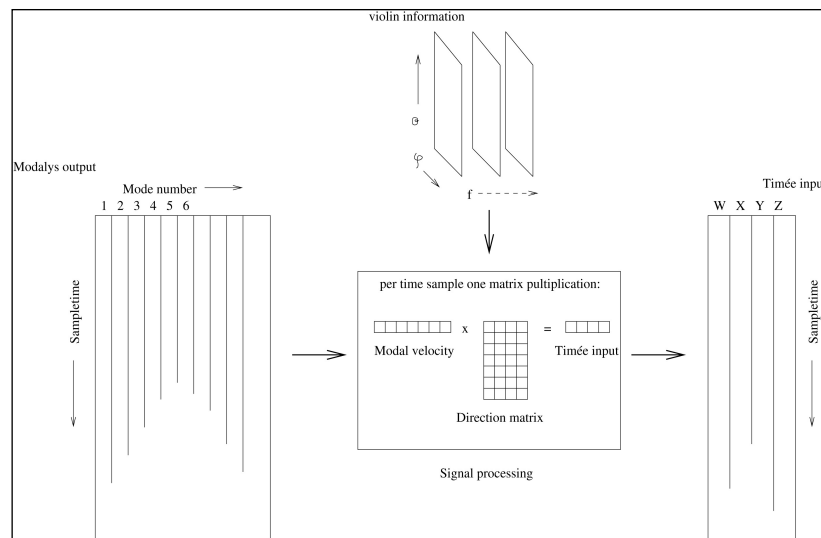


Figure 74. Schematic representation of signal processing.

The played tunes that have been constructed in *Modalys* are a short tango and a plucked-string tune. During the playing of these tunes, the directivity information can just be heard after close listening. The reason for this faint effect was already seen in figure 71. Here, it was shown that only in the range of 0.7 kHz to 1.5 kHz the dipoles do have significant power, but still no more than the monopole.

II.5.1.1.2.6 Conclusions

In this research, the spectrum and directivity of the violin body are measured and captured into parameters concerning the frequencies of the body resonances, damping, relative amplitude and directivity information.

Monopoles and dipoles up to 1.8 kHz can describe the violin directivity, above this frequency the directivity includes also quadripoles and higher spherical harmonical functions.

This information is used in a reproduction system consisting of instrument modeling software and a cubic loudspeaker set-up. Weighted sums over the modeled modes of the violin body composed four different tracks, which acted as input sound tracks for the reproduced monopole and dipoles.

This research showed that the sound of a violin can be reproduced in all of its aspects, but only for frequencies up to 1.8 kHz. The spectrum showed a large amount of high frequencies, and few directivity changes below 0.7 kHz. Resonances up to 1.8 kHz

showed more directivity changes, which could be heard during reproduction stage. The used cubic actuator does not support much higher frequencies and no higher order harmonics, which should be topic of further research.

This research was the first step for the mode-based instrument measurement, characterisation, synthesis and reproduction. A next step can be set to identify all modes of the instrument, using more sensors than one, and exporting the results to the synthesis program in a more complex structure, thus enlarging the resemblance between the violin 3D structure and its physical model.

II.5.1.2 Mesure et contrôle du rayonnement appliquée à la synthèse sonore du violon

II.5.1.2.1 Préambule

Dans la continuité du travail précédent, une seconde étude a instruit la question du rayonnement du violon – dans la perspective d’une application en Synthèse Additive Spatiale – par une méthode plus classique et directe comportant des mesures par vibrométrie laser, pour la partie vibratoire, et par antennerie, pour la partie acoustique. Les données expérimentales recueillies ont pu être comparées notamment avec les analyses modales effectuées par Marshall [Marshall, 1985b] et ont pu servir à implémenter une seconde instance de SAS dans le logiciel Modalys.

Ces travaux se sont déroulés dans le cadre du stage de Master I de Guillaume Defrance (Université du Maine) et ont fait l’objet d’une communication scientifique au VIII^{ème} Congrès Français d’Acoustique (CFA) [Misdariis et al., 2006].

II.5.1.2.2 Publication [Misdariis et al., Congrès Français d’Acoustique, 2006]

Contrôle du Rayonnement Appliqué à la Synthèse Sonore : Description de "Synthèse Additive Spatiale"

Nicolas Misdariis, Guillaume Defrance, Olivier Warusfel, René Caussé
Ircam-CNRS-STMS, 75004 Paris, France

II.5.1.2.2.1 Introduction

Pour cette étude, nous partons du principe que la pertinence de simulation d'un instrument de musique dépend à la fois du moteur de synthèse (production) et du moteur de rendu (diffusion). Nous faisons également le constat que les technologies de diffusion actuelles ne prennent pas ou peu en compte les propriétés spatiales du son et nous nous proposons d'intégrer ce degré de liberté supplémentaire dans le cadre générale de la synthèse sonore numérique.

L'étude s'appuie donc, d'une part, sur un moteur de synthèse sonore par modélisation physique basé sur le formalisme modal, et d'autre part, sur un système de diffusion capable de reproduire un champ acoustique donné en utilisant le principe de décomposition sur la base des harmoniques sphériques. Nous faisons par ailleurs l'hypothèse que le couplage de ces deux éléments permettra une recombinaison à la fois spectrale et spatiale des propriétés sonores du modèle, au lieu de réaliser cette synthèse uniquement dans le domaine spectral, comme c'est le plus souvent le cas (par exemple, dans la synthèse additive).

Ainsi, soit un instrument donné, la procédure de travail peut se résumer par les quatre principales étapes suivantes : a) observation expérimentale de l'instrument – à partir d'un ou plusieurs points significatifs du point de vue vibro- acoustique –, b) analyse spectrale et spatiale des mesures pour définir les modes de vibration et leur fonction de

directivité associée, c) modélisation mécanique et acoustique à partir des résultats précédents, d) couplage des deux modèles *i.e.* réalisation de la synthèse sonore "augmentée" des propriétés spatiales. Ces différentes étapes sont présentées et détaillées dans la figure 75.

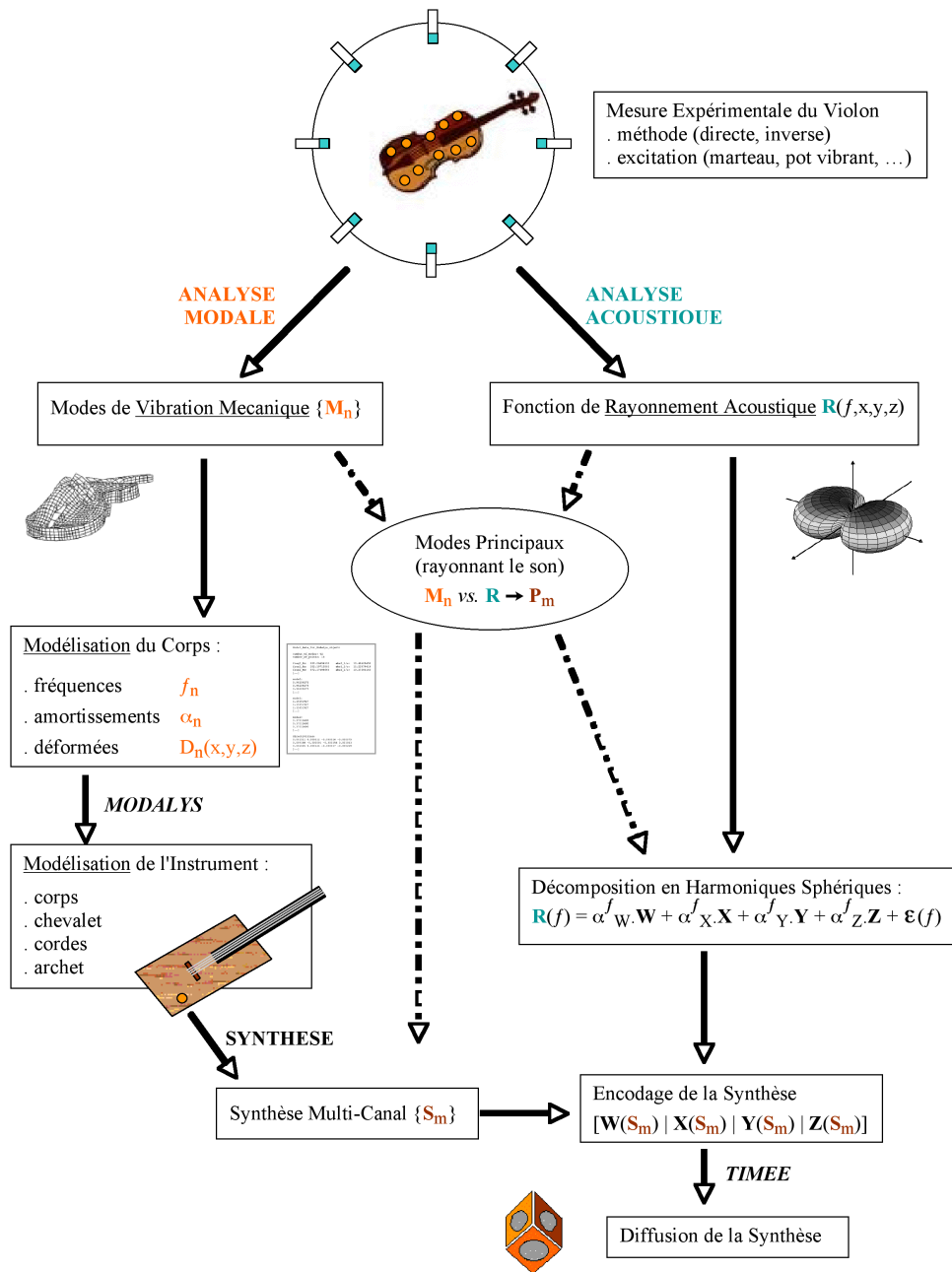


Figure 75. Détail des étapes de la *Synthèse Additive Spatiale*

La SAS pouvant finalement être vue comme un procédé de synthèse spatio-opérationnelle telle que définie dans [Topper et al., 2002], développé dans le cadre de la synthèse modale et d'une technique de diffusion spécifique.

II.5.1.2.2.2 Outils de l'étude

II.5.1.2.2.2.1 Synthèse sonore

Le moteur de synthèse sonore retenu pour notre application est *Modalys*, développé par l'Ircam à partir des travaux de J. M. Adrien [Adrien, 1988]. Il appartient à la catégorie

des synthétiseurs par modélisation physique et possède, en outre, la spécificité de reposer sur le formalisme modal. Ainsi, dans cet environnement, le comportement vibratoire (mécanique ou acoustique) d'une structure est décomposé sur les fonctions de déformations élémentaires de la base modale (modes) et utilise – en faisant l'hypothèse de linéarité – le principe de superposition pour reconstituer la fonction de déformation complexe de la structure. En conséquence, la définition des modèles s'effectue au moyen d'une description modale uniforme : nombre de modes et, pour chacun d'eux, paramètres modaux (fréquence, amortissement et déformée).

Cela étant, l'instrument de synthèse le plus simple dans Modalys se compose de trois éléments : deux objets en interaction (excitateur et résonateur) et une connexion définissant cette interaction, dans lequel sont localisées les non-linéarités du modèle. Des procédures annexes, notamment la définition des accès et des contrôleurs, permettent, par ailleurs, de mettre en oeuvre la synthèse (voir [Adrien, 1991] et [Eckel et al., 1995]).

II.5.1.2.2.2 Synthèse du rayonnement

II.5.1.2.2.2.1 Généralités

La technique de reproduction exploite le principe de Huygens (voir § 3.1 de [Nicol, 1999]) et repose sur une méthodologie classique : décrire sur une même surface entourant la source, la contribution élémentaire de chaque transducteur composant le système ainsi que la fonction spatiale cible à reproduire, puis effectuer une optimisation (au sens mathématique) pour induire le meilleur jeu de filtres qui approxime l'égalisation du champ de pression sur cette frontière. La méthode produisant, *in fine*, un ensemble de filtres correspondant à des fonctions spatiales statiques (voir [Warusfel et al., 1997]). La figure 76 résume ce principe pour un système de N haut-parleurs disposés sur les faces d'un polyèdre régulier.

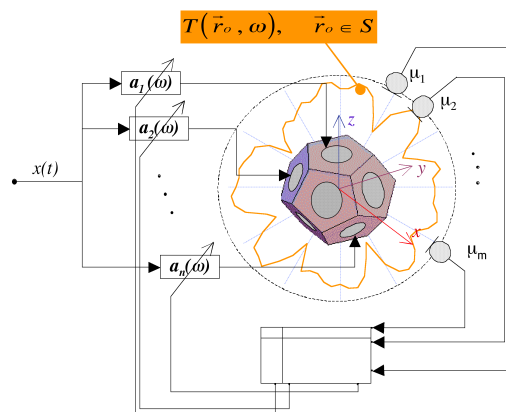


Figure 76. Synthèse du rayonnement (principe)

II.5.1.2.2.2.2 Implémentation et Propriétés

Cela étant, d'un point de vue pratique, cette implémentation "statique" présente de fortes contraintes pour des applications audio et musicales. En effet, d'une part, le jeu de filtres associé à une directivité-cible dépend par nature des propriétés géométriques et acoustiques du système de transducteurs, et d'autre part, les capacités de modification en temps-réel – cruciales pour ce genre d'application – imposent des procédures d'interpolation de filtres qui conduisent inévitablement à des artefacts de traitement du signal, notamment à cause de la phase des filtres.

Une approche plus pragmatique (issue d'autres applications audio telles que Ambisonics ou l'implémentation multi-canal du rendu binaural [Jot et al., 1998]) consiste à synthétiser en permanence un sous-ensemble de directivités élémentaires à partir desquelles une directivité composite sera ensuite reconstruite au moyen, cette fois-ci, d'un simple jeu de gains. La procédure s'effectue alors en deux étapes : encodage de la directivité sur un format multi-canal intermédiaire, puis décodage spécifique au système de reproduction, utilisant les filtres adaptés à la technologie et au type de décomposition spatiale choisis. Nous nous attacherons ici à la famille des harmoniques sphériques, chacun des filtres de décodage étant calculé suivant la procédure décrite précédemment (voir § II.5.1.2.2.2.1) en considérant l'harmonique sphérique concernée comme directivité-cible à approximer (voir [Misdariis, 2001a]). Globalement, cette implémentation présente deux avantages majeurs : indépendance de l'encodage vis-à-vis du système de reproduction et possibilité de traiter plusieurs voies simultanément (à coût de calcul quasi-constant), avec, notamment pour conséquence directe, le contrôle de la directivité en fonction de la fréquence (figure 77). Ce qui, considérant le fait que les instruments de musique présentent presque tous cette propriété, nous relie directement au sujet de cette étude : la *Synthèse Additive Spatiale*.

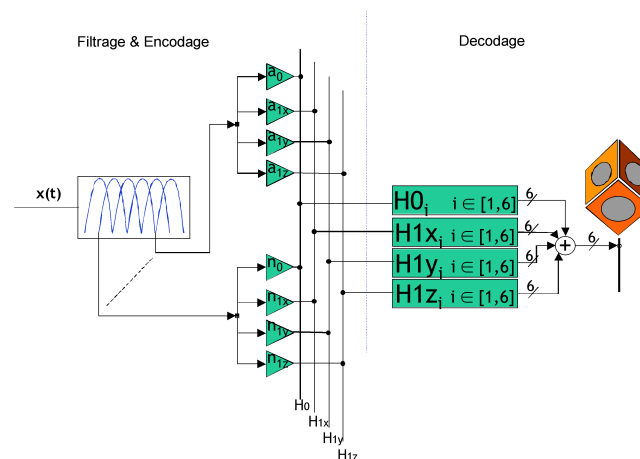


Figure 77. Synthèse du rayonnement (contrôle de la directivité en fonction de la fréquence)

II.5.1.2.2.3 Objet de l'étude : le violon

II.5.1.2.2.3.1 Présentation

Le violon fait l'objet de multiples recherches depuis de nombreuses années et ses caractéristiques ont été étudiées de manière approfondie tant sur le plan fonctionnel (voir [Jansson, 2002]), mécanique, acoustique que du couplage entre comportement vibratoire et rayonnement. Nous mentionnons ici certains des résultats obtenus, en grande partie à la lumière de trois références majeures dans le domaine ([Marshall, 1985b], [Weinreich, 1997a], [Bissinger, 2004]).

Du point de vue mécanique, dans une expérience impliquant un marteau d'impact frappant le chevalet et un accéléromètre placé sur 190 points de mesure, Marshall [Marshall, 1985b] met évidence 35 modes propres entre 100 et 1330 Hz. Il les classe, en outre, en 3 grandes catégories : i) les modes de flexion (B), ii) les modes de cavité (A), iii) les modes de plaque (P). Certains de ces principaux modes sont répertoriés dans le Tableau 1 (voir § II.5.1.2.2.3.4).

Du point de vue acoustique, Weinreich [Weinreich, 1997a] démontre, dans une expérience mettant en jeu le principe de réciprocité, le comportement du violon globalement omnidirectionnel en basse fréquence (100 - 1000 Hz) et très fortement dépendant de la fréquence au-delà de 1 kHz. Il en déduit, par ailleurs, d'éventuelles conséquences sur la perception auditive de l'instrument, comme par exemple, sa capacité "d'éclairer" la salle à chaque note différente jouée au-delà de cette limite.

II.5.1.2.2.3.2 Expérience préliminaire

Dans une phase préliminaire de l'étude, une mesure acoustique du violon a été réalisée par la méthode inverse, en utilisant le paradigme expérimentale de Weinreich [Weinreich, 1997a]. Le violon a donc été sondé de manière indirecte en plaçant le récepteur (cellule piézo-électrique) au niveau du chevalet et en scrutant l'espace au moyen d'un émetteur (système Tannoy 600) délivrant le signal d'excitation (voir [Vos et al., 2003]).

Outre une évaluation des paramètres (fréquence et amortissement) de certains modes propres, cette première expérience a permis de confirmer le principe de décomposition du champ de pression sur la base des harmoniques sphériques et a mis en évidence l'effet de troncature : d'une part la prédominance de la composante monopolaire en-dessous de 700 Hz confirme le comportement omnidirectionnel du violon en basses fréquences, d'autre part la comparaison entre le spectre de puissance évalué à partir des mesures et celui reconstruit après décomposition montre qu'au delà de 1.8 kHz, les ordres supérieurs seront nécessaires pour décrire le champ acoustique rayonné (fig. 78).

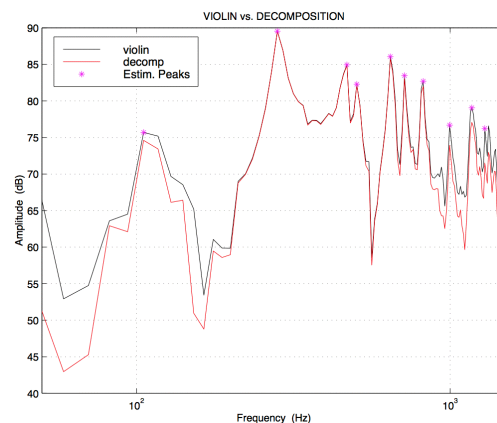


Figure 78. Spectre de puissance du violon mesuré (noir) et reconstruit (rouge), à partir des harmoniques sphériques d'ordre 0 et 1.

Cette première approche s'est cependant avérée incomplète dans la mesure où les données recueillies ne permettaient pas de décrire correctement un modèle physique associé, et notamment de caractériser les déformées modales.

II.5.1.2.2.3.3 Dispositif expérimental

Un protocole expérimental a donc été mis en oeuvre pour réaliser des mesures vibratoires et acoustiques comparables. Dans les deux cas, la partie "excitatrice" de la chaîne de mesure est constituée d'un pot vibrant B&K 4810, relié à l'interface RME FireFace 800 d'un Macintosh PowerBook doté de l'environnement Max/MSP dans lequel est généré le signal d'excitation : un sinus glissant balayant la plage 20Hz - 22 kHz (échelle logarithmique, 185 msec.). Par ailleurs, une micro-perforation est réalisée sur chacun des pieds du chevalet pour recevoir la pièce métallique, supposée rigide,

effectuant la liaison entre le pot vibrant et le violon. Enfin, le violon est maintenu en position par deux pinces situées au niveau de la base du cordier et de la volute (fig. 79).

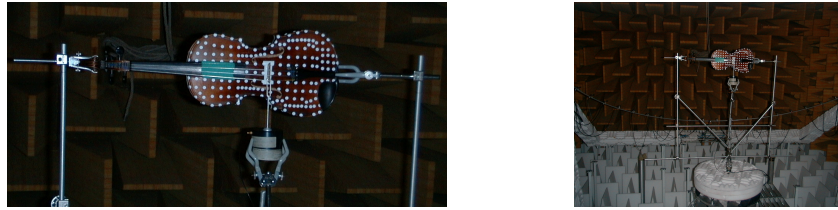


Figure 79. Violon - dispositif expérimental de mesure

II.5.1.2.2.3.3.1 Mesures vibratoires

Les mesures vibratoires ont été effectuées en chambre calme au moyen d'un vélocimètre laser Ometron VH300+. Elles s'effectuent sur un maillage préalablement matérialisé par des pastilles réfléchissantes sur les deux parties principales du violon : la table d'harmonie (199 points) et le fond (203 points). Pour chaque point, l'observation est répétée plusieurs fois jusqu'à obtenir un indice de focalisation satisfaisant, au sens des préconisations du vélocimètre (au minimum, 3 diodes vertes allumées).

II.5.1.2.2.3.3.2 Mesures acoustiques

Les mesures acoustiques ont été effectuées en chambre sourde, au moyen d'un banc de microphones à électret (marque Monacor, type MCE2000) disposés en arc de cercle. Le violon est situé au centre d'un appareillage mécanique automatisé, constitué d'une table tournante B&K assurant les déplacements dans le plan horizontal (0/360°) et d'un bras assurant les déplacements dans le plan vertical (0/90°) (figure 79) : ainsi, en n'opérant qu'un seul retournement, les deux parties du violon peuvent être caractérisés sur une sphère complète entourant la source. Par ailleurs, l'échantillonnage spatial de cette mesure est imposé par l'écartement des microphones sur l'antenne : 7.5°.

II.5.1.2.2.3.4 Premiers résultats

Dans un premier temps, une analyse fréquentielle est réalisée sur l'ensemble des réponses impulsionnelles du maillage (échantillonnage à 48 kHz, FFT sur 4096 points) et la recherche des modes propres de vibration s'effectue en recoupant les données spectrales de déformation de plusieurs points caractéristiques de la structure (figure 80).

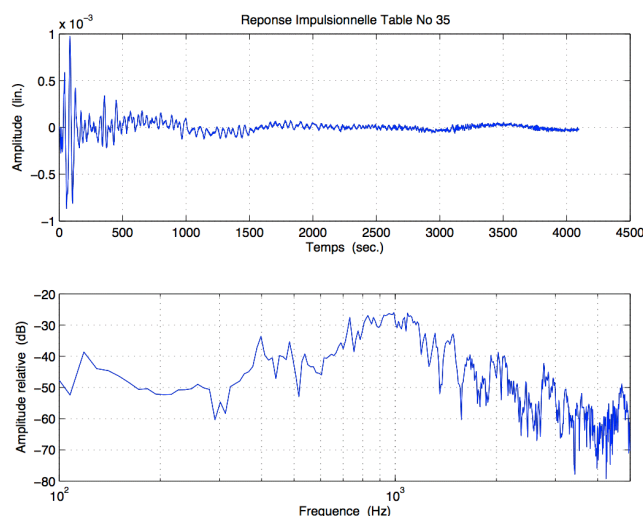


Figure 80. Exemple de réponse impulsionnelle (en temporel et en fréquentiel) au point N° 35 du maillage de la table

Les résultats, pour la plupart d'entre eux, sont placés dans le Tableau 1, en regard de ceux de Marshall [Marshall, 1985b].

Marshall [8]	Freq. mesurées	Marshall [8]	Freq. mesurées
163.2 (B)	151	574.1 (B)	600
187.3 (B)		656.1 (P)	
225.0 (B)		665.6 (B)	670
278.2 (A ⁰)	269	690.8 (P)	690
303.4 (B)	300	760.3 (B)	780
410.5 (P)	388	839.5 (A ³)	840
434.9 (B)	420	882.0 (B)	900
466.1 (B)	460	930.1 (P)	926
470.8 (P)		1173.5 (B)	1170
478.2 (A ¹)	495	1273.1 (P)	1260

Tableau 1. Fréquences propres du violon (Hz) – recoupement biblio.

Concernant la mesure de rayonnement, une synthèse des résultats peut s'illustrer par le produit de la décomposition en harmoniques sphériques à l'ordre 0 et 1 (figure 81) où l'on retrouve globalement la prédominance du rayonnement monopolaire en basses fréquences et l'émergence des contributions dipolaires à partir d'environ 700 Hz.

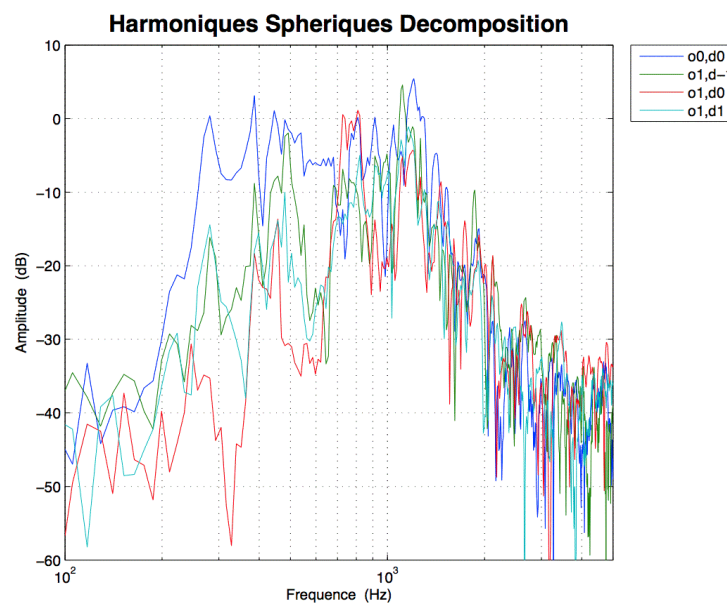


Figure 81. Décomposition du champ de pression rayonné par le violon ; base des harmoniques sphériques (ordre 0 et 1)

II.5.1.2.2.4 Perspectives

En terme d'analyse, ces premiers résultats doivent bien évidemment être affinés et complétés, notamment en considérant le comportement de la caisse du violon non contrainte, *i.e.* sans l'action couplée du chevalet et des cordes. Ils permettent néanmoins de franchir une étape supplémentaire dans le processus de *Synthèse Additive Spatiale*. Les points restant à traiter avant d'atteindre notre objectif consiste, d'une part, à exploiter l'intégralité des données vibratoires pour construire un modèle de violon cohérent dans *Modalys*, et d'autre part, à examiner les relations entre modes de vibration et modes de rayonnement afin de sélectionner les modes mécaniques qui interviennent de manière prépondérante dans les propriétés acoustiques de l'instrument.

Nous pourrions alors jouer séparément les modes propres les plus pertinents en terme d'énergie acoustique rayonnée, pondérés par leurs composantes dans la base des harmoniques sphériques, comme illustré figure 82.

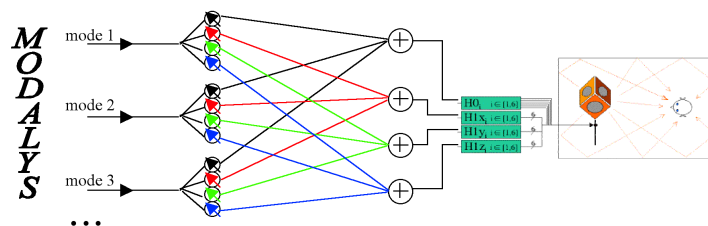


Figure 82. Procédure générale de *Synthèse Additive Spatiale*

II.5.1.2.2.5 Conclusion

Cette étude porte sur le développement d'une procédure de couplage d'un moteur de synthèse par modélisation physique avec une technologie de diffusion sonore capable de reproduire les composantes spatiales d'un champ acoustique.

Nous avons décrit l'approche générale ainsi que les premiers résultats consécutifs à sa mise en oeuvre, le but final étant de réaliser une *Synthèse Additive Spatiale*, à la place d'une "simple" synthèse "additive" réalisée uniquement dans le domaine spectral.

II.5.2 Evaluation perceptive du contrôle de la directivité

En parallèle au développement de l'approche de Synthèse Additive Spatiale, des travaux ont été initiés afin d'étudier d'un point de vue perceptif, le contrôle de la directivité. Ils s'inscrivent dans un cadre de recherche plus large qui s'intéresse à des mécanismes de perception et cognition spatiales mis en jeu dans la relation audition / espace et qui posent l'hypothèse générale suivante : le rayonnement acoustique est perceptible par le sujet humain.

En effet, en partant d'un principe de description des propriétés d'une source sonore selon trois dimensions : spectrale, temporelle et spatiale, et en faisant le postulat – néanmoins relativement fort – que les systèmes électroacoustiques actuelles restituent quasi fidèlement les deux premières dimensions (spectro-temporelles), on observe que l'oreille humaine est capable de distinguer un son provenant d'une source naturelle et ce même son "sortant" d'un haut-parleur. Autrement dit, si nous entendons une différence entre le son acoustique et le son électro-acoustique du violon, c'est que nous entendons des différences de propriétés spatiales – la signature spatiale – entre ces deux sources.

Cette réflexion fait suite à des travaux préliminaires sur la question, notamment ceux de Mzali qui s'est intéressée à l'influence de la directivité sur la perception d'une scène sonore et, plus précisément, sur le phénomène de ségrégation de flux auditifs [Mzali, 1997]. Dans cette étude, l'expérience centrale est une tâche de reconnaissance de la similitude de deux mélodies présentées successivement, la première étant jouée en présence d'un flux distracteur dont on varie l'indice de directivité. Elle s'intéresse à l'influence de cet indice sur la capacité à séparer les deux flux et donc à comparer les deux mélodies. Les premiers résultats montrent que même si l'indice de directivité a un effet sur la perception des signaux, il n'est cependant pas un critère suffisant pour différencier les deux flux.

II.5.2.1 Cadre général

L'étude s'est inscrite dans le cadre d'un stage de Master (Alexandre Lang, UTC) et a suscité une collaboration avec l'équipe d'O. Gapenne et C. Lenay (UTC/équipe CRED¹⁹ – *Cognitive Research & Enaction Design*) [Misdariis et al., 2008].

Elle se situe, par ailleurs, au terme de nombreuses études, expériences et réalisations techniques autour de la question du contrôle de la directivité ainsi qu'un certain nombre de collaborations et productions artistiques exploitant ce principe à des fins musicales²⁰. Elle se propose de compléter les diverses procédures de validation objective menées au cours des travaux antérieurs par une approche plus subjective se focalisant sur les percepts liés au rayonnement acoustique et à son contrôle par des moyens électroacoustiques.

Pour cela, elle mobilise plusieurs principes techniques de modélisation du rayonnement, dont certains ont été en partie validés dans les études précédentes :

- une référence (haut-parleur seul) qui ne permet aucun contrôle et émet le signal selon sa propre directivité ;
- un système simplifié (quatre haut-parleurs dos-à-dos) qui, au moyen d'un traitement du signal rudimentaire et en première approximation, permet de constituer des figures de directivité basiques dans le plan horizontal des oreilles d'un auditeur (azimuth) : monopole, dipôle et cardioïde – combinaison du monopole et du dipôle (voir figure 83, gauche) ;
- un système complet (*Timée* multi-voies) qui, associé à des filtres de contrôle optimisés, permet de reproduire ces mêmes figures de directivité basiques mais avec plus de précision, notamment en termes de dimensionnalité spatiale et de bande passante fréquentielle (voir figure 83, droite).

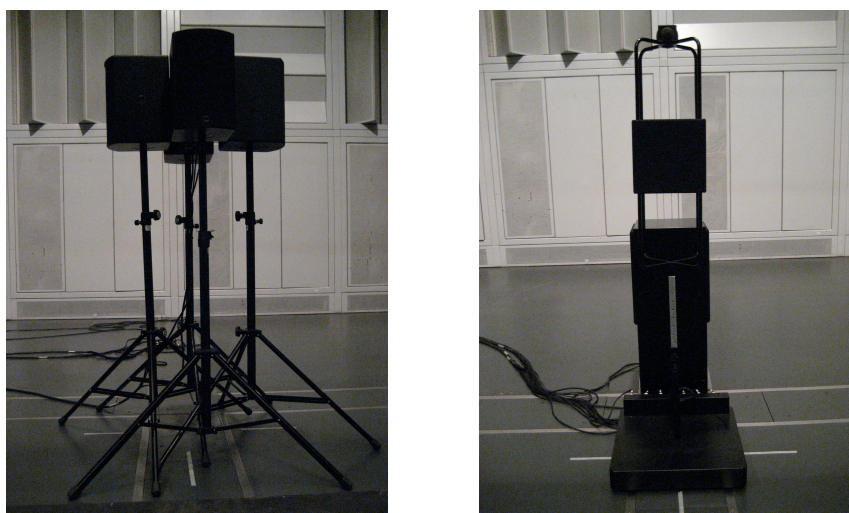


Figure 83. Système simplifié (4 haut-parleurs) et système complet (*Timée* multi-voies)

A noter que la configuration intermédiaire dite "simplifiée" est en premier lieu motivée par la structure du plan d'expérience qui croise deux facteurs principaux :

- le nombre de degrés de liberté du système de diffusion,
- le niveau de complexité – ou de précision – du contrôle de la directivité.

¹⁹ <http://www.utc.fr/cred/>

²⁰ parmi les principales productions : « Duell » (François Nicolas, 2001), « Five Imaginary Spaces » (Oliver Schneller, 2002), « Aliae Naturae » (Valerio Sannicandro, 2004), « Le Jardin des Fleuves » (O. Schneller, 2004)

Cependant, la seconde motivation liée à ce système simplifié est d'ordre musical. En effet, à l'époque de cette étude, le compositeur Marco Stroppa a fait part de son intérêt pour une utilisation effective d'un système de diffusion à directivité contrôlée en situation de production musicale – et donc potentiellement de tournée. D'où l'idée d'une « Timée simplifiée » qui utiliserait du matériel standard – donc plus facilement transportable, voire louable – et qui, associée à un traitement du signal adapté, permettrait un certain contrôle du rayonnement des sources électroacoustiques. Cette idée a donc été ébauchée et partiellement validée expérimentalement dans le cadre de cette étude, pour être ensuite effectivement exploitée en concert par M. Stroppa – associé à Arshia Cont en tant que Réalisateur en Informatique Musicale –, notamment pour sa pièce « ... of Silence », pour saxophone et électronique (création mondiale, au Japon, en 2007 ; création française, à l'Ircam, en 2009)²¹.

II.5.2.2 Objectifs

L'objectif est donc d'étudier l'écart perçu auditivement entre le son naturel et le son "artificiel", et de quantifier les variations perceptibles du rayonnement. Cette investigation est rendue possible d'un point de vue technologique par le contrôle du rayonnement acoustique qu'offrent des systèmes multi haut-parleurs associés à des algorithmes de traitement du signal – tels que la Timée présentée plus haut (§ II.3.3) – et qui permettent de maîtriser, dans une certaine mesure, les propriétés spatiales des sources électroacoustiques.

Les objectifs généraux de l'étude sont donc d'évaluer perceptivement le contrôle du rayonnement acoustique en identifiant les principaux facteurs influençant la perception de la directivité ou de ses effets. En particulier, l'étude tente de répondre à la question fondamentale suivante : le contrôle du rayonnement permet-il de réduire l'écart perçu entre un son provenant d'une source réelle et le son de cette même source réelle, enregistré puis rediffusé par un système électroacoustique ?

Indirectement, l'objectif est donc également musical dans la mesure où les résultats de l'étude pourront éventuellement militer en faveur d'une version simplifiée de Timée (si, par exemple, peu de différences sont perçues entre les systèmes complet et simplifié) qu'elle a par ailleurs contribué à développer et caractériser expérimentalement.

II.5.2.3 Expérience

L'étude comprend un important volet expérimental qui comporte principalement trois niveaux :

- concevoir les stimuli en faisant varier les facteurs contrôlés (variables indépendantes) : type de sources, niveau de précision du modèle de rayonnement et dimensionnalité du système de rendu ;
- conditionner les stimuli afin de les rendre compatibles avec une expérience en laboratoire (enregistrement binaural) ;
- conduire les tests d'écoute en fonction des facteurs mesurés (variables dépendantes) : largeur, distance et réalisme de la source.

II.5.2.3.1 Stimuli

Du point de vue des sources sonores, l'expérience considère deux corpus distincts : une voix parlée/chantée (préalablement enregistrée et mesurée spatialement)

²¹ <http://brahms.ircam.fr/works/work/19563/>

et une plaque frappée (synthétisée pour l'occasion et dont les données spectrales et spatiales sont calculées à partir de formulations théoriques).

II.5.2.3.1.1 Plaque frappée

Le stimulus de type plaque frappée est choisi pour pouvoir intégrer l'approche de Synthèse Additive Spatiale (voir § II.5.1) dans l'évaluation perceptive du contrôle de la directivité. Ainsi, le modèle de plaque choisi est du type : plaque rectangulaire isotrope, bafflée, en appui simple sur son contour.

En première approximation, cette structure rayonne en configuration dipolaire – s'apparentant ainsi à un instrument de type "gong" –, et est donc compatible avec les premiers niveaux de précision des deux systèmes de diffusion considérés dans l'étude ("simplifié" et "complet").

Plus précisément, la vibration et le rayonnement acoustique de cette structure peuvent être décrits au moyen d'une formulation analytique (moyennant, malgré tout, l'hypothèse de champ lointain pour l'acoustique) :

- les fonctions propres de propagation (Φ_{mn}) – et valeurs propres associées (ω_{mn}) – des ondes de flexion dans une plaque mince, isotrope, de longueur a , de largeur b , d'épaisseur h et constituée d'un matériau de masse volumique ρ , de module d'Young E et de coefficient de Poisson ν , sont données par [Lesueur, 1988] :

$$\omega_{mn} = \sqrt{\frac{D}{\rho h}} \left(\left(\frac{m\pi}{a} \right)^2 + \left(\frac{n\pi}{b} \right)^2 \right) \quad \text{où } D = Eh^3/12(1-\nu^2) \quad (\text{Eq. 27})$$

$$\Phi_{mn}(x, y) = \sin\left(\frac{m\pi x}{a}\right) \cdot \sin\left(\frac{n\pi y}{b}\right) \quad (\text{Eq. 28})$$

- la fonction de rayonnement d'une plaque isotrope, bafflée possédant les mêmes caractéristiques sur son mode $[m, n]$ est donnée par [Derogis et al., 1994 ; Derogis, 1997] :

$$\left| p(R, \theta, \phi) \right|^2 = \frac{4\gamma_{mn}^2}{R^2} \cdot \left(\frac{ab\rho}{mn\pi^3} \right)^2 \cdot \left[\frac{\left(\frac{\cos}{\sin} \right) \left(\frac{\lambda a}{2} \right) \cdot \left(\frac{\cos}{\sin} \right) \left(\frac{\mu b}{2} \right)}{\left(\left(\frac{\lambda a}{m\pi} \right)^2 - 1 \right) \cdot \left(\left(\frac{\mu b}{n\pi} \right)^2 - 1 \right)} \right]^2 \quad (\text{Eq. 29})$$

où γ désigne le module de l'accélération, les cosinus et sinus sont utilisés respectivement lorsque m (ou n) sont paires et m (ou n) sont impaires, et où λ et μ sont définis comme suit $\lambda = k \sin \theta \cos \phi$ et $\mu = k \sin \theta \sin \phi$ (k étant le nombre d'onde).

La synthèse sonore du modèle de plaque est effectuée dans Modalys à partir des données calculées analytiquement (fréquences et déformées modales) et d'une excitation standard de type "frappée" – le modèle de marteau est, lui aussi, générique car il ne rentre pas en compte dans le paramétrage de l'expérience. Par ailleurs, les coefficients d'amortissement sont ajustés afin d'obtenir deux types de stimuli : une plaque peu résonante ('PL3') et une plaque très résonante ('PL5'). Les données acoustiques sont, quant à elles, encodées sur les deux premiers degrés de la base des harmoniques sphériques (monopole et 3 dipôles), et les coefficients de pondération (W ,

X, Y, Z) associés sont assignés, pour chaque mode, au modèle de plaque implémenté dans Modalys.

Le modèle finalement obtenu est ainsi compatible avec le niveau de précision le plus élevé des systèmes de diffusion et, moyennant toutes les approximations effectuées, peut s'apparenter à un comportement vibro-acoustique réel – ou tout au moins réaliste – de la structure considérée.

II.5.2.3.1.2 Voix parlée/chantée

Le stimulus de type voix est choisi à la faveur d'une collaboration avec Brian Katz (LIMSI, CNRS) qui mène des recherches sur la voix parlée et chantée en partant du principe que le rayonnement de la voix est un paramètre important pour l'intelligibilité et la communication. Dans ce but, il a effectué de nombreuses mesures de directivité de la voix d'un locuteur et d'un chanteur prononçant une série de phonèmes [Prezat, 2006]. Les mesures ont été effectuées avec le dispositif équipant la chambre anéchoïque de l'Ircam (voir § II.4.2.1) ; et les analyses montrent des différences de rayonnement importantes entre phonèmes, surtout dans certaines bandes de fréquence (par exemple, 630-1250 Hz ou 2500-3150 Hz).

Ces données constituent le second corpus de l'étude. Les enregistrements bruts sont reconditionnés en séquences afin de respecter plusieurs critères d'ordre méthodologiques (durée et acceptabilité du stimuli, etc.). Les séquences choisies pour l'expérience – également sur des critères méthodologiques tels que cohérence globale, absence de contenu sémantique, répartition d'occurrence, etc. – sont répertoriées dans le tableau ci-dessous (Tableau 2) :

N°	Séquence	N°	Séquence
1	[m][o][s]	1	[a][A][a]
2	[an][s][o]	2	[A][an][i]
3	[o][f][an]	3	[an][A][I]
4	[o][ch][a]	4	[an][O][an]
5	[n][an][ch]	5	[o][I][o]
6	[f][i][m]	6	[I][O][a]
7	[ch][i][m]	7	[i][I][i]
8	[ch][a][m]	8	[o][a][O]
9	[s][a][an]	9	[o][O][o]

Tableau 2. Séquence des phonèmes parlés (droite) et chantés {min. : Lab2 ; maj. : Lab3} (gauche)

En première approximation, la voix rayonne soit en configuration monopolaire (plutôt en basses fréquences), soit en configuration cardioïde (plutôt en moyennes et hautes fréquences) et est donc compatible avec les premiers niveaux de précision des deux systèmes de diffusion considérés dans l'étude ("simplifié" et "complet").

Plus précisément, les mesures de rayonnement acoustique de la voix, effectuées par Prezats et Katz, permettent une modélisation de la directivité. La décomposition de ces données sur les deux premiers degrés de la base des harmoniques sphériques aboutit à la définition des coefficients de pondération (W, X, Y, Z), en fonction de la fréquence, à appliquer à chacun des phonèmes enregistrés. A noter que, dans ce cas et du fait de la nature des données (non issues d'une synthèse), le principe de Synthèse Additive Spatiale n'est que partiellement mis en œuvre mais peut malgré tout être appliqué et représenter, de nouveau, un comportement réel – ou tout au moins réaliste – du point de vue spectral et spatial, vis-à-vis du stimulus considéré.

II.5.2.3.2 Apparatus

Comme mentionné précédemment, l'outil de diffusion des directivités se décline en trois dispositifs de dimensionnalité spatiale croissante : un haut-parleur de référence rayonnant principalement selon une seule dimension dans l'axe du point d'écoute (D1), quatre haut-parleurs formant un système simplifié rayonnant dans le plan horizontal (D2) et une version de la Timée multi-voies ayant a priori la faculté de rayonner dans les trois dimensions de l'espace (D3).

Par leurs degrés de liberté respectifs, ces dispositifs déterminent plusieurs modalités spatiales qui spécifient le degré de précision du contrôle de la directivité :

- le rayonnement "brut" du haut-parleur seul, c'est-à-dire sans contrôle de directivité (R1) ;
- le rayonnement monopolaire générant une pression uniforme sur une sphère (ou un cercle) entourant le dispositif (R4) ;
- le rayonnement dipolaire présentant 2 lobes de directivité dans une direction et une réjection importante perpendiculairement à cette direction (R3) ;
- le rayonnement cardioïde, combinaison du monopole et du dipôle, présentant un seul lobe de directivité (R2) ;
- le rayonnement "réel" mettant en œuvre le principe de Synthèse Additive Spatiale tel que décrit au § II.5.2.3.1 ci-dessus (R5).

II.5.2.3.3 Plan d'expérience

La combinaison des 2 facteurs expérimentaux – système de restitution et mode de rayonnement – avec les différentes natures de stimuli aboutit donc à un plan d'expérience qui peut être illustré suivant la figure 84 ci-dessous ; l'intérêt d'utiliser plusieurs systèmes de diffusion étant de pouvoir vérifier, le cas échéant, si le gain perceptif lié au contrôle du rayonnement est indépendant du dispositif technologique.

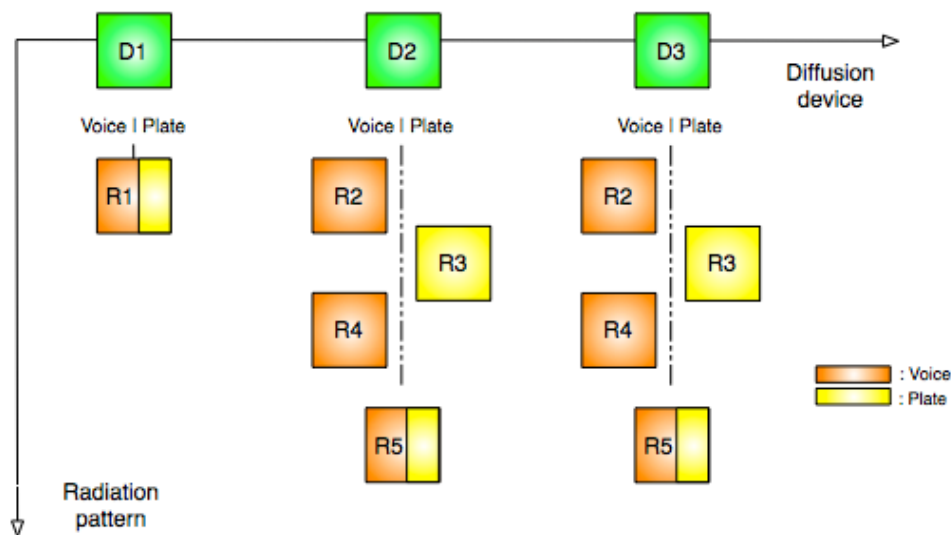


Figure 84. Etude perceptive du rayonnement : plan d'expérience
Combinaison des facteurs "système de restitution" et "mode de rayonnement"
en fonction du groupe de stimuli ; les pointillés séparent les modalités de rayonnement
utilisées pour la voix (gauche/orange) et pour la plaque (droite/jaune) –
R1 : sans contrôle ; R2 : cardioïde ; R3 : dipôle ; R4 : monopole ; R5 : "réel"

II.5.2.3.4 Protocole

Pour des raisons essentiellement pratiques et logistiques, le protocole expérimental privilégie une écoute en situation de laboratoire (cabine audiométrique)

plutôt qu'une écoute en situation "naturelle" où les sujets sont placés dans une salle, en présence des systèmes eux-mêmes (mais occultés).

Cette condition nécessite d'enregistrer la trace de chaque modalité de diffusion associée à un système – c'est-à-dire un couple (D_i , R_j) selon la nomenclature de la figure 84. Pour cela, la technique de captation et restitution binaurale est utilisée afin de pouvoir reproduire au mieux, sous casque, les informations spatiales délivrées par chacune des configurations, et notamment l'interaction de chaque système avec la salle de diffusion.

Les enregistrements binauraux sont effectués dans l'Espace de Projection (Espro) de l'Ircam au moyen d'une tête artificielle positionnée à une hauteur et une distance cohérentes par rapport à une écoute normale (voir figure 85).

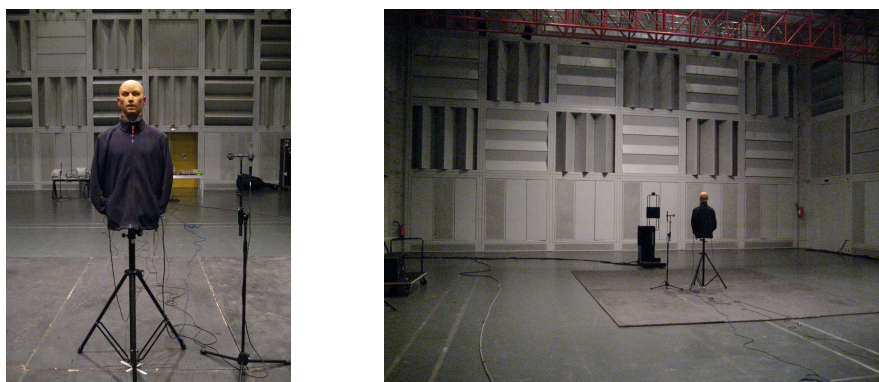


Figure 85. Tête artificielle pour enregistrements binauraux et position par rapport à la source (Espro)

Dans cette étape de pré-traitement, il est malgré tout important de noter que l'encodage binaural des effets de directivité repose sur les performances de la technique binaurale (voir [Hammershoi et al., 2002] pour une explication du principe) et fait, en particulier, une hypothèse forte par rapport à l'individualisation, à savoir : les filtres HRTFs (*Head-Related Transfer Functions*) du mannequin – buste + tête artificielle – sont compatibles avec ceux des sujets et ne provoquent pas d'artefacts susceptibles de dégrader la restitution (voir [Busson, 2006] pour une étude sur ce thème).

Cela étant, et à partir du corpus d'étude "binauralisé", le protocole expérimental utilise un paradigme classique de jugement sur des échelles sémantiques (voir § III.3.3 pour une présentation détaillée). L'un des problèmes inhérents à cette méthode étant la définition et la pertinence des attributs sémantiques associés à chacune des échelles, nous puisons dans la littérature sur les études de qualité sonore ou d'évaluation perceptive de grandeurs acoustiques, des dimensions qui s'avèrent faire sens et consensus pour notre cas d'étude. Parmi d'autres, les travaux de Lavandier sur l'évaluation physique et perceptive des enceintes acoustiques [Lavandier, 2005] et de Kahle sur la perception de la qualité d'un ensemble de salles de concert [Kahle, 1995] nous amène à définir trois critères d'évaluation des stimuli :

- la **largeur de source**, ou étendue de la source : renvoyant implicitement à la notion d'enveloppement ;
- la **distance perçue** : distance à laquelle le sujet positionne apparemment la source ;
- le **réalisme de la diffusion** : renvoyant à l'aptitude du système à reproduire le son provenant d'une source réelle – critère le plus difficile à expliquer aux sujets mais le plus à même de valider l'hypothèse de départ.

Le test d'écoute à proprement parlé comporte 28 sujets, à parité égale, possédant une expertise dans le domaine du son ou de la musique. Après la lecture d'une consigne d'expérience, les sujets évaluent 3 groupes de séquences sonores – plaque frappée, voix parlée et chantée – selon les 3 critères mentionnés au paragraphe précédent. L'expérience se déroule en cabine audiométrique, sous casque, au moyen d'une interface graphique programmée en *PsiExp* [Smith, 1995] (et couplée à Max/MSP pour la gestion de la partie audio de l'expérience). Chaque échelle sémantique est de type unipolaire et est matérialisée par un curseur allant de « peu » à « très » selon le critère considéré (par exemple de "peu réaliste" à "très réaliste"). Les ordres de présentation des 3 échelles, celui des 3 groupes de stimuli pour chaque échelle ainsi que des stimuli à l'intérieur des 3 groupes, sont gérés de manière aléatoire.

II.5.2.4 Résultats et Analyses

Les données brutes recueillies sont, pour chaque sujet, chaque groupe de stimuli (S), chaque système de diffusion (D) et chaque configuration de rayonnement (R), une valeur numérique continue comprise entre 0 et 1 correspondant à l'évaluation sur chacune des échelles sémantiques (E). Les différentes modalités de ces variables indépendantes (S, D, R) et dépendantes (E) sont répertoriées dans le Tableau 3 ci-dessous :

Modalité	Définition
S1	son de plaque frappée "peu résonante" ('PL3')
S2	son de plaque frappée "très résonante" ('PL5')
S3	séquences de phonèmes parlés
S4	séquences de phonèmes chantés
D1	système de diffusion à 1 dimension (haut-parleur)
D2	système de diffusion à 2 dimensions (Timée "simplifiée")
D3	système de diffusion à 3 dimensions (Timée multi-voies)
R1	rayonnement naturel (sans contrôle)
R2	rayonnement cardioïde
R3	rayonnement dipolaire
R4	rayonnement monopolaire
R5	rayonnement réel – ou réaliste
E1	évaluation du critère de « largeur »
E2	évaluation du critère de « distance »
E3	évaluation du critère de « réalisme »

Tableau 3. Modalités des variables et facteurs expérimentaux

A noter que dans ce plan d'expérience – par nature ou par hypothèse – des incompatibilités factorielles existent, notamment entre :

- D1 et {R2, R3, R4, R5} : le seul mode de rayonnement possible du haut-parleur de référence est le mode dit naturel, c'est-à-dire sans contrôle ;
- R3 et {S3, S4} : le modèle dipolaire n'est, a priori, pas cohérent avec le rayonnement de la voix
- {R2, R4} et {S1, S2} : les modèles monopolaire et cardioïde ne sont, a priori, pas cohérents avec le rayonnement de la plaque.

Cela étant, les analyses statistiques ont été effectuées en plusieurs étapes : i/ une statistique descriptive (moyenne, écart-type) donnant les grandes tendances de résultats ; ii/une analyse en composante principale (ACP) pour déterminer des

éventuelles sujets *outliers* ; iii/ une statistique inférentielle (analyse de la variance – ANOVA) complétée, le cas échéant, d’une analyse de contraste pour étudier précisément les interactions entre facteurs expérimentaux.

II.5.2.4.1 Moyennes et écarts-type

Les moyennes et écarts-type sont effectuées sur les 28 sujets. Un exemple de résultats est donné dans les figures 86, respectivement pour la plaque frappée ‘PL3’ "courte" (S1) et la voix chantée (S4). D’une manière générale, ces premiers résultats sont difficiles à interpréter tel quel, même si quelques tendances se dessinent notamment pour l’échelle de largeur. Cependant, ils ne permettent en aucun cas de conclure sur les effets perceptifs du contrôle de la directivité.

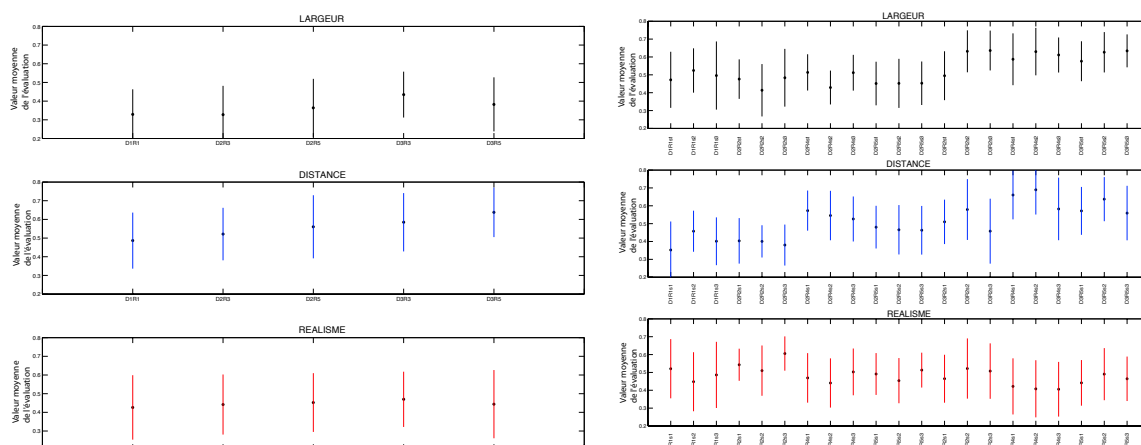


Figure 86. Etude perceptive du rayonnement : premiers résultats
Moyenne et écarts-type de l’évaluation suivant les 3 échelles : largeur, distance, réalisme des stimuli plaque frappée "courte" (gauche) et voix chantée (droite) selon les différentes configurations de rayonnement (en abscisse)

II.5.2.4.2 Définition des outliers

L’analyse en Composantes Principales (ACP) effectuée sur les sujets permet de vérifier leur cohérence mutuelle et d’éventuellement déceler certains d’entre eux, responsables d’une augmentation significative de la variabilité des résultats. Cette étape est d’autant plus importante que, pour obtenir des résultats significatifs par une analyse de variance de type ANOVA (étape suivante), la variabilité des données expliquée par les facteurs doit être supérieure à la celle due aux différences inter-sujets. Pour plus de précision, l’analyse ACP est effectuée en distinguant les groupes de stimuli (S) et les échelles (E) ; un exemple de résultats d’ACP sur les réponses respectives de réalisme (E3) sur la plaque frappée (S1) et de largeur (E1) sur la voix chantée (S4) est donné à la figure 87.

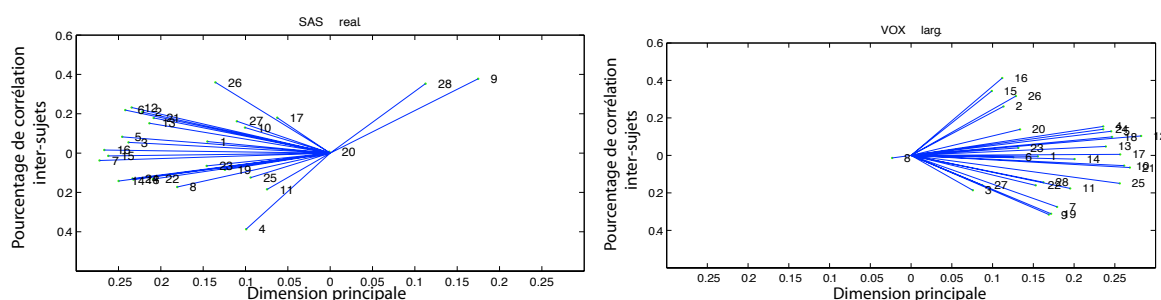


Figure 87. Analyse en Composantes Principales (ACP) sur les réponses des sujets
Couples {E3, S1} – réalisme de la plaque frappée (gauche) ; {E1, S4} – largeur de la voix chantée (droite)

La mise en œuvre systématique de l'ACP sur tous les couples (E_i , S_j) a abouti à la détection des 9 outliers dont les données ont été retirées pour la suite des analyses.

II.5.2.4.3 Analyses de la variance

Le plan d'expérience étant à la base incomplet (du fait de la présence du haut-parleur de référence n'acceptant qu'un seul mode de rayonnement), les analyses de variance (ANOVA – *Analysis of Variance*) ont été effectuées en deux étapes. De plus, l'analyse est en fait une MANOVA (*Multivariate Analysis of Variance*) [Abdi, 1987] afin de mieux prendre en compte d'éventuelles interactions entre les échelles d'évaluation et de possibles effets cumulés sur les différentes modalités de E.

La première analyse – ANOVA#1 – se restreint à 2 modalités de diffusion (D2 et D3) et teste l'hypothèse nulle suivante : le système de diffusion D et/ou le mode de rayonnement R n'ont pas d'effet simultané sur les jugements effectués selon les trois échelles. Cette analyse teste également l'indépendance des facteurs D et R. De plus, le facteur stimulus (S) y est réduit à une distinction entre "plaques" et "voix" et est introduit d'un point de vue expérimental pour accroître la robustesse de l'analyse (s'il s'avère que les effets de D et/ou R ne dépendent pas du type de stimulus, les données sur S pourront alors être moyennées).

L'ANOVA#1 effectuées sur les plaques – respectivement, les voix – indique globalement un effet significatif du système de diffusion (D) (lambda de Wilks²², $F(3,143)=4.34$, $p<0.01$ – resp., $F(3,664)=75.53$, $p<0.0001$) et du mode de rayonnement (R) (lambda de Wilks, $F(3,143)=3.74$, $p<0.05$ – resp., $F(3,664)=20.08$, $p<0.0001$) sur l'évaluation des critères, et plus particulièrement la distance (E2) ; les figures 88 et 89 illustrent ce résultat.

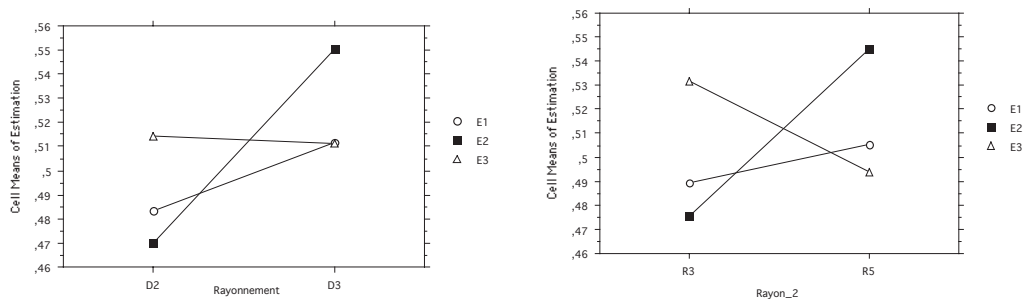


Figure 88. ANOVA#1 pour les plaques : effet des facteurs principaux sur la variable dépendante (E)
Syst. de diffusion (D), à gauche ; mode de rayonnement (R), à droite

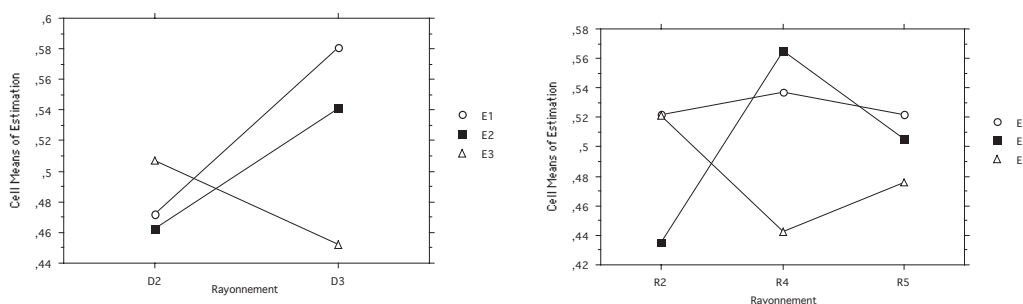


Figure 89. ANOVA#1 pour les voix : effet des facteurs principaux sur la variable dépendante (E)
Syst. de diffusion (D), à gauche ; mode de rayonnement (R), à droite

²² Le « lambda de Wilks » est la statistique associée à la MANOVA

Cependant, l'analyse montre également une différence significative du type de plaques – resp., de voix – dans l'évaluation des critères. C'est d'ailleurs, l'effet le plus important dans la mesure où les différences observées sur les échelles entre les 2 plaques – resp. les 2 voix – sont plus importantes que les différences liées aux facteurs D et R. Autrement dit, les effets des stimuli sont prépondérants par rapport à ceux des facteurs principaux ; la figure 90 illustre ce résultat pour les plaques (λ de Wilks, $F(3,143)=66.56$, $p<0.0001$).

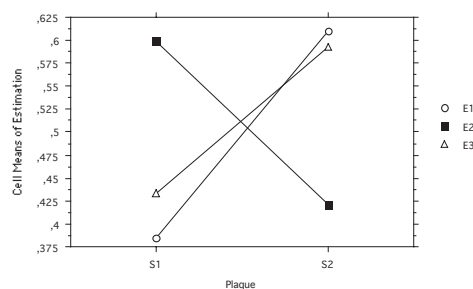


Figure 90. ANOVA#1 pour les plaques : effet du type de plaques sur la variable dépendante.
S1 : plaque "courte" ('PL3') ; S2 : plaque "longue" ('PL5')

Enfin, quelque soit le type de stimuli, l'analyse ne montre pas d'interaction entre le facteur système de diffusion (D) et mode de rayonnement (R) ce qui va permettre, par la suite, de coupler ces 2 facteurs principaux ; la figure 91 illustre ce résultat pour la voix.

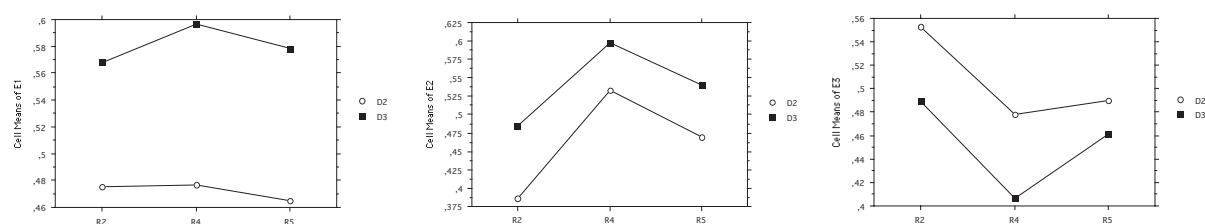


Figure 91. ANOVA#1 pour les voix : effet du mode de rayonnement (R) sur le système de diffusion (D) en fonction de la variable dépendante
(échelle "largeur" (E1), à gauche; échelle "distance" (E2), au milieu; échelle "réalisme" (E3), à droite)

La seconde analyse – ANOVA#2 – considère ainsi une réduction des facteurs système de diffusion (D) et mode rayonnement (R) en un facteur couplé (C) pouvant être interprété comme un facteur de complexification du système de diffusion ; la règle de couplage est explicitée dans le Tableau 4 ci-dessous :

Modalité	Définition
C1	couplage des modalités D1 et R1 (sans contrôle)
C2	couplage des modalités D2 et R2 (cardioïde)
C3	couplage des modalités D2 et R3 (dipôle)
C4	couplage des modalités D2 et R4 (monopole)
C5	couplage des modalités D2 et R5 (réel)
C6	couplage des modalités D3 et R2 (cardioïde)
C7	couplage des modalités D3 et R3 (dipôle)
C8	couplage des modalités D3 et R4 (monopole)
C9	couplage des modalités D3 et R5 (réel)

Tableau 4. Couplage des facteurs expérimentaux système de diffusion (D) et mode de rayonnement (R) en un facteur complexification du système de diffusion (C).

L'analyse ANOVA#2 teste donc l'hypothèse nulle suivante : La variable couplée C (complexification du système de diffusion) n'a pas d'effet sur les jugements effectués sur les trois échelles ; cette hypothèse étant en lien direct avec le postulat de départ sur l'effet perceptif du contrôle de la directivité.

L'ANOVA#2 effectuée sur les plaques – respectivement, les voix – indique globalement un effet significatif du facteur couplé (C) (lambda de Wilks, $F(3,143)=3.35$, $p<0.0005$ – resp., $F(3,664)=21.06$, $p<0.0005$), du stimulus (S) (lambda de Wilks, $F(3,143)=80.66$, $p<0.0005$ – resp., $F(3,664)=4.34$, $p<0.005$) ainsi qu'une interaction entre C et les stimuli (S) – mais uniquement pour les voix (lambda de Wilks, $F(3,664)=2.21$, $p<0.005$) ; la figure 92 illustre ce dernier résultat.

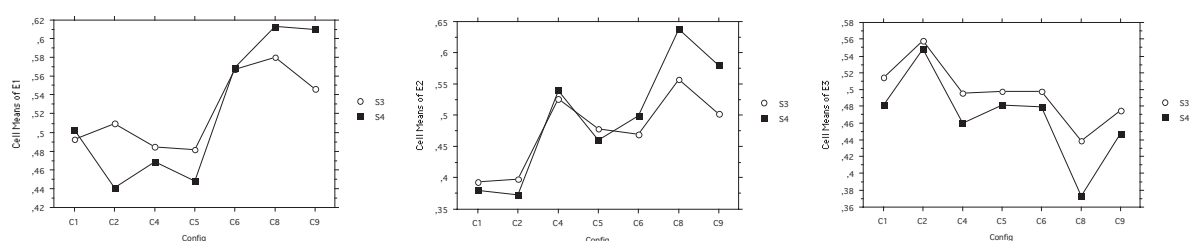


Figure 92. ANOVA#2, pour les voix : effet du facteur complexification du système de diffusion (C) sur les modalités du stimulus (S) en fonction de la variable dépendante (échelle "largeur" (E1), à gauche; échelle "distance" (E2), au milieu ; échelle "réalisme" (E3), à droite)

Sur la base de ces premiers résultats, une analyse complémentaire de contraste est effectuée pour tester 2 configurations particulières :

- Contr.#1 : haut-parleur seul {C1} vs. autres modalités, {C2 à C9}
- Contr.#2 : modes de rayonnement réels {C5,C9} vs. théoriques {C2,C3,C4,C6,C7,C8}

Les résultats de cette analyse peut globalement se résumer dans le Tableau 5 suivant :

	Largeur (E1)	Distance (E2)	Réalisme (E3)
Contr. #1	Pl.* / Vo.*	Pl. * / Vo.*	Pl. / Vo.
Contr. #2	Pl. / Vo	Pl. * / Vo.	Pl. / Vo.

Tableau 5. Récapitulatif des résultats des 2 analyses de contraste selon les stimuli (Pl., plaques ; Vo., voix) et les échelles d'évaluation (E) – les contrastes significatifs sont notés *

En résumé, le bilan de ces analyses valide en partie l'hypothèse de départ en montrant un effet global du rayonnement acoustique sur la perception. Cet effet, matérialisé par la complexification du système de diffusion – et donc la précision du contrôle du rayonnement – est surtout notable pour les critères de largeur et de distance. Dans l'état actuel de l'étude, et tel qu'il a été défini dans l'expérience, le critère de réalisme ne permet pas quant à lui de caractériser la perception du rayonnement. Par ailleurs, il est également important de noter que, dans cette expérience, certains groupes de stimuli ont un effet plus important sur le résultat que les facteurs principaux eux-mêmes, ce qui milite en faveur d'une modification et amélioration du corpus de test.

II.5.2.5 Conclusion et Perspectives

L'objectif initial de démontrer que le contrôle de la directivité tendait à réduire l'écart entre une source réelle et une source diffusée par un système électroacoustique à été partiellement atteint. En effet, les jugements effectués entre la référence (haut-parleur seul) et des systèmes offrant une directivité complexe présentent des résultats significativement en faveur de la complexification du système de diffusion.

Cela étant, ces résultats généraux doivent être nuancés dans la mesure où plusieurs contre-résultats apparaissent à l'issue des analyses statistiques : d'une part, la significativité n'est effective que pour certaines échelles (largeur et distance), voire n'est pas du tout vérifiée concernant la perception du réalisme de la scène. Et d'autre part, le type de stimulus semble, dans certains cas, jouer un rôle plus important que les facteurs expérimentaux eux-mêmes.

Ce bilan contrasté permet néanmoins d'envisager des perspectives d'amélioration et d'approfondissement du paradigme en explorant, par exemple, les quelques pistes suivantes : définir de manière plus pertinente les échelles d'évaluation (avec éventuellement des expériences préliminaires permettant de dégager des attributs cohérents), s'affranchir de l'étape d'encodage/décodage binaural en effectuant le test en situation d'écoute naturelle, réduire la diversité des stimuli et les confronter à une source réelle, et enfin améliorer les performances techniques du système de diffusion en permettant d'accéder à une modélisation du rayonnement aux ordres supérieurs et en affinant le principe de Synthèse Additive Spatiale sur lequel repose l'hypothèse de départ.

II.6 Conclusion

Cette seconde partie relative à la thématique de la **reproduction du son** a donc permis de poser les éléments fondamentaux – autant historiques que techniques – d'un dispositif de diffusion sonore pour rapidement s'intéresser à la question du rapport entre rayonnement naturel d'une source acoustique et directivité contrainte d'un haut-parleur ainsi qu'à la problématique du contrôle de cette directivité.

L'introduction de modèles pour le rayonnement du haut-parleur (piston plan) et la description du champ sonore (harmoniques sphériques) ont permis d'aborder cette problématique sous un angle théorique. Un état de l'art portant sur les questions de mesure et de contrôle de la directivité d'une source a, par ailleurs, mis en évidence la complexité du problème, notamment du point de vue expérimental.

Partant de là, le développement de travaux personnels ont présenté mes contributions à ce domaine de recherche à la fois en termes de mesure expérimentale du rayonnement instrumental (saxophone, cor de basset, violon), mais également sous la forme d'une approche originale du contrôle de la directivité reposant sur les propriétés élémentaires de la base des harmoniques sphériques. Une composante technologique de ces travaux m'ont par ailleurs permis de concevoir des dispositifs prototypes (système multi haut-parleurs à directivité contrôlable) qui seront utilisés autant d'un point de vue scientifique – pour valider nos hypothèses de recherche – que d'un point de vue artistique – pour confronter le paradigme à des configurations musicales mixtes –, donnant ainsi lieu à la création de plusieurs œuvres et installations d'envergure.

Le lien avec la thématique précédente (synthèse) s'est opéré par une extension du travail sous la forme d'une proposition originale de synthèse augmentée, appelée « Synthèse Additive Spatiale ». Ce paradigme combine une synthèse modale avec un dispositif de reproduction sonore capable de contrôler les propriétés spatiales de chaque mode rayonné ; la synthèse globale s'effectuant alors dans le milieu de propagation lui-même. L'idée, mise en œuvre techniquement, a en outre, été évaluée de manière perceptive, permettant ainsi de combiner à nouveau deux grandes thématiques abordées dans le document, à savoir la reproduction et la perception des sons.

Troisième partie

Perception

PARTIE III. Perception

III.1 Problématique

La **perception** est, par définition, l'acte de percevoir ou de « prendre connaissance », en mobilisant une, ou plusieurs, modalités sensorielles associées aux cinq sens : la vue, l'ouïe, le toucher, l'odorat ou le goût. Plus largement, elle désigne une « opération psychologique complexe par laquelle l'esprit, en organisant les données sensorielles, se forme une représentation des objets extérieurs et prend connaissance du réel »²³. La **cognition** est quant à elle, par définition, « l'ensemble des structures et activités psychologiques dont la fonction est la connaissance »²⁴ ou bien encore « [...] *the activity of knowing: the acquisition, organization and use of knowledge* » [Neisser, 1976]

D'un point de vue théorique, ces deux notions ont divergé pendant très longtemps en considérant, que la perception exploite des informations issues des organes sensoriels et aboutit à une représentation mentale des événements ou objets présents dans l'environnement d'un individu ; cette image mentale constituant le point de départ de processus psychologiques plus élaborés (attention, organisation, mémoire) censés être attribués au concept de cognition. La divergence exprime donc un processus de transformation de l'information de type modulaire, avec une première étape (perceptive) de captation de l'information et une seconde étape (cognitive) d'activation de différentes fonctions telles que langage, mémoire ou pensée ; ces deux modules correspondant à des systèmes et des fonctions cérébrales distinctes et indépendantes. Cela étant, cette divergence tend progressivement à disparaître notamment grâce aux travaux de Barsalou qui développe le concept de cognition « incarnée » (*grounded cognition*) en ancrant la cognition dans les états perceptifs et en rejetant la notion de représentation amodale (c'est-à-dire, non spécifique à une modalité sensorielle : vision, audition, etc. ...) [Barsalou, 1999] [Houix, 2003].

De manière consensuelle, on peut donc considérer la perception comme une collecte d'informations provenant du monde extérieur, une interprétation de ces données d'un point de vue mental et une mise en œuvre de processus liés à la connaissance (mémoire, apprentissage, réflexion, jugement, etc. ...). De plus, dans ce paradigme, « nos processus cognitifs, en fin de compte, sont guidés aussi bien par un traitement ascendant que par un traitement descendant » [Matlin, 2001] :

- le processus **ascendant** (*bottom-up*) part des informations issues du stimulus dont il traite les attributs en termes de données physiques à des fins de reconnaissance et d'identification ;
- le processus **descendant** (*top-down*) part des connaissances acquises (catégories d'objets, schémas d'actions, etc) pour orienter les mécanismes de perception en effectuant ainsi une sorte de rétroaction de l'expérience sensible (expérience passée, vécue) sur les données sensorielles ou leur interprétation.

Ce double mouvement pouvant être illustré, dans le domaine de l'audition, par la citation de Howard et Ballas : « *it is not only what we hear that tells us what we know; what we know tells us what we hear* » [Howard et al., 1980].

²³ <http://www.cnrtl.fr>

²⁴ <http://www.larousse.fr>

Dans le même ordre d'idées, pour Matlin, la perception est l'un des processus mentaux – avec, entre autres, la mémoire, le langage, l'apprentissage, la résolution de problèmes, le raisonnement ou la prise de décision – qui constitue le champ de la cognition, « activité mentale [qui] comprend l'acquisition, le stockage, la transformation et l'utilisation des connaissances » [Matlin, 2001].

La cognition – ou psychologie cognitive, toujours selon Matlin – s'inscrit par ailleurs, plus largement, dans le cadre des **sciences cognitives** regroupant, quant à elles, un ensemble de disciplines scientifiques qui s'intéressent au fonctionnement de la pensée en général, incluant la psychologie, la philosophie, la linguistique, l'anthropologie, les neurosciences et même l'intelligence artificielle lorsqu'il s'agit d'étudier des systèmes complexes, tels que le cerveau ou l'ordinateur, selon des processus issus de modèles de traitement de l'information [Eysenck, 2000].

Cela étant, pour revenir à la **psychologie cognitive**, elle est parfois assimilée à l'une des grandes approches théoriques de la psychologie – l'approche cognitive – qui se distingue d'autres courants comme le behaviorisme, la psychologie de la forme (gestaltisme) ou l'approche psychophysique ; tous ces courants s'inscrivant, en outre, dans le cadre de la **psychologie expérimentale**, démarche basée sur une méthodologie scientifique qui consiste à rechercher des preuves empiriques (expérimentales) dans un contexte de recherche en psychologie [Guastavino, 2003].

Le **behaviorisme** (ou comportementalisme) se base sur l'étude du comportement considéré comme manifestation objective de l'activité globale d'un individu (ou d'un animal) qu'il considère donc comme une réponse permettant d'étudier les mécanismes psychiques sous-jacents. Le paradigme associé est celui de l'association stimulus-réponse qui permet de décomposer une forme complexe de comportement en unités élémentaires mesurables (concept associationniste) et considère l'individu comme un système "entrée-sortie" ou comme une "boîte noire" à laquelle l'expérimentateur n'a pas accès. De fait le behaviorisme a principalement eu deux impacts sur la psychologie :

- d'une part minimiser l'étude des processus mentaux – notamment ceux relatifs à la pensée – au profit de l'étude du comportement objectif des individus dans leur environnement par des méthodes expérimentales ;
- d'autre part, grâce à cette même démarche méthodologique, poser les fondements de la psychologie expérimentale en formalisant l'observation des réactions d'un individu à l'exposition d'une stimulation.

On distingue donc généralement le behaviorisme théorique qui ne fait aucune hypothèse sur le traitement cognitif des stimuli (contrairement au cognitivisme – cf. ci-dessous) et le behaviorisme méthodologique qui est implicitement mis en œuvre dans toute approche de psychologie expérimentale ([Le Ny, 1989], cité dans [Guastavino, 2003]).

Ce courant, formalisé par J. B. Watson (1913), s'est largement développé aux Etats-Unis au cours de la première moitié du XX^{ème} siècle et a fini par être largement critiqué, notamment sur les points suivants :

- ne pas prendre en compte les états intérieurs des individus n'est valable que dans des cas de conduite automatique de type réflexes, mais ne peut pas s'étendre aux activités ou situations du quotidien ;
- l'approche associationniste (ou moléculaire) ne permet probablement pas d'expliquer le processus global.

La **psychologie de la forme** (gestaltisme – issue de la théorie de la Gestalt) se pose en alternative au behaviorisme en se développant à peu près à la même période (fin du XIX^{ème} siècle) mais essentiellement en Europe (initialement, Autriche et Allemagne). Il rejette globalement le paradigme stimulus-réponse et les approches associationnistes. Pour la Gestalt, le système perceptif met en place des processus de traitement de l'information structurés et auto-organisés qui permettent d'interpréter les objets ou les situations issus d'une observation. De plus, cette approche est basée sur le postulat que la totalité est supérieure à la somme des parties, cette totalité étant la "forme" – littéralement définie comme « autre chose ou plus que la somme de ses parties » ([Ehrenfels, 1890], cité dans [Guastavino, 2003]) – qui émerge d'un "fond". La théorie de la Gestalt repose donc sur deux principes :

- la distinction figure / fond qui correspond à l'organisation du système perceptif en une figure qui porte toute la signification, mise en avant d'un fond, moins bien structuré et moins précis (voir figure 93, vase de Rubin);
- le principe de regroupement qui postule que les informations perçues sont mises en forme grâce à une organisation des relations entre elles, existante de manière intrinsèque à l'individu (voir figure 93, image du dalmatien).



Figure 93. Théorie de la Gestalt : vase de Rubin (gauche) – image du dalmatien (droite)

Par ailleurs, ces principes d'organisation sont régis par des règles, appelées lois de la forme, dont les énoncés sont les suivants :

- loi de proximité : les éléments proches les uns des autres constituent une forme ;
- loi de similitude : les éléments ayant des caractéristiques communes (couleur, forme, taille, etc.) sont perçus comme un ensemble ;
- loi de délimitation : tendance à combler les vides pour percevoir des formes ou séquences d'action complètes ;
- loi de continuité : les lignes forment un ensemble prolongé dans le temps et dans l'espace

Cette dernière règle est notamment à mettre en relation avec l'approche de la Gestalt appliquée à la modalité auditive avec des travaux comme ceux de Bregman sur la ségrégation de l'information sonore en flux auditifs [Bregman, 1994] ou plus largement ceux de Deutsch sur la psychologie de la musique [Deutsch, 1982].

L'approche **psychophysique** cherche à établir des relations entre une grandeur physique (par exemple, la température) et la perception que l'on en a. Elle se développe à partir du milieu du XIX^{ème} siècle, sous l'impulsion de Gustav Theodor Fechner qui réalise une première application de méthodes scientifiques dans le champ de la psychologie. L'intention de Fechner était de « mesurer des états mentaux, des sensations, provoqués par des éléments physiques déclenchés intentionnellement par

l'expérimentateur et dérivés des réponses manifestes des sujets, réponses obtenues selon des modèles spécifiés ». Pour l'essentiel, la psychophysique étudie donc les relations quantitatives entre des événements physiques mesurables et des réponses évoquées selon « une règle expérimentale avérée » [Bonnet, 1986].

Le cadre de la psychophysique est donc l'étude du système perceptif qui reçoit, en entrée, une stimulation et délivre, en sortie, une sensation ; son objectif étant la modélisation des lois régissant les relations entre stimuli et réponse (par exemple, la loi de Stevens qui relie une intensité subjective à la mesure physique de cette même intensité par une fonction de puissance). Pour cela, la psychophysique part d'une hypothèse sur l'existence d'un processus décisionnel entre sensation (issue d'une stimulation) et réponse (voir figure 94) ; elle rentre donc globalement dans le cadre de la théorie de la détection du signal (TDS).

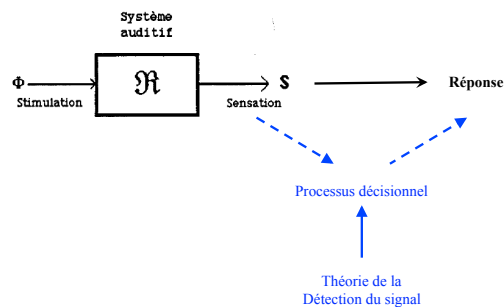


Figure 94. Hypothèse expérimentale de la psychophysique

Pour la modalité auditive, la discipline qui s'intéresse plus particulièrement aux relations entre signaux sonores et sensations auditives induites est la **psychoacoustique** qui inclue également la science de l'audition, ou la physiologie (mécanismes du système auditif).

Par ailleurs, la relation à la TDS motive la mise en œuvre de méthodes expérimentales spécifiques à la notion de détection (ou discrimination) de seuil, en premier lieu de nature unidimensionnelle, comme par exemple dans le cas de l'étude de l'intensité sonore perçue (sonie) :

- méthode des limites : production d'une réponse binaire (oui/non) à la présentation d'un stimulus d'intensité croissante – et décroissante ; c'est le paradigme utilisé notamment dans les procédures d'audiogramme (type Békésy) ;
- méthode d'ajustement : égalisation du niveau du stimulus étudié en fonction d'une référence ;
- estimation de grandeur : production d'une valeur proportionnelle à l'intensité perçue d'un stimulus ;
- estimation de rapport : production d'une valeur proportionnelle à la différence d'intensité perçue entre 2 stimuli ;
- méthode adaptative : production d'un seuil différentiel – ou plus petite valeur perçue (JND – *Just Noticeable Difference*) – entre 2 stimuli, au moyen d'une procédure itérative alternant augmentation/diminution d'une petite quantité de la grandeur mesurée.

A noter que ces mesures peuvent être réalisées au moyen de méthodes multidimensionnelles, comme par exemple dans le cas de l'étude de la grandeur physique du timbre (voir § III.2.2 pour des précisions sur cette notion) – on se place alors plutôt dans un paradigme de discrimination. Dans ce cas, une méthode de jugement de dissemblance permet de révéler les dimensions perceptives qui sous-tendent la perception des différences de timbre entre les sons d'un corpus puis de les

relier à des paramètres acoustiques des stimuli (voir § III.3.1 pour des détails sur cette méthode).

Cela étant, les principales critiques faites à la psychophysique, et donc à la psychoacoustique, se révèlent être finalement de trois ordres : i/ la difficulté d'identification des facteurs expérimentaux, c'est-à-dire des paramètres variables pertinents ; ii/ l'étude de stimuli complexes, chargés de sens (comme dans le cas de certains sons environnementaux) qui nécessite la prise en compte d'attributs liés à l'individu (sujet), d'une autre nature que ceux liés au stimulus mesuré (objet physique) ; iii/ la validité des tests en laboratoire, à l'opposé d'une situation d'écoute en condition naturelle par rapport à l'activité du sujet.

L'approche **cognitive** peut globalement être considérée – et ce malgré des points de vue différents – comme une approche exploitant l'analogie entre le cerveau et l'ordinateur [Eysenck, 2000]. Elle se développe depuis la seconde moitié du XX^{ème} siècle à partir notamment de l'ouvrage fondateur de Neisser « Cognitive Psychology » [Neisser, 1967]. L'un des facteurs de développement de la psychologie cognitive a été le paradigme du **traitement de l'information** à la base du développement de l'informatique et des sciences de la communication. Ce paradigme – et donc l'approche cognitive dans son ensemble – repose sur deux hypothèses essentielles vis-à-vis des processus mentaux [Matlin, 2001] :

- ils peuvent être appréhendés en les comparant aux opérations d'un ordinateur,
- ils peuvent être interprétés comme une progression d'informations à l'intérieur d'un système, par étapes, de la stimulation à la réponse ; comme l'explique l'illustration extraite de l'ouvrage de McAdams et Bigand] (figure 95) traitant de la psychologie cognitive de l'audition [McAdams et al., 1994].

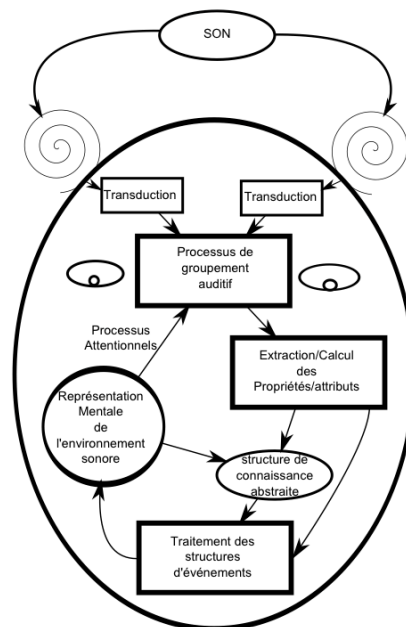


Figure 95. Principales étapes du processus de traitement de l'information (selon McAdams et Bigand [McAdams et al., 1994])

Néanmoins, dans l'histoire de l'évolution des sciences cognitives, le modèle du traitement de l'information, à la base du cognitivisme, est progressivement mis en concurrence avec d'autres approches, notamment l'approche écologique, le connexionnisme (intégrant le cerveau) ou la cognition située (intégrant le mouvement) :

- le **cognitivism** est donc l'approche basée sur le modèle du traitement de l'information qui compare l'esprit humain à l'ordinateur le plus complexe qui soit et qui tente d'en comprendre le fonctionnement. Autrement dit, l'approche cognitive (ou approche computationnelle) part de l'ordinateur comme métaphore de l'esprit humain et postule que les éléments extérieurs captés par le système sensoriel sont traduits en représentations internes symboliques ; ces représentations étant ensuite manipulées selon des règles prédéterminées pour fournir les réponses (ou sorties) adaptées. Au cours de ce processus, le système n'interagit, a priori, qu'avec la forme des symboles (leurs propriétés physiques) et non leur sens.
Cette approche peut également s'interpréter comme une forme de **perception indirecte** dans la mesure où elle implique des traitements internes qui permettent de transformer les entrées sensorielles – par définition incomplètes – et concevoir les représentations mentales qui donnent la conscience du monde externe.
- l'approche **écologique** se réfère, quant à elle, à une **perception directe**. Elle est associée à la théorie écologique développée par Gibson [Gibson, 1978 ; Gibson, 1979] et qui stipule une spécification précise des objets dans la stimulation sensorielle, disponible pour l'organisme. Dans cette approche, la perception est active, liée au mouvement et capte les variations et invariants structuraux d'un flux reçu par les organes sensoriels. L'idée de représentation mentale interne n'existe plus puisque les informations sont déjà contenues dans la structure de la sensation (l'énergie lumineuse ou le timbre sonore, par exemple) ; les mécanismes de perception étant alors plutôt du type centré sur l'objet et ses données intrinsèques (*data-driven*). Il n'est donc plus utile de traiter l'information provenant des capteurs sensoriels pour construire une représentation du monde.
De ce point de vue, découle également la notion d'affordance définie par Gibson lui-même comme « l'information potentiellement utile et pertinente pour l'action d'un individu » [Gibson, 1977] ; cette notion ayant ensuite été reprise, dans le domaine de l'ergonomie, pour signifier le caractère intuitif de l'utilisation d'un objet [Norman, 1999].
- Le **connexionnisme** apparaît réellement au début des années 80 en réaction au cognitivism – mais est également issu de la réflexion cybernétique des années 50/60. Il prend en compte le cerveau et tente d'expliquer la cognition grâce au modèle des réseaux de neurones, sur la base d'une métaphore d'un réseau complexe d'unités interconnectées entre elles. A partir de ce modèle, les mécanismes perceptifs s'interprètent par l'exploration des ressources de ce réseau, plutôt sur le schéma d'un processus descendant (*top-down*). L'une des premières implémentations de cette approche a été effectuée par F. Rosenblatt : le « *Perceptron* », système de reconnaissance basé sur le changement de connectivité entre des composants semblables à des neurones (neurones artificiels) [Rosenblatt, 1958].
Le connexionnisme repose également sur le formalisme d'agents (ou micro-agents) liés par une architecture distribuée, coopérative et auto-organisée. De plus, le modèle algorithmique séquentiel, propre au cognitivism, est abandonnée au profit d'une conception du traitement de l'information en parallèle. L'un des grands postulats du connexionnisme est également la notion d'apprentissage (ou d'entraînement) : en effet, cette approche fait l'hypothèse que la structure du réseau

se modifie en fonction de l'expérience – à l'image de certains mécanismes cérébraux. La connectivité du système devient alors dépendant de son histoire, c'est-à-dire de ses états antérieurs qui renforcent ou affaiblissent les liens entre les unités [Upop, 2011].

- La **cognition située** (ou cognition incarnée – *embodied cognition*) – ou approche par systèmes dynamiques embarqués – est l'un des autres grands paradigmes de la psychologie cognitive. A la différence du cognitivisme et du connexionnisme, elle prend en compte le fait que l'organisme qui perçoit le monde se trouve dans une situation particulière qui concerne la totalité du système formé par : le cerveau, le corps et le monde. Elle débouche sur le concept de l'**énaction** [Varela et al., 1992] qui postule fondamentalement que l'environnement extérieur est perçu au travers des interactions que l'individu entretient avec lui et qui se base notamment sur les points suivants:
 - le couplage sensori-moteur entre l'organisme et son environnement permet de faire émerger la signification ;
 - la cognition s'incarne dans un corps doté d'un système nerveux dynamique capable de générer des patterns d'activité ;
 - l'expérience vécue est un phénomène central aux facultés cognitives d'un individu.

Malgré la relative divergence de ces différentes approches, il est important de noter le rôle crucial que joue la **représentation cognitive** (ou mentale) du monde autour du sujet "percevant" ; le système cognitif humain reposant plus ou moins directement – selon les théories – sur la mise en œuvre et la manipulation de ces objets mentaux. Pour autant, la notion de représentation reste difficile à définir et peut varier suivant la discipline considérée (psychologie, philosophie, etc.). Michel Denis en fournit une définition appropriée dans le « vocabulaire des sciences cognitives » [Houdé et al., 1998] : « la représentation mentale peut être vue comme une entité interne, le correspondant cognitif individuel des réalités externes expérimentées par un sujet ». En outre, ces représentations mentales en tant qu'éléments probables de médiation des mécanismes perceptifs – éléments intermédiaires entre les données captées par les organes sensoriels et l'information traitée par le cerveau – possèdent principalement deux propriétés ; elles sont :

- le plus souvent, multimodales ; pour preuve, les études sur l'interaction vision/audition (p. ex., [Viollon, 2000]) ou audition/vibration (p. ex., [Frere et al., 2014]) ;
- partiellement subjectives, c'est-à-dire spécifiques à un individu parce qu'issues d'expériences sensorielles propres, mais également partagées, au moins au sein d'une même culture ou d'un même groupe social.

Cela étant, quoique (individuellement ou collectivement) subjectives – donc par essence, non directement observables – les représentations mentales sont malgré tout accessibles par des expériences d'objectivation. Les méthodologies employées peuvent être classiques (par exemple, tâche de catégorisation libre – voir § III.3.2) ou bien utiliser des supports tels que le langage (description sémantique) [Dubois, 1997], le dessin (croquis, esquisses) [Maffiolo, 1999], les imitations vocales [Lemaitre et al., 2013], la production de gestes [Caramiaux et al., 2011] ou bien encore par un procédé actif de synthèse par corpus [Lafay et al., 2014].

D'un point de vue scientifique, ce modèle de représentations internes suscite un intérêt quant à la manière dont elles sont structurées. Une première question se pose alors d'emblée sur la nature même du monde : est-il « continu ou rempli d'objets » [Faure, 2000] ? Il semble que les différentes théories de la perception répondent chacune à leur manière à cette interrogation en donnant une place plus ou moins grande à la notion de mémoire (notamment la mémoire à long terme – MLT) et à la façon dont les connaissances y sont organisées. Il s'avère cependant que pour la plupart des modèles, la MLT est structurée en catégories et relations inter-catégorielles [Eysenck, 2000].

Quoiqu'il en soit, les bases théoriques de la notion de **catégories** – et de catégorisation – sont intéressantes à poser dans la mesure où elles s'intègrent dans un processus de discrétisation du monde qui intervient dans l'interprétation de nombreux mécanismes perceptifs (voir [Houix, 2003] pour un développement détaillé de la question). Pour cela, partons d'abord de la définition de Sloutsky : « *Categories are equivalence classes of different (i.e. discriminable) entities and categorization is the ability to form such categories and treat discriminable entities as members of an equivalence class.* » ([Sloutsky, 2003], cité dans [Houix, 2012]) et notons ensuite qu'il existe principalement trois grandes théories comptant, d'une part, la théorie classique (ou *rule-based*) basée sur un ensemble de règles, et d'autre part, la théorie prototypique et la théorie des exemplaires, basées toutes les deux sur le principe de similarité :

- la **théorie classique** repose sur un fondement logique et utilise un ensemble de conditions nécessaires et suffisantes pour déterminer si un objet appartient ou non à une catégorie. Ainsi, à partir du moment où ils satisfont à ces conditions, les objets d'une catégorie possèdent exactement le même statut, de manière indépendante du contexte, et sont déterminés par un concept commun (un canari et un pingouin font partie de la catégorie "oiseaux").

L'exigence des conditions nécessaires et suffisantes induit, *de facto*, des limitations qui excluent notamment la notion de typicalité – même si certains développements de cette théorie tentent de faire la distinction entre des attributs de définition (*defining features*) et des attributs de caractère (*characteristic features*).

En revanche, ces mêmes conditions fournissent un schéma de catégorisation très bien défini dont s'inspirent, entre autres, certains réseaux sémantiques ou modèles de classification.

- la **théorie prototypique** a été initialement formalisée par Eleanor Rosch [Rosch, 1978] et repose sur le principe de similarité (ou « air de famille ») à partir de propriétés partagées par les membres d'une catégorie ; l'ensemble des propriétés n'étant pas nécessairement partagées par l'ensemble des membres de la catégorie, ce qui, par ailleurs, confère aux catégories un caractère non exclusif.

Elle repose en outre sur la notion de prototype, élément central d'une catégorie qui représente une abstraction (stéréotype) d'une majorité de propriétés partagées par une majorité d'éléments ou « qui condense les propriétés les plus saillantes d'une catégorie ». Le prototype est aussi « l'élément moyen » qui, par contraste, permet de distinguer deux classes situées au même niveau d'abstraction.

Cette théorie trouve son fondement dans le postulat de deux principes psychologiques relatifs à la catégorisation :

- le principe d'économie cognitive consistant à extraire l'information avec un "cout" cognitif minimum ;

- le principe de la structure du monde perçu qui « n'est pas un ensemble totalement non structuré d'attributs de co-occurrences équiprobables » [Rosch, 1978].

Elle postule également l'existence d'une organisation inter-catégorielle se déclinant sur trois niveaux hiérarchisés : le niveau superordonné (p.ex., "animal"), le niveau intermédiaire ou basique (p. ex., "oiseau") et le niveau subordonné (p. ex., "canari").

- la **théorie des exemplaires** (ou de l'instance) qui relâche essentiellement les contraintes sur les critères d'appartenance à une catégorie en permettant à chaque élément (exemplaire) d'une catégorie de posséder ses propres attributs. En première approximation, cette théorie peut être vue comme une approche prototypique à plusieurs prototypes.

Le modèle du contexte (*context model*) représente une version simplifiée de la théorie des exemplaires et postule notamment que tous les membres d'une catégorie sont stockés en mémoire ; les similarités entre les instances pouvant varier en fonction de phénomènes d'attention sélective, autrement dit les attributs étant pondérés par des poids attentionnels qui dépendent du contexte.

III.2 Perception sonore : timbre et identification des sources

Les approches théoriques présentées précédemment couvrent de très larges champs disciplinaires et permettent d'expliquer en grande partie les mécanismes de perception dans les différentes modalités sensorielles ; elles ont d'ailleurs été, pour certaines, initialement établies pour la vision – par exemple, le théorie de la Gestalt ou bien l'approche gibsonienne.

Dans le domaine du sonore, les études perceptives portent essentiellement sur la question du timbre, de l'identification et de la reconnaissance des sources mais aussi de la préférence ou de l'évaluation de la qualité sonore sur des critères de type hédonique (agréable, gênant, etc.), voire esthétique (la caractérisation du « beau son »).

De plus, les études en perception sonore manipulent généralement deux corpus distincts qu'il est rare de voir combiner :

- les **sons musicaux** naturellement issus des instruments de musique acoustiques ou synthétiques ;
- les **sons environnementaux**, par exclusion, autres que musicaux et respectant la définition donnée par Vanderveer : « (...) *any possible audible acoustic event which is caused by motions in the ordinary human environment* (...). » [Vanderveer, 1979] (voir § III.6.2.1 pour la citation complète)

Même si la définition de Vanderveer est suffisamment large pour y inclure aussi les sons musicaux, par ailleurs, même si un son musical peut revêtir un caractère environnemental – et parfois pas du tout musical au sens premier du terme (p. ex., une note de violon jouée 'col legno tratto') – et même si un son environnemental peut posséder un caractère musical (p. ex., un klaxon ou un tintement de verre), cette distinction reste néanmoins de rigueur tant la source de l'instrument de musique demeure spécifique et spécifiquement étudiée dans le monde de l'acoustique.

III.2.1 Introduction

La perception d'un signal sonore provenant d'une source peut s'effectuer de plusieurs manières et à plusieurs niveaux de signification.

Initiateur du courant dit de « musique concrète », Pierre Schaeffer a repensé le phénomène perceptif et a redéfini la notion d'écoute en formalisant quatre modes qui reflètent chacun une attitude perceptive différente [Schaeffer, 1966] :

- « OUIR », écouter le son sans se poser de question, un acte de perception élémentaire, brut et passif ;
- « ENTENDRE », pratiquer l'écoute réduite, c'est-à-dire, sélectionner dans le signal sonore des éléments permettant de qualifier le perçue ;
- « ECOUTER », rechercher dans les indices du son sa provenance possible, sa source, la causalité de l'événement qui l'a produit ;
- « COMPRENDRE », percevoir la signification, la valeur d'un son en le considérant comme le signe d'un langage.

Sur une même base, une terminologie semblable a été proposée par W. Gaver [Gaver, 1993] à propos des sons environnementaux tels que définis par Vanderveer (voir la définition au § III.2) :

- l'écoute musicale (*musical listening*) se focalise sur les propriétés du son,
- l'écoute quotidienne (*everyday listening*) se focalise sur la caractérisation de la source qui est à l'origine du son.

Le terme « musical » étant cependant potentiellement ambigu dans la mesure où, comme le font remarquer Tuuri et coll. dans une étude de 2007, il est a priori possible d'avoir une écoute musicale d'un son non-musical, et inversement, de se focaliser sur les sources d'un signal musical (en cherchant, par exemple, à séparer les instruments d'un orchestre) [Tuuri et al., 2007].

De son côté, M. Chion, dans un ouvrage traitant de la relation du son à l'image, définit trois modes d'écoute – ou attitudes de perception sonore – en faisant la synthèse des formalismes de Schaeffer et de Gaver et y adjoignant une troisième notion [Chion, 1994] :

- l'écoute réduite, reprise de la terminologie de Schaeffer,
- l'écoute causale, équivalente à l'écoute quotidienne de Gaver,
- l'écoute sémantique, relative au sens véhiculé par le son.

Depuis des travaux de recherche ont permis de démontrer partiellement cette structure, pour la perception des sons environnementaux. Dans le cadre du projet européen CLOSED²⁵ (Closing the Loop of Sound Evaluation and Design), Lemaitre, Houix, Misdariis et Susini ont constitué puis étudié un corpus de sons issus de l'environnement quotidien de la cuisine et ont mis à jour expérimentalement, deux niveaux principaux d'écoute des sons, similaires à ceux définis par Chion :

- acoustique : écoute analytique des propriétés "abstraites" des sons,
- causal : écoute de la production des sons.

De plus, l'étude a également montré l'influence de deux facteurs sur ces modes d'écoute : i/ l'identifiabilité du son – moins le son est reconnaissable, plus la stratégie d'écoute se focalise sur ses propriétés acoustiques ; ii/ l'expertise du sujet – plus le sujet est novice, plus il développe une stratégie d'écoute causale [Lemaitre et al., 2010].

²⁵ <http://www.closed.ircam.fr>

A ce stade, et comme le fait remarquer Faure [Faure, 2000], il est par ailleurs intéressant de noter que ces deux principaux types d'écoute peuvent facilement renvoyer à deux théories de la perception (voir § III.1) : la théorie cognitive du traitement de l'information, démarche analytique de reconstruction d'une représentation à partir d'attributs pouvant être associée au niveau "acoustique" et l'approche écologique de Gibson, perception directe sans intermédiaire, pouvant être associée au niveau "causal".

III.2.2 Etude du timbre

La notion de timbre musical – et, par extension, timbre sonore – est un concept important dans les études sur la perception auditive. Il est communément défini – par défaut – comme un attribut perceptif n'étant ni la hauteur, ni la durée, ni l'intensité d'un son et permettant néanmoins de juger de la similarité entre deux sons (ANSI, 1960). Dans cette approche de définition par la négative, un complément donné par McAdams [McAdams, 2012], considère également que les paramètres d'espace (comme la localisation) et d'environnement (comme la réverbération) ne constituent pas, non plus, des indices de timbre.

Par ailleurs, toujours pour McAdams, il est maintenant admis que le timbre possède deux principales caractéristiques :

- un ensemble d'attributs perceptifs élémentaires, certains étant continus (attaque, brillance, nasalité, richesse, etc.), d'autres discrets (p. ex., le rebond caractéristique du sautereau de clavecin) ;
- un élément primaire de reconnaissance, identification et repérage (*tracking*) de la nature d'une source sonore, impliquée, de fait, dans un mécanisme psychologique de catégorisation.

Plus largement, le timbre est donc à la fois un percept de bas niveau complexe et puissant, un vecteur d'identité sonore mais aussi un élément de composition musical, notamment en termes de structuration et de groupement [Wessel, 1979] (voir [McAdams, 2012], pour un développement complet et réactualisé de la question).

De fait, de nombreuses études perceptives se sont intéressées à caractériser le timbre en le reliant à des paramètres acoustiques ou mathématiques du signal sonore. Elles ont montré, entre autres, qu'il s'agissait d'un percept multidimensionnel, c'est-à-dire résultant de la covariation de plusieurs paramètres physiques, aussi bien dans le domaine spectral que temporel ; elles ont donc, pour la plupart, abouti à la définition d'un **espace des timbres** par rapport à un corpus d'étude donné (voir [Hajda et al., 1997] pour une revue détaillée des différentes méthodologies).

Historiquement, Helmholtz est le premier scientifique à s'être intéressé au timbre des instruments de musique. Il a tenté de décrire les relations entre le contenu spectral des sons et leur qualité perçue (voir figure 96, interprétation des résultats d'Helmholtz par E. Leipp [Leipp, 1971]).

Par la suite, Lichte effectue une étude psychophysique de comparaison par paires, à partir de sons de synthèse et met en évidence trois attributs du son : la « brillance » (centre de gravité spectral), la « rugosité » (localisation des partiels de rangs élevés) et la « plénitude » (homogénéité du spectre) ([Lichte, 1941], cité dans [Faure, 2000]).

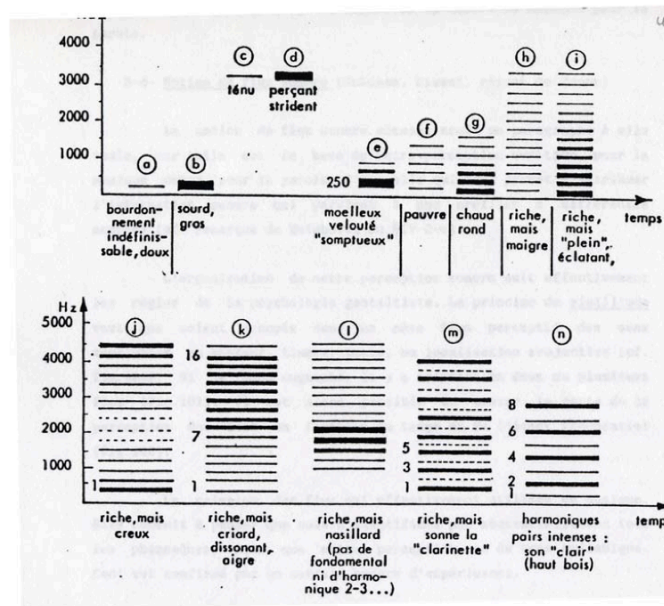


Figure 96. Résultat des travaux de Helmholtz sur le timbre interprétés par E. Leipp

De manière plus convergente, les premières méthodes expérimentales utilisées pour étudier le timbre ont exploité les techniques de différentiel sémantique (voir § III.3.3). Solomon, en 1958, a ainsi étudié un corpus de sons de sonar à partir d'une cinquantaine d'échelles sémantiques (« lourd-léger », « mat-brillant », etc.) et a abouti à une caractérisation verbale selon 7 dimensions : magnitude (taille volume, hauteur), esthétique, clarté, sécurité, relaxation, familiarité, autre (richesse, couleur, ...) [Solomon, 1958].

Presque vingt ans plus tard, Bismarck effectue une étude similaire sur des sons de synthèse à base de sons harmoniques et bruit blanc (synthèse de type source/filtre) qu'il fait évaluer sur une trentaine d'échelles sémantiques ; il en ressort deux dimensions explicites : la « brillance » (*sharpness*) et la « compacité » (*compacity*) [Bismarck, 1974].

Depuis les années 70, le paradigme généralement employé pour étudier les attributs du timbre est la méthode de comparaison par paires suivant un jugement de dissemblance – et l'analyse multidimensionnelle qui lui est associée. Cette approche s'affranchit des contraintes du langage et permet, a priori, de faire émerger les dimensions perceptives qui sous-tendent la perception de cette grandeur multidimensionnelle qu'est le timbre. De nombreuses études ont exploité cette technique soit sur des sons de synthèse [Miller et al., 1975 ; Caclin et al., 2005], des sons instrumentaux simulés ou resynthétisés [Wessel, 1979 ; Grey, 1977 ; Krumhansl, 1989 ; McAdams et al., 1995] ou bien encore sur des sons instrumentaux enregistrés [Iverson et al., 1993 ; Lakatos, 2000]. Parmi ces études de référence :

- Grey [Grey, 1977] étudie un corpus de 16 sons resynthétisés d'instruments à vent et à cordes réels, joués à la même note (Mib3) : hautbois (2), cor anglais, basson, clarinette (2), saxophone (3), flute, trompette, cor, trombone, violoncelle (3). Il met en évidence un espace à 3 dimensions (voir figure 97) expliqué par :
 - la distribution d'énergie dans le spectre, qualifiée de « brillance » par Wessel [Wessel, 1979] ;
 - la densité et l'inharmonicité des transitoires d'attaque, qualifiée de « qualité d'attaque » par McAdams [McAdams et al., 1994]

- une combinaison du synchronisme des partiels pendant l'attaque et de la fluctuation de l'enveloppe spectrale, qualifiée de « flux spectral » par Krumhansl [Krumhansl, 1989] ;

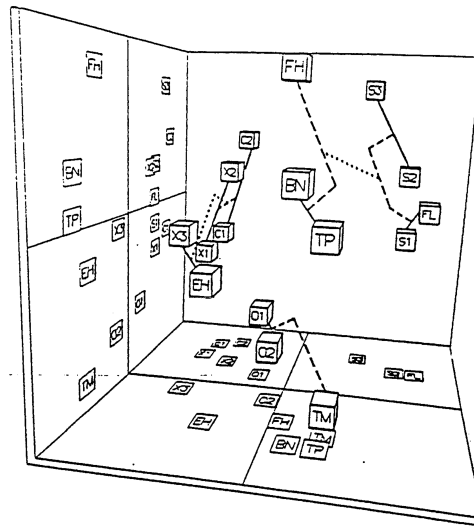


Figure 97. Espace des timbres d'après Grey [Grey, 1977]

Instruments : hautbois (O1, O2), clarinette (C1, C2), saxophone (X1, X2, X3), cor anglais (EH), cor (FH), cordes (S1, S2, S3), trompette (TP), trombone (TM), flute (FL), basson (BN)

- McAdams et coll. [McAdams et al., 1995] étudient un corpus de 18 sons de synthèse, incluant des hybrides (synthèse croisée de 2 sources) : vibraphone, harpe, oboleste (hautbois/celesta), guitare, piano, vibron (vibraphone/trombone), clavecin (*harpsichord*), obochord (hautbois/clavecin), guitarnet (guitare/clarinette), clarinette, striano (corde/piano), trombone, cor d'harmonie (*French horn*), corde frottée (*bowed string*), trumpar (trompette/guitare), cor anglais (*English horn*), basson, trompette. Il met en évidence également un espace à 3 dimensions expliqué par : le temps d'attaque, le centre de gravité spectral et le flux spectral (figure 98).

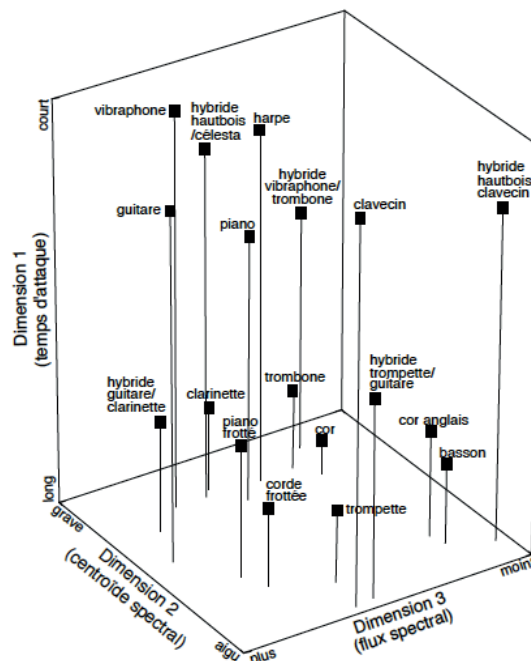


Figure 98. Espace des timbres d'après McAdams et coll. [McAdams et al., 1995]

En résumé, la structure de l'espace des timbres semble converger vers trois dimensions, l'interprétation des deux premières faisant relativement consensus : brillance et qualité de l'attaque. En revanche, l'interprétation de la troisième dimension semble être davantage sujette à variation mais également sensible aux facteurs expérimentaux (corpus de sons, groupes de sujets, etc.) : « flux spectral » pour McAdams et coll. [McAdams et al., 1995] ou bien « structure fine du spectre » (irrégularité par rapport à une enveloppe globale) pour Krimphoff et coll. [Krimphoff et al., 1994].

Il est également important de noter que cette définition de l'espace des timbres se restreint généralement à une classe particulière de sons (sons musicaux entretenus ou percussifs, etc.) et nécessiterait d'être généralisée à un ensemble plus large, ce qui pose notamment des problèmes méthodologiques importants (taille et hétérogénéité du corpus pour le jugement de dissemblance). Dans cette optique, nous avons récemment proposé une approche de "méta-description" du timbre des sons environnementaux qui permet de dégager des dimensions communes – et d'autres spécifiques – à différents types de sons [Misdariis et al., 2010b] (voir § III.6.2 pour la teneur de cette étude).

Par ailleurs, comme mentionné précédemment, l'objectif de ces études sur le timbre est non seulement de faire émerger ses dimensions perceptives sous-jacentes mais également de déterminer des corrélats acoustiques permettant de qualifier ces percepts élémentaires tels que brillance ou attaque. Le principe consiste à trouver un descripteur acoustique du signal (*audio feature*) dont la variation est proportionnelle à la position des sons – et donc des signaux – sur une dimension de l'espace, de sorte à modéliser la perception suivant cet axe par une grandeur acoustique (ou mathématique) calculée sur le son ; cette modélisation pouvant, le cas échéant, être le résultat d'une combinaison (linéaire ou non) de plusieurs descripteurs. La plupart des études de référence mentionnées ci-dessus comportent donc une partie "computationnelle" de recherche des meilleurs corrélats acoustiques ; parmi ceux-ci, certains sont largement standardisés :

- la brillance est corrélée avec le centre de gravité spectral (CGS) dont la formulation élémentaire est définie comme suit :

$$CGS = \frac{\sum_{k=1}^N k \cdot A_k}{\sum_{k=1}^N A_k} \quad (\text{Eq. 30})$$

où les $\{A_k\}$ sont les amplitudes des composantes spectrales (généralement obtenues par transformée de Fourier du signal).

A noter que des formulations plus élaborées intégrant des concepts de psychoacoustiques – comme, par exemple, l'échelle de Bark – peuvent aboutir à un centre de gravité spectral perceptif, ou « acuité » (*sharpness*) tel que défini par Zwicker et Fastl [Zwicker et al., 1990].

- la qualité de l'attaque est corrélée avec le logarithme du temps d'attaque (LTM), défini comme suit :

$$LTM = \log_{10}(t_{\max} - t_{\text{seuil}}) \quad (\text{Eq. 31})$$

où t_{\max} et t_{seuil} sont respectivement les temps correspondant au maximum d'amplitude (100%) et au seuil de perception (10%) de l'enveloppe temporelle du signal.

- la structure fine du spectre est corrélée avec l'irrégularité – ou la déviation – spectrale (IRR), définie comme suit :

$$IRR = \log_{10} \left(\sum_{k=2}^{N-1} \left| 20 \log_{10}(A_k) - \frac{20 \log_{10}(A_{k-1}) + 20 \log_{10}(A_k) + 20 \log_{10}(A_{k+1})}{3} \right| \right) \quad (\text{Eq. 32})$$

- la fluctuation de l'enveloppe spectrale est corrélée avec la variation spectrale – ou flux spectral – dont la formulation élémentaire est définie comme suit (corrélation entre 2 spectres instantanés) :

$$VAR = \frac{1}{M} \sum_{p=2}^M \left[\frac{\sum_{k=1}^N A_k(p-1) \cdot A_k(p)}{\sqrt{\sum_{k=1}^N |A_k(p-1)|^2} \cdot \sqrt{\sum_{k=1}^N |A_k(p)|^2}} \right] \quad (\text{Eq. 33})$$

où, p est le pas temporel de fenêtrage du signal sur lequel sont calculés les coefficients de Fourier $\{A_k\}$ et M le nombre total de fenêtres d'analyse.

A noter les fluctuations temporelles des propriétés spectrales peuvent également être décrites de manière plus réduite, par l'écart entre un CGS instantané – $CGS(p)$ – et un CGS moyen – CGS_m –, défini comme suit :

$$FLU = \frac{1}{M} \sum_{p=1}^M [CGS(p) - CGS_m]^2 \quad (\text{Eq. 34})$$

De nombreuses variations – ou nouvelles définitions – de ces descripteurs acoustiques ont émergé des différentes études menées sur le timbre depuis plusieurs dizaine d'années ; cette approche intéressant également de près la communauté MIR²⁶ (*Music Information Retrieval*) et les procédures d'indexation et de recherche par le contenu dans des bases de données sonores.

Ainsi, depuis peu, un effort de normalisation a été entrepris pour le développement de bibliothèques mathématiques recensant le calcul d'un grand nombre de descripteurs ; parmi elles :

- La MIR Toolbox²⁷ développée au Département de Musique de l'Université de Jyväskylä (Finlande) [Lartillot et al., 2007] ;
- La Timbre Toolbox, développée conjointement à l'Ircam et à l'Université de McGill (Montréal), et issue d'une série de travaux menés depuis plus de vingt ans dans le domaine. Elle contient un ensemble de 54 descripteurs modélisant au mieux les caractéristiques énergétiques, spectrales, temporelles, spectro-temporelles des signaux sonores. Elle a récemment fait l'objet d'une publication collective établissant la formulation stabilisée de ces descripteurs [Peeters et al., 2011].

III.2.3 Identification et classification des sources

En reprenant la définition du timbre de McAdams (§ III.2.2) – ou bien les notions d'écoute réduite et causale développées au § III.2.1, ou bien encore les différentes théories de la perception énoncées dans l'introduction (§ III.1) de ce chapitre –, il apparaît clairement qu'un autre niveau de traitement que celui des attributs acoustiques du signal peut être mis en jeu dans les processus de perception sonore.

²⁶ <http://www.ismir.net>

²⁷ <https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox>

En effet, lors de la production sonore d'une source (« événement sonore ») et la réception de cette information acoustique par le système auditif (« événement auditif »), de nombreuses études montrent – dans la lignée de l'approche écologique de Gibson – que ce que perçoit l'auditeur n'est pas nécessairement les propriétés de l'onde acoustique (« stimulus proximal ») mais peut aussi bien être directement les propriétés de l'événement lui-même (« stimulus distal ») [Gaver, 1988]. Dans certaines conditions, l'auditeur est ainsi à même d'identifier les matériaux et propriétés de la source, les actions mises en jeu pour produire le son, etc. Deux invariants sont alors principalement utilisés pour la reconnaissance et la classification des sources :

- les **invariants structuraux** qui décrivent les propriétés de structure (forme, volume, matériau, etc.) et indiquent la nature de la source – ils sont reliés à la perception de l'**objet** ;
- les **invariants transformationnels** qui décrivent les interactions opérées sur la source et indiquent la nature de la modification qu'elle subie – ils sont reliés à la perception de l'**action**.

Un exemple représentatif de cette distinction entre nature d'invariants est donnée par Houix [Houix, 2003] : la source "violon" sera toujours identifiable en tant que tel quelque soit sa provenance (Stradivarius ou violon d'étude), son registre (grave ou aigu) ou son mode de jeu (pincé ou frotté) ; de même, le mode d'action "pincé" sera reconnaissable quelque soit la source sur laquelle il s'opère : violon, guitare, clavecin, etc.

De nombreux travaux ont étudié spécifiquement ces deux types d'invariants à partir de différentes sources sonores dans le but général de relier les paramètres mécaniques des sources à des attributs perceptifs, se plaçant ainsi dans le champ thématique de la **psychomécanique** – concept développé entre autres par McAdams et Roussarie [McAdams et al., 2004]. Les objets et résultats (globaux) d'étude sont, parmi d'autres, les suivants (voir [Houix et al., 2007] pour une revue détaillée sur la question – dont sont extraites la plupart des références citées ci-après) :

- barres frappées en verre ou métal (définis par la densité et le module d'Young) : mauvaise discrimination des 2 classes de matériaux [Lufti et al., 1997].
- barres frappées en verre, métal, caoutchouc et bois (définis par l'amortissement) : meilleur résultat de discrimination du matériau dans une tâche d'identification à choix forcé, et bonne corrélation avec le paramètre mécanique d'amortissement [Klatzky et al., 2000].
- modèles physiques de barres et plaques frappées variant en densité, élasticité et différents types d'amortissement : bonne corrélation entre dimensions perceptives (révélées par un jugement de dissemblance) et dimensions mécaniques contrôlées [Roussarie, 1999].
- synthèse de bois, verre et métal (modélisés par un ensemble de fonctions sinusoïdes amorties) : définition d'un « espace de matériaux » expliqué par une combinaison linéaire de plusieurs descripteurs acoustiques (centre de gravité spectral, largeur de bande, rugosité et amortissement spectral) qui discrimine des zones distinctes dans l'espace [Aramaki et al., 2011].
- baguettes en bois de longueur variable jetées sur une surface rigide : bonne performance dans le jugement de la longueur relative des baguettes, mais absence de corrélats acoustiques significatifs [Carello et al., 1998].

- barres en métal et bois, de longueur variable : bonne reconnaissance du matériau mais mauvaise performance concernant la dimension des barres ; il s'avère, en outre, que ces deux facteurs ne sont pas indépendants [Gaver, 1988].
- barres en métal et bois, de section variable : bonne performance dans l'appariement des sections, en outre partiellement expliqué par des propriétés géométriques (rapport épaisseur/largeur) ou mécaniques (modes de flexion et torsion) [Lakatos et al., 1997 ; Houix, 2003].
- plaque fines de différentes formes (rectangle, triangle et cercle) et matériaux (métal, bois, plexiglas) frappées : bonne discrimination des formes et types de matériau, avec néanmoins une interaction entre ces 2 derniers facteurs (cercle/bois, triangle/métal, rectangle/plexiglas). L'étude ne dégage pas, par ailleurs, de corrélats acoustiques ou mécaniques [Kunkler-Peck et al., 2000].
- balle de taille variable rebondissant sur une plaque d'épaisseur variable : perception de la taille de la balle dépendante de l'épaisseur de la plaque [Grassi, 2005].
- plaques frappées en verre, métal, bois, plexiglass, de taille variable : identification groupée des familles verre/métal et bois/plexiglass et association significative entre tailles et matériaux : les plaques en verre/métal sont plus petites que celles en bois/plexiglass [Giordano et al., 2006].
- rebond et casse d'un objet en verre lâché : bonne performance de discrimination entre ces deux événements pouvant, de surcroît, être expliquée par des paramètres de type temporel (synchronicité d'impacts élémentaires) [Warren et al., 1984].
- frappe de casseroles de tailles différentes avec plusieurs maillets variant en dureté : les jugements perceptifs (échelle de dureté perçue) sont indépendants de la taille de l'objet frappé [Freed, 1990].
- remplissage d'un récipient par un liquide : mise en évidence de la capacité à remplir un verre d'eau, "en aveugle" en distinguant le niveau normal du niveau de trop-plein. L'étude conclue sur une affordance perceptive de l'action de remplissage et suggère la présence d'invariants transformationnels mais sans les expliciter [Cabe et al., 2000].
- applaudissement : pas de reconnaissance de la source et confusion liée à des stéréotypes (applaudissements aigus, rapides et de faible amplitude associés à des mains féminines) [Repp, 1987].

En résumé, ce panel d'études montre des résultats mitigés, parfois contradictoires ou stéréotypés, et avant tout relativement localisés, c'est-à-dire spécifiques à un type de source ou d'action. Ils militent, à nouveau, en faveur d'un double flux de traitement de l'information incluant un processus descendant (*top-down*) qui donne de l'influence au contexte d'écoute, ainsi qu'aux attentes et connaissances préalables de l'auditeur.

Par ailleurs, et d'un point de vue plus large, un autre point de départ à l'étude de l'identification des sons environnementaux peut être constitué de la taxonomie proposée par Gaver [Gaver, 1993]. Cette structure propose une organisation hiérarchique basée sur la physique des événements sonores et est globalement articulée selon 2 niveaux :

- un 1^{er} niveau impliquant des classes génériques de "matériaux", ou plutôt d'états de la matière : liquide, solide, gaz ;
- un 2nd niveau impliquant, pour chaque classe, différentes excitations – ou interactions – élémentaires de cette matière :
 - pour les solides : déformation (*deformation*) impact (*impact*), frottement, (*scraping*) et roulement (*rolling*) ;

- pour les liquides : goutte (*drip*), écoulement (*pour*), éclaboussement (*splash*) et ondulation (*ripples*) ;
- pour les gaz, explosion, souffle (*whoosh*) et vent (*wind*).

La structure complète de Gaver se complexifie ensuite par la combinaison de ces composantes élémentaires (voir figure 99). Elle définit ainsi, de manière intuitive mais relativement exhaustive, une topographie des sons environnementaux – voire un modèle d’organisation de notre monde sonore quotidien – qui fait depuis référence dans de nombreuses études traitant de la perception des sons environnementaux.

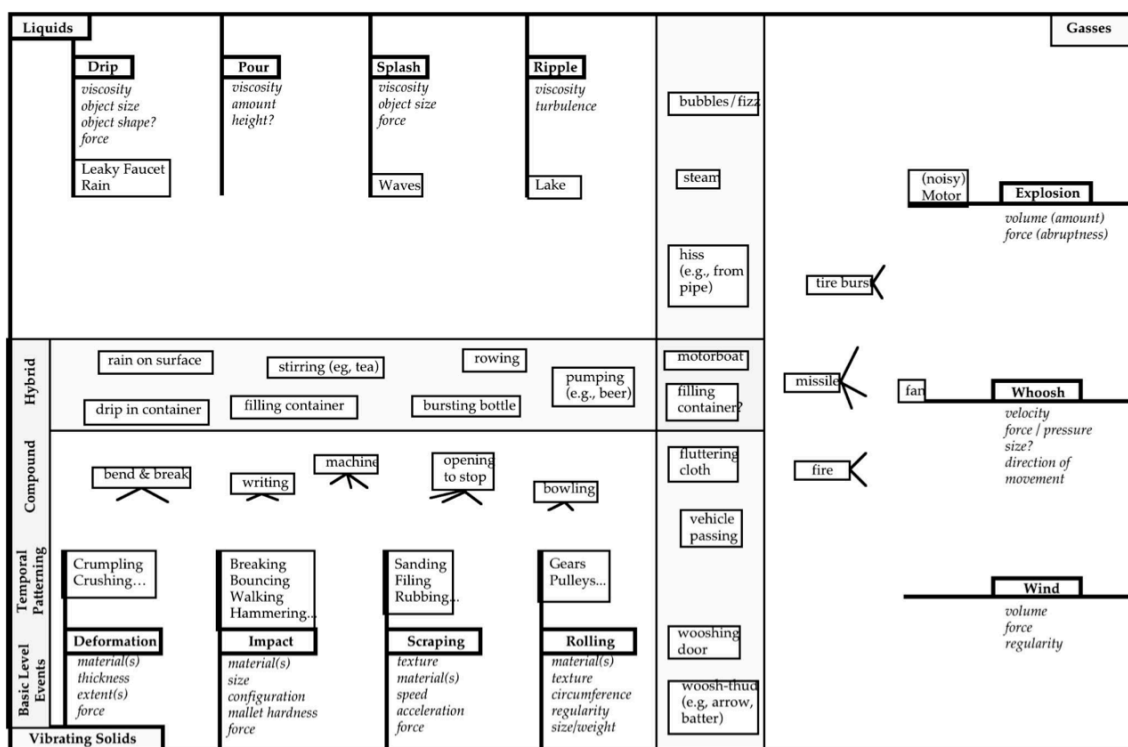


Figure 99. Taxonomie de Gaver pour la classification des sons environnementaux.

Les composantes élémentaires sont en gras et leurs attributs en italique ;
les combinaisons complexes se trouvent au centre de la figure – extrait de [Gaver, 1993]

Sur cette base, des travaux de recherche récents (dans le cadre du projet CLOSED²⁸) ont permis de valider expérimentalement cette approche phénoménologique en apportant de nouvelles données sur la structure perceptive des sons environnementaux [Houix et al., 2012], incluant notamment :

- la mise à jour d’une 4^{ème} classe d’"état de la matière" : « machines », en complément de la confirmation des classes « solide », « liquide » et « gaz » ;
- la description des interactions pour la classe « solide » et la mise en évidence d’une sous-structure hiérarchique distinguant les interactions discrètes et continues, et proposant une typologie à l’intérieur de chacune de ces sous-catégories (figure 100).

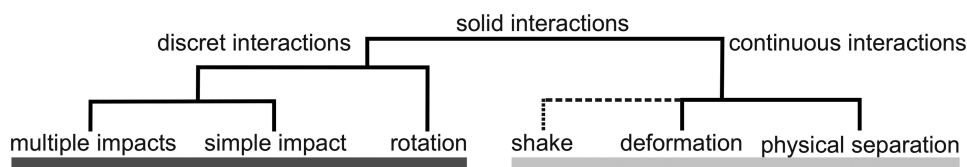


Figure 100. Structure des interactions de la classe des « solides » – extrait de [Houix, 2012]

²⁸ <http://www.closed.ircam.fr>

III.3 Paradigmes expérimentaux

Du point de vue de la psychologie expérimentale, l'étude de la perception sonore nécessite donc la mise en œuvre d'expériences d'écoute qui permettent de qualifier les sons et recueillir des données sur les sensations ou le ressenti subjectif du sujet. Cette section présente trois principaux paradigmes, couramment utilisés – et effectivement mis en œuvre dans la plupart des travaux expérimentaux relatifs à la perception présentés tout au long du document –, ainsi que les analyses de données associées : le test de dissemblance et l'analyse multidimensionnelle, le test de catégorisation et l'analyse de cluster et la méthode du différentiel sémantique et la production de portraits verbaux (voir [Susini et al., 2011] et [Giordano et al., 2013]) pour un développement complet de la question).

III.3.1 Test de dissemblance

Dans le test de dissemblance, les échantillons sonores sont présentés au sujet par paires et celui-ci doit se prononcer sur la dissemblance perceptive entre les deux sons d'une même paire. Pour cela, il possède un curseur qui se déplace sur une échelle continue allant de "très similaire" à "très dissemblable".

Les jugements sont codés sur une échelle continue allant de 0 (similarité maximum) à 1 (similarité minimum). Les paires, construites aléatoirement pour chaque sujet, ne sont généralement pas ordonnées ($\{a,b\}$ équivalent à $\{b,a\}$) et contiennent des éléments distincts (pas de présentation de $\{a,a\}$). De fait, un corpus d'expérience comportant N sons nécessitera $N(N-1)/2$ présentations de paires différentes.

Il est par ailleurs important de noter qu'au préalable, le sujet a écouté tous les échantillons dans un ordre aléatoire, afin de se faire une idée de la gamme de variation possible, et s'est familiarisé avec la tâche sur des paires d'entraînement.

Pour l'analyse, les données d'expérience sont arrangées sous forme matricielle $D(i,j)$ (demi-matrice triangulaire, sans diagonale) qui donne accès, pour chaque sujet, à la valeur δ_{ij} de la dissemblance entre les objets i et j ; cette matrice constitue la donnée d'entrée principale du logiciel d'analyse.

Puis des algorithmes de calcul basés sur le principe de l'**analyse multidimensionnelle** (MDS – *Multi-Dimensional Scaling*) permettent d'obtenir un espace géométrique où les objets sont positionnés en fonction de leurs dissemblances mutuelles. Pour construire cette représentation mathématique, la MDS suppose l'existence d'une relation monotone entre les valeurs expérimentales de dissemblance δ et les distances calculées d afin de conserver une cohérence dans la variation des deux ensembles. En d'autres termes, lorsque deux objets sont jugés éloignés, le modèle les écarte d'une distance en rapport avec cette perception et suit cette même logique avec deux objets jugés très proche.

D'un point de vue mathématique, il existe plusieurs modèles de calcul pour exprimer la distance entre deux points de l'espace d'analyse, les principaux sont :

- Le modèle euclidien classique (**MDSCAL**) [Shepard, 1962a,b ; Kruskal, 1964] où l'on suppose que les stimuli présentent des attributs communs et où la distance d_{ij} est calculée grâce aux projections x_{in} et x_{jn} des objets i et j sur les N dimensions de l'espace, suivant la formule :

$$d_{ij} = \sqrt{\sum_{n=1}^N (x_{in} - x_{jn})^2} \quad (\text{Eq. 35})$$

- Le modèle euclidien pondéré (**INDSCAL**) [Carroll, 1970] qui inclut dans son code de calcul des pondérations propres à un sujet et une dimension. En fait, par l'intermédiaire de coefficients de pondération w_{ns} , cette approche tient compte de l'importance perceptive que le sujet s attribue à la dimension n mais, par la même occasion, fait disparaître l'invariance rotationnelle qui existe dans le modèle précédent. On définit donc, pour le sujet s , la distance sous la forme :

$$d_{ijs} = \sqrt{\sum_{n=1}^N w_{ns} \cdot (x_{in} - x_{jn})^2} \quad (\text{Eq. 36})$$

- Le modèle euclidien étendu (**EXSCAL**) [Winsberg et al., 1989] où l'on suppose que les stimuli ont à la fois des dimensions communes et une spécificité propre S . La distance entre deux objets s'écrit alors (en tenant compte, ici encore, de l'effet pondérateur v_s induit par chaque sujet) :

$$d_{ijs} = \sqrt{\sum_{n=1}^N w_{ns} \cdot (x_{in} - x_{jn})^2 + v_s \cdot (S_i + S_j)} \quad (\text{Eq. 37})$$

l'algorithme EXSCAL pouvant, le cas échéant, être utilisé avec/sans pondération et avec/sans spécificité.

- Le modèle hybride **CLASCAL** [Winsberg et al., 1993], qui permet, quant à lui, de prendre à la fois en compte un modèle spatial comme ceux décrits plus haut, mais aussi d'évaluer l'existence de classes latentes de sujets en calculant des pondérations par classes t et non plus des pondérations par sujets isolés s , en faisant l'hypothèse qu'un sujet appartient à une seule classe.

En outre, l'indice statistique qui permet de déterminer le modèle spatial correspondant le mieux adapté aux données perceptives est le BIC proposé par Schwarz [Schwarz, 1978] ; il s'exprime de la manière suivante :

$$BIC_{\Omega} = -2 \log L_{\Omega} + v_{\Omega} \log \left(K \cdot \frac{J(J-1)}{2} \right) \quad (\text{Eq. 38})$$

où : K est le nombre de sujets
 J est le nombre de stimuli
 L_{Ω} est la fonction de vraisemblance (*likelihood*)
 v_{Ω} est le degré de liberté du modèle spatial Ω

Les conditions nécessaires à une approche de type jugement de dissemblance peuvent se résumer par : i/ égaliser les sons en intensité, durée et hauteur ; ii/ concevoir un corpus de sons homogène ; iii/ restreindre la taille du corpus ; iv/ le cas échéant, collecter des données verbales de la part des sujets [Susini et al., 2011].

III.3.2 Test de catégorisation

Il existe fondamentalement deux manières d'effectuer une catégorisation selon que celle-ci soit "libre" ou "orientée". Dans une approche de catégorisation orientée, le sujet est amené à assigner chaque stimulus à une classe prédéfinie ; cette définition pouvant être faite de manière sémantique – le plus courant – mais aussi au moyen d'une représentation graphique (p. ex., le symbole d'une morphologie) ou bien en référence à un autre son (prototype) – voir § III.6.2.3.2 pour exemple. Dans ce cas, l'analyse consiste

en un calcul de performance (% de bonnes réponses) ou bien de probabilité d'appartenance à une classe donnée et fournit des résultats de type matrice de confusion.

Dans une approche de catégorisation libre, les sons sont présentés au sujet en bloc et celui-ci doit effectuer des regroupements par similarité, en formant autant de classes qu'il le souhaite (le plus souvent, le nombre de classes n'est pas limité). L'interface de test est généralement graphique et numérique (écran d'ordinateur) et donne accès à l'ensemble des sons en termes d'écoute et de position ; le sujet pouvant ainsi effectuer ses regroupements dans l'espace de la fenêtre, définir ses classes et, le cas échéant, les spécifier sémantiquement ou en désignant un prototype pour chacune d'elles (voir, p. ex., l'environnement TCL/LABx²⁹ [Gaillard, 2009]). Cette approche méthodologique permet ainsi de traiter des corpus plus grands et plus variés que dans le cas du test de dissemblance.

L'analyse des similarités est alors menée en termes de traits communs plutôt qu'en termes de dimensions continues grâce à une **analyse de cluster**. Une telle conception aboutit à représenter les relations de proximité entre des objets non plus par des distances les séparant dans un espace géométrique mais en déterminant des groupes de stimuli (clusters) qui partagent les mêmes traits. Cela revient soit à diviser l'ensemble initial d'objets en plusieurs sous-ensembles relativement homogènes – éventuellement hiérarchiquement organisés – à l'intérieur desquels ceux-ci sont perçus comme similaires, soit à déterminer la meilleure structure en arbre dans laquelle la proximité correspond à la hauteur ou longueur du lien joignant deux objets. L'interprétation de telles structures consiste ensuite à identifier les traits que partagent les objets appartenant à un même sous-groupe, c'est-à-dire déterminer les traits ou propriétés d'un domaine de stimuli donné.

Pour cela, deux principales approches, correspondant à des algorithmes de regroupement différents, peuvent être utilisées [Barthélemy et al., 1988] :

- les **clusters additifs** correspondent à une organisation non hiérarchique des clusters mais dont les sous-groupes d'objets peuvent présenter des recouvrements partiels [Arabie et al., 1973 ; Shepard et al., 1979]. L'hypothèse de base de cette méthode est que tous les objets ont des propriétés auxquelles sont associés des poids (ou pondérations). La similarité perceptive S_{ij} entre deux stimuli i et j correspond à la somme des poids W_k associés à ces propriétés discrètes $\{P\}$ que les deux stimuli partagent. Dans ce cas, un cluster d'objets est la réunion des objets possédant le même trait ou la même propriété. Un tel modèle peut s'écrire sous la forme :

$$S_{ij} = \sum_k W_k P_{ik} P_{jk} \quad (\text{Eq. 39})$$

où $P_{ik} = 1$ si l'objet i possède la propriété k et W_k est une valeur de pondération non négative représentant la saillance de la propriété k . Si les deux objets sont dans un sous-ensemble donné (c'est-à-dire qu'ils partagent le même trait k), alors $P_{ik}P_{jk} = 1$, mais si l'un des deux objets appartient à un autre sous-ensemble, ce produit est nul.

- les **clusters hiérarchiques**, appelée également arbres ultramétriques, où dans ce cas, le recouvrement des sous-groupes est possible et les objets peuvent appartenir à plusieurs clusters de différents niveaux hiérarchiques [Johnson, 1967]. Un arbre

²⁹ <http://w3.octogone.univ-tlse2.fr/lordat/p-gaillard/tcl-labx/>

ultramétrie représente la similarité entre deux objets – correspondant aux noeuds extérieurs uniquement – par la hauteur du lien les joignant. Les objets les plus similaires sont reliés au plus bas niveau de l'arbre et les objets les plus dissemblables au niveau le plus haut. Un tel arbre est appelé 'ultramétrique' car il doit satisfaire l'inégalité ultramétrique selon laquelle pour tout x, y et z :

$$d(x, y) \leq \max[d(x, z), d(y, z)] \quad (\text{Eq. 40})$$

où $d(x, y)$ est la distance entre x et y .

Autrement dit, deux objets reliés à un noeud donné doivent être tous deux également distants d'un troisième objet auquel ils sont reliés à un noeud supérieur. Ceci implique que tous les objets sont à égal distance de la racine, correspondant au noeud le plus haut de l'arbre (voir figure 101).

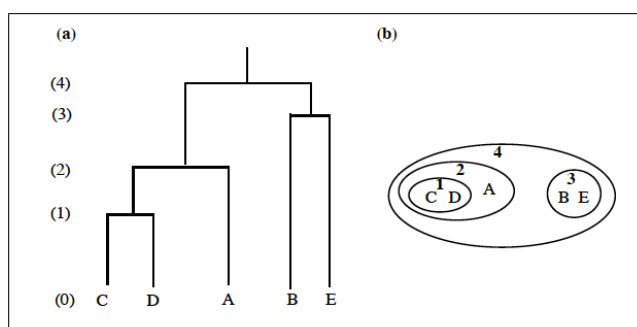


Figure 101. Représentation graphique d'un arbre ultramétrique

Les conditions nécessaires à une approche de type catégorisation libre peuvent se résumer par : i/ concevoir un large corpus de sons de natures diverses ; ii/ recueillir des informations de la part des sujets sur le type de similarité utilisé lors de la catégorisation : ii/ sélectionner un type de similarité donné [Susini et al., 2011].

III.3.3 Test de différentiel sémantique

Dans la méthode du différentiel sémantique, les échantillons sonores sont présentés au sujet séquentiellement ou en bloc et celui-ci doit effectuer des jugements sur des échelles dites "différentielles sémantiques" correspondant à des attributs perceptifs prédéfinis. Cette méthodologie, introduite par Osgood [Osgood, 1952], a été largement utilisée par la suite ; ses principes ayant été repris pour élaborer des approches de type analyse sensorielle [Bassereau et al., 1996].

Les échelles peuvent être de deux formes :

- bipolaire, c'est-à-dire définie par un couple d'attributs a priori opposés (par exemple, grave / aigu)
- unipolaire, c'est-à-dire définie par un seul attribut pondéré par un adjectif d'amplitude (par exemple, peu aigu / très aigu) ; cette variante, appelée VAME (Verbal Attribute Magnitude Estimation) a été proposée par Kendall et coll. [Kendall et al., 1993] pour palier un éventuel défaut de symétrie d'une échelle bipolaire.

Par ailleurs, la définition des échelles peut être effectuée soit arbitrairement, soit sur la base de données bibliographiques, soit au moyen d'une expérience préalable de verbalisation libre sur les sons et d'une analyse lexicale faisant ressortir les termes les plus pertinents par rapport au corpus d'étude. Cette étape constitue un point crucial de la méthode dans la mesure où l'on doit s'assurer que les échelles sont correctement et uniformément comprises par l'ensemble des sujets.

Pour l'analyse, les données issues de l'expérience se présentent, pour chaque sujet, sous la forme d'un vecteur de scores relatifs aux échelles de jugement. Moyennant les précautions d'usage (distribution gaussienne des données, suppression d'éventuels sujets *outliers*, voire test de χ^2 si plusieurs groupes de sujets), une réduction des données peut être effectuée par un calcul statistique simple de type moyenne / écart-type. Ces résultats sont ensuite généralement présentés sous la forme d'une représentation multidimensionnelle appelée **profil sémantique** – ou portrait sémantique (voir figure 102, pour exemple).

En outre des analyses plus poussées (analyse en composantes principales, régression multiple, etc.) peuvent être menées pour révéler les éventuelles interactions perceptives ayant eu lieu au cours de l'expérience entre les échelles sémantiques.

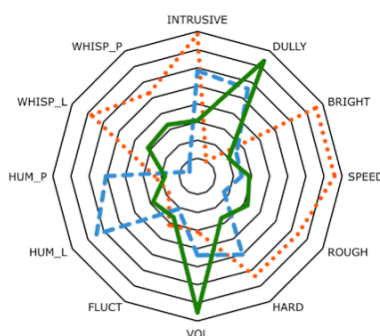


Figure 102. Exemple d'un profil (portrait) sémantique

3 sons de systèmes d'air conditionné sur 12 échelles différentielles sémantiques :
les labels 'L' et 'P' correspondant respectivement à l'intensité et la hauteur des parties souffle (*whisp*) et moteur (*hum*) – extrait de [Siekierski et al., 2001]

Les conditions nécessaires à une approche de type différentiel sémantique peuvent se résumer par : i/ contrôler le corpus en hauteur, durée et intensité ; ii/ sélectionner un nombre optimal d'échelles ; iii/ définir des attributs sémantiques pertinents et leur signification associée [Susini et al., 2011].

III.4 Timbre des sons musicaux : développement d'un moteur de recherche (projet SOL)

III.4.1 Préambule

Les résultats d'études perceptives sur le timbre musical et la recherche de corrélats acoustiques associés ont été mis à profit dans le cadre du développement d'un moteur de recherche adossé à une importante base de données de sons instrumentaux enregistrée et produite à l'Ircam.

L'application SOL (*Studio On Line*) – devenue par la suite *SoundPalette* – est issue de deux projets de recherche collaboratifs CUIDADO³⁰ (*Content-based Unified Interfaces and Descriptors for Audio/music Databases available Online*) et ECRINS³¹ (Environnement de Classification et de Recherche Intelligente des Sons). C'est un

³⁰ <http://anasynt.ircam.fr/home/projects/cuidado>

³¹ <http://anasynt.ircam.fr/home/projects/écrins>

développement qui répond aux besoins des professionnels du son et de l'audio ainsi que des compositeurs et des interprètes. Il donne accès à une collection de plus de 20.000 échantillons sonores issus de 16 instruments (clarinette, flûte, violon, trompette, violoncelle, cor, hautbois, trombone, alto, contrebasse, basson, saxophone, harpe, guitare, tuba et accordéon) enregistrés note par note, sur toute leur tessiture, dans plusieurs nuances, selon la plupart des modes de jeu traditionnels ou contemporains, et suivant 4 prises de sons différentes (micro interne, micro de proximité, couple stéréo standard, prise de son "dipolaire" dans le champ réverbéré de la salle).

L'environnement comprend une interface de classification et de recherche par le contenu (la "Palette Sonore") qui permet d'accéder à la base de données d'échantillons en ligne, y ajouter ses propres échantillons, classer l'ensemble en s'appuyant sur des fonctionnalités de classification automatique et, donc, effectuer des recherches sur la base de similarités acoustiques entre les sons représentés par des descripteurs. Actuellement, le déploiement "en ligne" de l'application n'est plus maintenu et il n'en reste qu'une version utilisable uniquement sur les serveurs internes de l'Ircam.

L'objectif de cette étude est donc d'élaborer un outil de recherche « intelligente » qui permette de guider la sélection des sons dans la base SOL. Le moteur de recherche est conçu initialement sur deux niveaux : une recherche par comparaison délivrant les sons les plus proches d'un son de référence donné et une recherche par description permettant de définir sémantiquement les caractéristiques du son désiré – mais qui n'a pas été implémenté jusqu'à présent.

Les études antérieures fournissent les dimensions perceptives saillantes du timbre musical et les paramètres physiques qui leurs sont corrélés. A partir de ces données, un modèle de distance est formulé et permet de calculer théoriquement la distance perceptive entre deux sons. L'exportation de ce modèle théorique au sons réels de la base SOL impose une optimisation du calcul des paramètres physiques et un ajustement de l'influence de chacun des critères perceptifs sur le calcul de distance au moyen de fonctionnalités complémentaires.

Pour ce faire, de nouveaux paramètres viennent compléter la description des sons de la base, constituée initialement de 4 paramètres de référence intervenant dans le modèle de distance. Ces critères additionnels permettent de construire des classes restreintes disjointes ou recouvrables : sons entretenus ou percussifs, sons de hauteur et/ou de nuance voisine du son de référence. L'accès à ces options de recherche oriente le calcul par similarité et aboutit à des résultats plus précis. En outre, le degré d'ajustement des paramètres de référence dans le calcul de distance a été normalisé : la focalisation concerne 3 dimensions perceptives (Brillance, Richesse, Attaque) et un "zoom" est défini par 3 positions (normal, plus et moins) permettant ainsi de traduire l'influence donnée à chaque paramètre (voir figure 103).



Figure 103. Interface du moteur de recherche par similarité de la base SOL.

Ces travaux de recherche et de développement se sont principalement poursuivis par une participation à la rédaction d'un brevet, déposé en 2001, sur un « procédé de caractérisation du timbre d'un signal sonore selon au moins un descripteur » [Peeters et al., 2001] puis une contribution à la rédaction d'un article de référence publié en 2011 sur la définition exhaustive d'une librairie de calcul de descripteurs acoustiques, la « *Timbre Toolbox* » [Peeters et al., 2011].

Ils ont par ailleurs fait l'objet de communications scientifiques au 16th *International Congress on Acoustics / 135th Acoustical Society of America meeting* (ICA/ASA) [Misdariis et al., 1998a], au 4^{ème} Journées Françaises sur la Psychologie de la Musique [Misdariis et al., 1998b] ainsi qu'à la conférence *Human Centered Process* [McAdams et al., 1999].

III.4.2 Publication [Misdariis et al., International Congress on Acoustics, 1998a]

Validation of a Multidimensional Distance Model for Perceptual Dissimilarities among Musical Timbres

Nicolas Misdariis^(a), Bennett K. Smith^(a), Daniel Pressnitzer^(a), Patrick Susini^(a), Stephen McAdams^(a,b)

^(a) Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), 1 place Igor Stravinsky, F-75004 Paris, France.

^(b) Laboratoire de Psychologie Expérimentale (CNRS), Université René Descartes, EPHE, 28 rue Serpente, F-75006 Paris, France.

III.4.2.1 Theoretical Basis

Musical timbre is a complex auditory attribute for which multidimensional scaling techniques (MDS) have usually been employed to reveal the main underlying perceptual dimensions: from dissimilarity judgments on pairs of sounds, an MDS analysis builds a low-dimensional space where sounds are located with regards to their mutual timbral differences. In this context, the study of Krumhansl [Krumhansl, 1989] on a set of 21 synthesized sounds of musical instruments and the derived study of McAdams et al. [McAdams et al., 1995] — with a more elaborated MDS model — on 18 stimuli of the same set are used as the basis for the present work. In both cases, the analysis leads to a timbre space with three common dimensions ; these will be referred to as 'Space 1' and 'Space 2', respectively.

Krimphoff et al. [Krimphoff, 1993 ; Krimphoff et al., 1994] attempted to quantify the perceptual axes of Space 1 in terms of acoustic parameters and the parameters derived were tested on Space 2. Both spaces gave similar results for the two of the dimensions, which fit with Spectral Centroid (average over the sound duration of the instantaneous spectral centroid within a running time window of 12 ms) and Log Attack-Time (rise time measured from the time the amplitude envelope reaches a threshold of 2% of the maximum amplitude to the time it attains this maximum amplitude). As for the third dimension, the most significant correlation occurs, for Space 1, with Spectral Irregularity (log of the spectral deviation of component amplitudes from a global spectral envelope derived from a running mean of the amplitudes of three adjacent harmonics) and, for Space 2, with Spectral Flux (average of the correlations between amplitude spectra in adjacent time windows) . The values of the correlation coefficients are shown below in Table 6.

	Spectral Centroid	Log Attack-Time	Spectral Irregularity	Spectral Flux
Space 1	+ 0.93	– 0.94	– 0.87	
Space 2	– 0.94	– 0.94		+ 0.54

Table 6. Correlation between perceptual dimensions and physical parameters (Krimphoff et al. [4,5])

III.4.2.2 Construction of the Distance Model

The main objective of this study was to build a multidimensional distance model from the relations among the perceptual criteria obtained by MDS analysis and the calculated physical parameters. First, the subjective data, i.e. the coordinates of the sounds along each perceptual axis, were normalized to make the results of the two initial studies comparable: a multiplicative factor of 1.93 was applied to Space 1. Afterwards, a mathematical relation between these subjective data and the values of the corresponding parameter was obtained by a linear regression method: the slopes of the resulting six straight lines could then be considered as proportionality coefficients between the perceptual and the physical dimensions. Finally, these coefficients were used as weighting factors for the four acoustic correlates in a Euclidean distance formula representing the perceptual distance, as described in Eq. 41:

$$DIST = \sqrt{3.5385 \cdot 10^{-5} \times SC^2 + 15.5236 \times LT^2 + 0.011881 \times SI^2 + 2728.7 \times SF^2} \quad (\text{Eq. 41})$$

where SC, LT, SI, and SF are the differences in Spectral Centroid (Hz), Log Attack-Time (\log_{10} (s)), Spectral Irregularity (dB) and Spectral Flux (unitless), respectively, and DIST is the resulting distance between two sounds in this acoustic parameter space. Note that the coefficients for SC and LT are derived from the mean of the values from Space 1 and Space 2.

III.4.2.3 Application of the Model to a Real Database

The global aim of this work is the development of a search engine within a database of recorded musical sounds. However, the initial studies from which the model is derived were based on a restricted set of sounds controlled in pitch, duration, and loudness. To the contrary, the database upon which the model is expected to make predictions covers at present eleven instruments recorded over their entire pitch range, at all dynamics and using almost all playing techniques of each instrument (normal, flutter tonguing, staccato, multiphonics, etc.). Therefore, a validation step is necessary both in the definition of the perceptual dimensions and in the calculation method of the parameters.

For the time being, we have focused on the problems of parameter extraction when processing acoustically produced sounds rather than synthesized ones, such as a non negligible signal-to-noise ratio, natural or forced inharmonicity, high non-stationarity, etc.. To illustrate a typical problem often encountered, Figure 104 presents the Spectral Centroid as a function of the physical level (RMS) over the duration of a Bb0, staccato, trombone sound. The existence of areas where the Spectral Centroid information is quite false seems obvious: taking with reference the frame where RMS level is the highest (approximately between 0.4 and 0.6 seconds) and the theoretical pitch of a Bb0 in the tempered scale (55 Hz), we can notice an over-estimation of the parameter elsewhere.

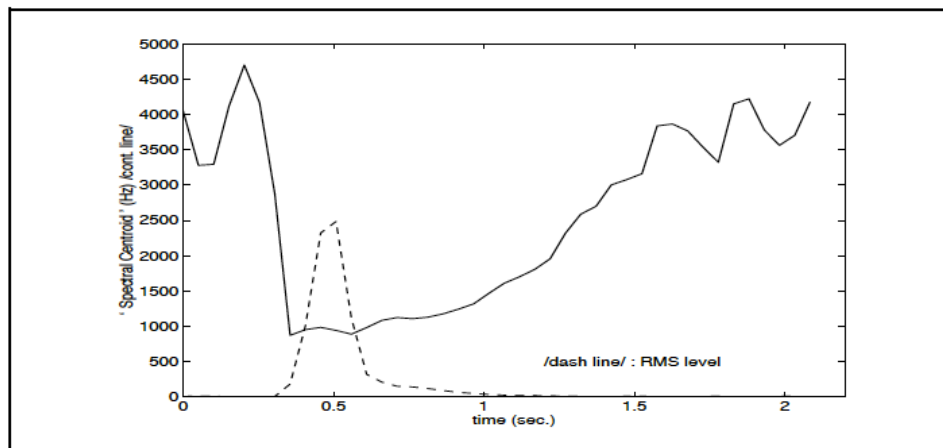


Figure 104. Spectral Centroid vs RMS level of Bb0, staccato trombone sound

The adjustments suggested by this kind of investigation have been made directly within the parameter calculation procedures: different ways to calculate a given parameter have been tested and elements like noise-gate modules in the generation of spectral data have even been implemented. Since the modified model in its present state gives encouraging results, the next step of development will be to quantify empirically different perceptual scales, such as roughness, in order to establish more accurately the relation between the distances in perceptual and physical parameter spaces.

III.5 Evaluation objective de la qualité perçue d'un instrument de musique

III.5.1 Préambule

L'étude de la qualité d'un instrument de musique – p. ex., le violoncelle [Petiot et al., 2007] –, ou d'une partie constituant l'instrument – p. ex. les anches de saxophone [Gazengel et al., 2012] – est depuis plusieurs dizaines d'années au centre des préoccupations des chercheurs en acoustique musicale [Campbell, 2013]. Cet axe de recherche est motivé, d'une part, par la volonté d'accroître les connaissances générales sur le fonctionnement des instruments de musique, voire comprendre certains mythes entretenus de longue date (p. ex., la qualité des violons fabriqués par les facteurs italiens de Crémone, notamment Stradivarius, au cours du XVIII^{ème} siècle [Fritz et al., 2012]), et d'autre part, par la perspective de fournir des aides efficaces aux luthiers en termes de prédiction et de contrôle de la qualité de leurs instruments. Cette section d'études se situe dans cette lignée de recherche en proposant deux approches distinctes mais complémentaires.

La première s'intéresse de manière "classique" à la caractérisation de la qualité d'un ensemble de trompettes et développe une méthodologie expérimentale basée sur l'interaction avec l'instrument – et non l'écoute passive en cabine audiométrique –, la verbalisation du ressenti des musiciens et la corrélation avec des paramètres physiques mesurables sur l'instrument lui-même ou le son qu'il a produit. Cette étude a fait l'objet d'une communication scientifique au 4^{ème} Congrès Français d'Acoustique (CFA) [Caussé et al., 1997] et au 137^{ème} meeting de l'*Acoustical Society of America* [Caussé et al. 1999b].

La seconde s'intéresse, de manière plus originale, aux conséquences perceptives d'un procédé technique de contrôle actif appliqué à un pseudo-instrument à vent de type clarinette basse, constitué d'un tuyau cylindrique excité par un bec de clarinette (anche simple). Le contrôle actif permettant, par définition, de modifier les propriétés acoustiques du résonateur, l'étude s'intéresse en fait – et toujours *via* un paradigme interactif prenant en compte le système instrument+musicien – à la pertinence perceptive des paramètres de contrôle et à la détermination d'éventuels seuils en-deçà ou au-delà desquels les effets de ce procédé innovant ne sont plus ressentis par le musicien.

Cette étude s'est déroulée dans le cadre du stage de Master I de Marguerite Jossic (ENS Ulm) et a fait l'objet d'une communication scientifique au XII^{ème} Congrès Français d'Acoustique [Jossic et al., 2014].

III.5.2 Publication [Caussé et al., Congrès Français d'Acoustique, 1997]

Caractérisation Objective de la *Qualité* d'un Instrument de Musique

René Caussé, Nicolas Misdariis, Sébastien Thoreton

IRCAM – 1, place Igor Stravinsky – 75004 Paris

III.5.2.1 Introduction

L'objet de cette étude est de tenter d'interpréter les processus d'évaluation de la qualité d'un instrument de musique : c'est-à-dire, relier les critères subjectifs d'appréciation qu'utilise le musicien pour juger la qualité d'un instrument avec des paramètres physiques mesurables sur l'instrument lui-même. Cette étude présente de fortes similitudes avec les travaux menés, par exemple, en acoustique des salles et qui visent à caractériser la qualité acoustique d'une salle de concert [Kahle, 1995].

Notre approche s'effectue ici de manière préliminaire et prospective sur une série de trompettes en Ut. Elle comprend une procédure expérimentale (tests perceptifs avec des trompettistes professionnels) qui permet de recueillir les données nécessaires à une analyse statistique aboutissant à la définition des dimensions perceptives communes utilisées par l'ensemble des musiciens. Parallèlement à cela, on effectue un travail d'analyse physique des trompettes qui permet de dégager un ensemble de paramètres caractérisant de manière objective tous les instruments. La mise en corrélation de ces données de deux natures différentes - perceptive et objective - fournit ensuite les résultats qui servent de base à l'interprétation de la relation entre qualité musicale et propriétés physiques d'un même instrument.

III.5.2.2 L'évaluation subjective des instruments

III.5.2.2.1 Les tests perceptifs

Dans cette phase expérimentale, on confronte 8 trompettistes professionnels avec un ensemble de 10 trompettes en Ut haut-de-gamme.

L'objectif du test est de recueillir des valeurs de dissemblance entre toutes les trompettes. Pour ce faire, le protocole expérimental consiste à présenter les instruments par paire et à demander aux sujets d'évaluer, sur une échelle de 1 à 10, la 'distance perceptive' qui sépare les objets d'une même paire.

La séance est soumise à des conditions expérimentales particulières motivées par le souci de minimiser les biais de jugement qui sont ici potentiellement importants du fait

du contact direct entre le sujet et les objets du test : le musicien a les yeux bandés et porte des gants de coton; les éléments caractéristiques de chaque trompette (anneaux, embouts de piston) sont enlevés ou camouflés en prenant soin toutefois de ne pas altérer la tenue générale de l'instrument et, par ailleurs, le lieu d'expérience est choisi en fonction des habitudes de travail des instrumentistes.

Le test de dissemblance s'effectue ainsi sur 45 paires d'instruments et permet de recueillir, pour chaque sujet, une matrice de dissemblance qui constitue la donnée d'entrée principale du processus d'analyse multidimensionnelle. En outre, cette première étape est suivie d'une seconde, beaucoup plus informelle, au cours de laquelle le musicien s'exprime verbalement sur chaque trompette et peut donner les attributs sémantiques qui lui semblent caractéristiques de chaque instrument : "compacte", "timbrée", "équilibrée", "fermée", "chère dans l'aigu", "très ouverte", etc. ...

III.5.2.2.2 L'analyse multidimensionnelle

Le procédé d'analyse multidimensionnelle utilise des algorithmes qui assurent la conversion des dissemblances en distances géométriques équivalentes construisant ainsi un 'espace perceptif' qui sous-tend les jugements subjectifs des trompettistes [Schiffman et al., 1981].

Pour construire cette représentation, la théorie de l'analyse suppose l'existence d'une relation monotone entre les valeurs expérimentales de dissemblance et les distances calculées afin de conserver une cohérence dans la variation des deux ensembles : lorsque deux objets sont jugés éloignés, le modèle les écarte d'une distance en rapport avec cette perception et suit la même logique avec deux objets jugés très proches.

D'un point de vue mathématique, il existe plusieurs modèles de calcul pour exprimer la distance entre deux points de l'espace d'analyse : nous utilisons ici le modèle euclidien étendu qui est implanté dans le programme EXSCAL [Winsberg et al., 1989]. Ce modèle suppose à la fois l'existence de dimensions communes à tous les objets du test mais aussi d'éventuelles spécificités propres à chacun d'eux (par exemple, le poids anormalement élevé d'un instrument par rapport aux autres); de plus, il tient également compte de l'effet pondérateur induit par chaque sujet qui traduit l'importance que le sujet attribue aux critères perceptifs du jugement. La dimension optimale de l'espace est définie grâce à un indice statistique qui permet d'estimer l'erreur entre l'espace construit et la situation réelle. Dans nos résultats expérimentaux, l'espace perceptif est à 3 dimensions.

La définition sémantique des axes est effectuée à posteriori en exploitant les informations recueillies lors de la seconde partie du test perceptif et après avis auprès de certains des trompettistes. Ainsi, les trois dimensions de l'espace correspondent respectivement à "la largeur du son" (1), la "touche" (qui définit la réponse de l'instrument notamment dans des phases de jeu précises telles que le staccato) (2) et la "facilité de jeu" (3).

III.5.2.3 La caractérisation physique des instruments

III.5.2.3.1 Approche morphologique

On effectue un relevé de cotes en plusieurs points sur les deux éléments les plus influents de l'instrument : la branche avec l'embouchure (partie de la trompette en amont des pistons) et le pavillon; il comprend :

- la mesure du diamètre intérieur en plusieurs points de la branche, de l'épaisseur moyenne de métal et de la longueur de branche directe, c'est-à-dire avant la première

discontinuité (exceptée celle due à l'embouchure).

- la mesure du diamètre extérieur en plusieurs points du pavillon en prenant soin de concentrer les points de mesure à l'endroit où la courbure est la plus grande.

Ces données directes constituent un premier ensemble de grandeurs physiques spécifiques; ils permettent d'accéder à des paramètres plus complexes tels que le coefficient d'évasement du pavillon qui rend compte de l'évolution de son profil.

III.5.2.3.2 Approche acoustique

Des mesures d'impédance d'entrée sont effectuées pour les sept positions de jeu de chaque instrument. Elles servent de base au calcul de plusieurs paramètres acoustiques :

- les fréquences réduites de la série harmonique à partir desquelles on calcule plusieurs critères d'inharmonicité.
- le facteur de qualité des résonances à -3 dB.

- le centre de gravité spectral des résonances : $CGS = \sum_{p=1}^P p \cdot h_p / \sum_{p=1}^P h_p$ (Eq. 42)

où h_p est l'amplitude du partiel p et P le nombre total de partiels considérés.

- la déviation d'amplitude des résonances : $DEV = \sum_{p=1}^P |h_p - (h_{p-1} + h_p + h_{p+1})/3|$ (Eq. 43)

obtenue à partir de l'écart à une enveloppe spectrale moyenne.

Par ailleurs, quelques mesures de réponse impulsionnelle ont été mises en œuvre mais la disponibilité réduite des trompettes n'a pas permis de généraliser cette tâche à l'ensemble des instruments de l'étude, de ce fait, les paramètres acoustiques issus de la réponse temporelle de l'instrument ne sont pas pris en compte dans la caractérisation physique.

III.5.2.4 Relation critères subjectifs / paramètres physiques

Les deux ensembles corrélés sont {coordonnées dans l'espace perceptif} et {valeurs des paramètres physiques}. Ce calcul permet d'évaluer le degré de correspondance linéaire entre un élément de chaque nature.

Dans le cas présent, l'examen des coefficients de corrélation met à jour un paramètre physique influent pour chacun des axes perceptifs de l'espace d'analyse; ce sont :

- le facteur de qualité du 2ème pic d'impédance (pour la position 'à vide' des pistons) pour la "largeur du son" [$R = 0.686$];
- le diamètre moyen de la branche pour la "touche" [$R = -0.736$];
- le paramètre $\text{var}(f_p / p)$ (position 'à vide') pour la "facilité de jeu" [$R = -0.689$].

Ces premiers résultats sont encourageants puisqu'ils permettent d'ébaucher un modèle objectif complet de caractérisation de l'instrument sous la forme du jeu des trois paramètres physiques décrits précédemment.

Cependant, ceux-ci ne sont pas uniques et plusieurs autres grandeurs physiques présentent aussi une assez forte corrélation avec l'une des trois dimensions de l'espace : par exemple, la déviation spectrale (pour la position 2) avec la "facilité de jeu" [$R = 0.668$] ou bien le facteur de qualité du 6ème pic d'impédance (pour la position 'à vide') avec la "largeur du son" [$R = -0.693$]. Cette redondance de résultats acceptables remet en question la pertinence du choix des paramètres et donc la fiabilité du modèle proposé.

III.5.2.5 Conclusion et perspectives

L'étude qui a été menée amène plusieurs observations. D'une part, le traitement multidimensionnel semble bien adapté au type de résultats que nous voulons tirer de l'expérience, et ceci malgré la complexité des objets du test. D'autre part, les mesures physiques pourraient être encore plus complètes et précises, en généralisant notamment la mesure de réponse impulsionnelle des instruments testés. Enfin, malgré un premier résultat encourageant, la pertinence de nos interprétations restent à définir plus solidement.

Pour cela, la suite logique de ce travail passe par des séries d'expériences mieux maîtrisées au niveau de la variation des grandeurs physiques des instruments testés : comme, par exemple, une nouvelle étude avec des trompettes identiques sur tous les points mais dont seul le profil de pavillon varie régulièrement entre le 'très fermé' et le 'très ouvert'. Ce genre d'étude nécessiterait une collaboration étroite avec les facteurs d'instruments.

Dans le futur, ce projet peut intéresser à double titre la facture instrumentale. En premier lieu, la connaissance précise et fiable des paramètres physiques influant sur le comportement musical de l'instrument peut devenir une aide précieuse à la facture en permettant une évaluation de la qualité au cours de la fabrication ainsi qu'une appréciation finale basée sur des critères objectifs et systématiques. A l'inverse, la mise à jour des liens entre facteurs perceptifs du jeu instrumental et grandeurs physiques mesurables sur l'instrument lui-même peut permettre de s'orienter vers une facture innovante en matière de comportement musical et, par application, de répondre aux attentes et aux interrogations des musiciens.

III.5.3 Publication [Jossic et al., Congrès Français d'Acoustique, 2014]

Perception du Contrôle Actif sur Instruments à Vent

M. Jossic, T. Meurisse, N. Misdariis, A. Mamou-Mani
STMS Ircam-CNRS-UPMC, 1 place Igor Stravinsky, 75004 Paris, France

III.5.3.1 Introduction

Le contrôle actif, dont le tout premier brevet est attribué à Paul Lueg (1934) pour ses travaux sur la réduction du bruit dans les tuyaux, est aujourd'hui une technique largement utilisée dans l'industrie afin de réduire les nuisances sonores. Le principe de base est simple : l'élimination d'un bruit indésirable dans une structure se fait par l'introduction de sources de bruit secondaires. Celles-ci sont calculées de sorte que leur addition au bruit initial réduise l'amplitude de ce dernier.

Dans un cadre non plus industriel mais musical, l'utilisation du contrôle actif ne se fait plus dans une optique de réduction de vibrations - ce qui est compréhensible - mais plutôt de modifications de ces vibrations, afin d'agir sur le timbre de l'instrument. Dans notre cas, un système de contrôle actif a été greffé sur une clarinette simplifiée. Des travaux précédents [Meurisse et al., 2013] ont montré qu'un contrôle actif analogique induisait des modifications acoustiques nettes dans les caractéristiques modales du système instrumental. La question qui se pose alors est de savoir dans quelle mesure ces modifications acoustiques de l'instrument entraînent des modifications perceptives pour la personne qui en joue. Les objectifs généraux de cette étude concernent la quantification des différences perçues par les musiciens lorsqu'ils jouent d'un instrument contrôlé, et la qualification de ces différences par un vocabulaire musical.

III.5.3.2 Système de contrôle actif numérique

Le choix a été fait d'effectuer un contrôle actif numérique de l'instrument. En effet, le contrôle analogique, déjà réalisé dans une étude précédente [Meurisse et al., 2013], se prête peu aux procédures automatisées qui sont en général nécessaires à une démarche expérimentale contrôlée. De plus, le contrôle numérique permet d'affiner le type de contrôle que l'on souhaite effectuer sur l'instrument, via le logiciel Max/MSP, dont on détaille l'utilisation dans la suite de l'article.

III.5.3.2.1 Fonction de transfert en boucle fermée

Le système de contrôle numérique, réalisé sous Max/MSP, contrôle toutes les sorties et entrées d'une carte son analogique et convertisseur A/N MOTU-828. Le système global contenant le musicien, l'instrument, ainsi que le système de contrôle est représenté sur la figure 105.

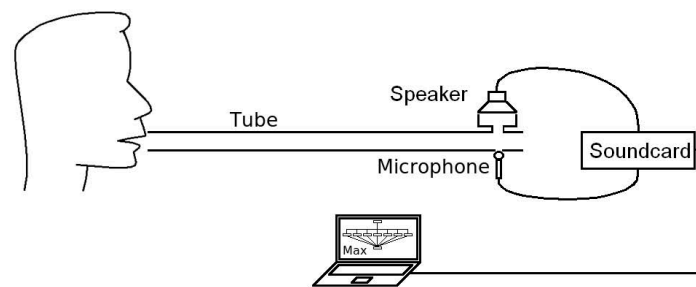


Figure 105. Schéma simplifié du système de contrôle numérique monté sur le système tube+anche

La première étape de l'étude consiste à modifier les paramètres de la fonction de transfert en boucle fermée du système. Celle-ci est calculée à l'aide du montage représenté sur la figure 106 : Un sweep fréquentiel est généré via un programme MATLAB sur ordinateur (voir ①). Il est, d'une part, directement relié à l'ordinateur de mesure ② en tant que signal d'entrée pour la mesure de la fonction de transfert (en vert), et de l'autre envoyé dans le tube via un haut-parleur (Tymphany Peerless PLS-P830983) ③. Le signal modulé par le tube est récupéré par le microphone (Endevco modèle 8507C-5), puis modifié via le système de contrôle ④, avant d'être réinjecté dans le tube par le même haut-parleur ③. Le microphone capte donc la somme du signal initial U et de ce même signal U modifié. La somme des deux signaux est alors envoyée à l'ordinateur de mesure, en tant que sortie du système (⑤, en bleu).

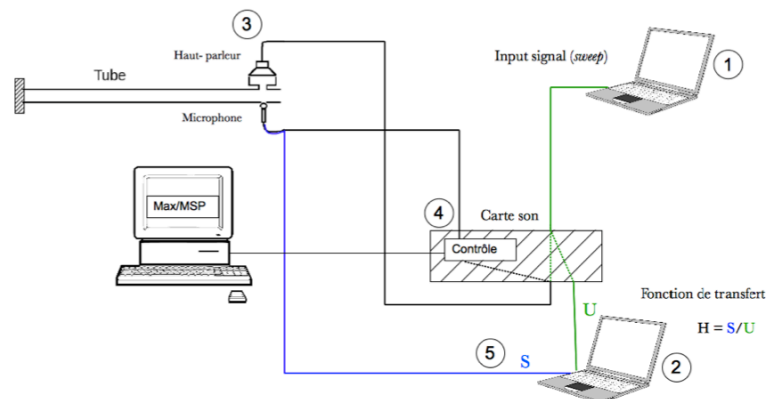


Figure 106. Dispositif expérimental pour mesure de la fonction de transfert en boucle fermée du système

III.5.3.2.2 Phase en boucle ouverte

La chaîne des composants du système de contrôle tel qu'il a été décrit précédemment possède un retard non nul. Ce retard entre l'onde initiale et l'onde initiale modifiée est appelée phase en boucle ouverte, ou phase en BO. Il est dû à plusieurs retards temporels intrinsèques à la chaîne HP+micro+système de contrôle. Dans le cas où aucun contrôle n'est effectué ($G=0$), la chaîne de retards est composée d'un retard Δt_1 dû à la distance entre le HP et le micro, et d'un retard numérique ou de traitement Δt_2 (de 25 ms) dû au système de contrôle Max/MSP. Cette chaîne de retards successifs impose que chaque résonance n de fréquence f_n subit un déphasage ϕ_n entre le moment où le sweep émet la fréquence f_n via le haut-parleur (③), et le moment où elle est réinjectée par le même haut-parleur après être passée dans le système de contrôle (④). Cette phase ϕ_n est telle que : $\phi_n = (\Delta t_1 + \Delta t_2) \cdot 2\pi \cdot f_n$. Le retard Δt_2 entraîne une variation rapide de la phase en boucle ouverte de la fonction transfert du système (voir figure 108). Ceci aura son importance dans l'établissement des configurations (voir § III.5.3.4.1).

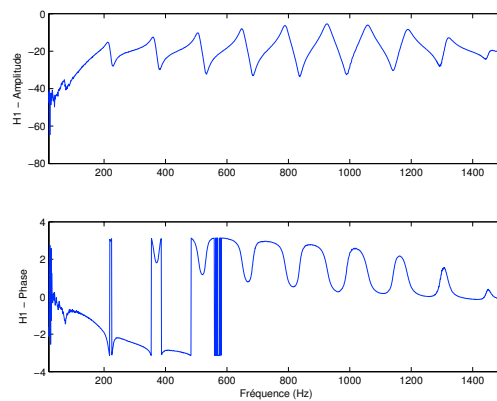


Figure 107. Fonction de transfert en boucle fermée du système avec contrôleur numérique (configuration sans contrôle actif)

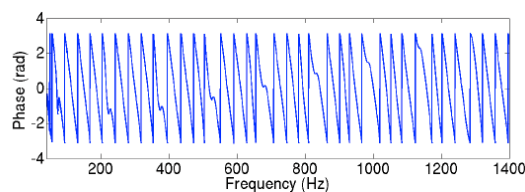


Figure 108. Phase en boucle ouverte de la fonction de transfert du système

III.5.3.2.3 Description du contrôle numérique

A l'aide du logiciel Max/MSP, un ensemble de 10 filtres passe-bandes est monté en parallèle (on ne réalise pas de contrôle au delà du dixième mode, où l'amplitude de la fonction de transfert est faible). La fréquence, le gain ainsi que le facteur de qualité du filtre peuvent être ajustés (voir figure 109). La valeur du facteur de qualité a été fixé à $Q = 50$. Afin de compenser les effets de la phase en boucle ouverte du système, un objet "delay" est ajouté à chacun des filtres : il permet d'imposer le retard temporel désiré afin de modifier la phase de la résonance comme souhaité. En effet, si l'on note ϕ_n la phase en boucle ouverte de la résonance n , et Δdelay le retard supplémentaire introduit par l'objet "delay" à cette même résonance n , celle-ci possède non plus une phase ϕ_n mais une phase ϕ_c telle que $\Delta \text{delay} = (\phi_n - \phi_c) \cdot T_n / 2\pi$ où T_n est la période de la résonance n ($T_n = 1/f_n$).

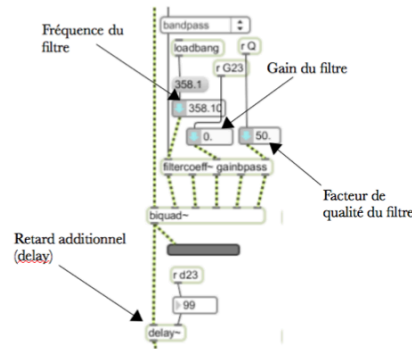


Figure 109. Filtre passe-bande dans Max/MSP utilisé pour le système de contrôle numérique

Inversement, on peut choisir que la résonance n de phase ϕ_n en boucle ouverte ait une certaine phase cible ϕ_c – par exemple, pour que l’onde modifiée et l’onde initiale s’ajoutent en phase, il faut choisir $\phi_c = 0$. Il suffit alors de calculer le Δdelay nécessaire à appliquer dans Max/MSP à l’aide de l’équation précédente.

On dispose donc d’un système de contrôle numérique qui permet d’augmenter l’amplitude et la fréquence de chaque résonance du système, indépendamment les unes des autres. On donne sur la figure 110 un exemple des modifications de la fonction de transfert (FT) en boucle fermée qu’il est possible de réaliser sur la deuxième résonance.

Les phases ϕ_c choisies ont pour valeurs 0 , $\pi/2$, π , et $3\pi/2$. On observe par exemple qu’une phase $\phi_c = 0$ augmente l’amplitude de la résonance, et qu’une phase de $\phi_c = \pi$ la diminue. De même, les déphasages $\phi_c = \pi/2$ et $\phi_c = 3\pi/2$ ont respectivement pour conséquence d’augmenter ou de diminuer la fréquence de la résonance. Ces résultats sont en accord avec ceux issus du contrôle analogique réalisé dans l’étude précédente [Meurisse et al., 2013]. Sur la figure 110, les courbes vertes représentent les résultats issus de simulations MATLAB.

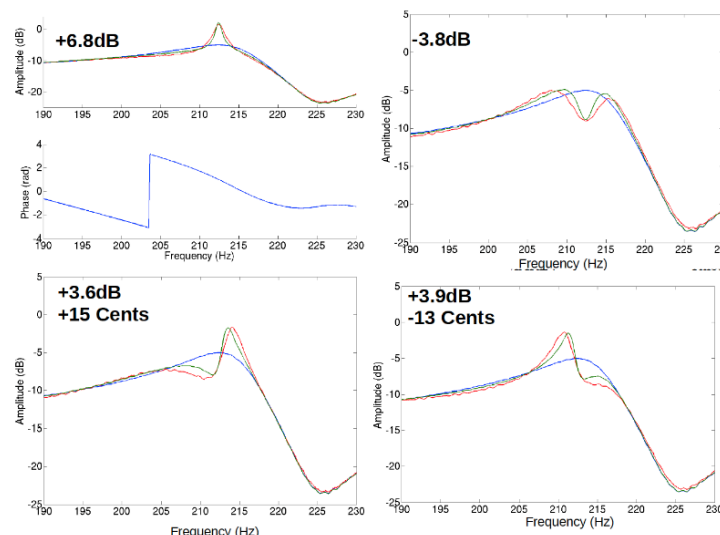


Figure 110. FT mesurée avec différents contrôles appliqués à la deuxième résonance de l’instrument.

Sans contrôle (bleu), avec contrôle (rouge) et simulation de contrôle (vert).

Les contrôles sont tous réalisés avec $G=1$, $Q=50$ et quatre phases ciblées différentes.

Haut-gauche : $\phi_c=0$. Haut-droite : $\phi_c=\pi$. Bas-gauche : $\phi_c=\pi/2$. Bas-droite $\phi_c=3\pi/2$.

Ces simulations sont réalisées en identifiant les fréquences de résonance de la fonction de transfert mesurée sans contrôle et contient, comme le système réel, un ensemble de dix filtres passe-bande permettant de contrôler les mêmes paramètres (fréquence f_i , facteur de qualité Q_i et retard τ_i). On observe que les effets du gain sont très localisés : cela est dû à la fois à la variation rapide de la phase en BO, et à la valeur élevée choisie pour le facteur de qualité des filtres passe-bande afin de limiter cet effet de la phase du système. Cela a par exemple pour conséquence l'apparition de deux pics voisins de la fréquence de résonance lorsque la phase cible est de π . De même, les décalages en fréquence s'accompagnent d'augmentations d'amplitude. Cette localisation des effets de gain sur les pics de résonance devra être prise en compte dans la suite.

III.5.3.3. Approche perceptive de l'étude

III.5.3.3.1 Approche psychophysique et écologique de l'étude

L'élaboration du protocole s'est faite par une approche à la fois psychophysique et écologique. Concernant l'aspect psychophysique, plusieurs travaux, dont ceux de McAdams [McAdams et al., 1995] et de Grey [Grey, 1977], ont tenté de démontrer l'aspect multidimensionnel de la perception sonore. Par ses travaux sur le timbre, Grey suppose l'existence d'un espace perceptif multidimensionnel, dont les dimensions perceptives (brillance, flux spectral, et densité de transitoires) sont révélées à l'aide d'expériences de dissemblance entre des paires de stimuli. Ces données de dissemblance sont ensuite "cartographiées" dans un espace à plusieurs dimensions via une analyse multidimensionnelle.

A l'inverse, les travaux de Gibson en écologie de la perception [Gibson, 1979], à la fin des années 70, veulent profondément remettre en question les fondements de la psychophysique et de la théorie de l'information. Le stimulus n'est plus un ensemble de variables physiques que le cerveau traite en vue de l'interpréter, mais il est lui-même l'information : celle-ci est le produit de l'interaction forte qui existe entre l'individu et l'environnement tel qu'il le perçoit. Ces nouveaux points de vue utilisent d'autres méthodes d'analyse que celles utilisées en psychophysique : les verbalisations et le langage permettent ainsi d'autres accès aux modes de représentations.

III.5.3.3.2 choix des tests

Dans la perspective de répondre à la problématique posée par l'étude, à savoir la perception de musiciens sur le contrôle actif d'instruments, le choix a été fait d'effectuer des tests où le musicien est placé en situation active, c'est-à-dire en situation de jeu. En effet, mettre en place des tests passifs requiert en général de disposer d'une banque de données sonores adéquate. Ce corpus sonore doit être proprement défini physiquement, c'est-à-dire qu'il nécessite d'être contrôlable par l'utilisateur afin que celui-ci puisse faire varier les paramètres de façon découplée et maîtrisée. Il aurait donc fallu avoir à notre disposition une bouche artificielle spécifique à l'anche de clarinette basse, ce qui n'était pas le cas.

Il se trouve que la littérature concernant les protocoles de tests perceptifs en situation active [Placido et al., 2011 ; Fritz et al., 2012] se révèle peu pertinente vis à vis de notre étude. Dans le souci de quantifier les différences perçues par un musicien, il a donc été décidé d'établir un protocole original à partir d'expériences de dissemblances, qui semblent appropriées bien que souvent réalisées dans le cadre de tests passifs. En

plus de cette approche psychophysique, nous avons choisi d'ajouter au protocole une phase de verbalisation libre au cours de laquelle le musicien est amené à faire part oralement des différences qu'il perçoit, en les qualifiant à partir de termes qui lui sont propres. Ce genre de protocole mixte, entre approche psychophysique (expériences de dissemblance) et écologique (verbalisations) avait déjà été mis en place dans plusieurs études précédentes [Faure, 2000].

III.5.3.4 Protocole expérimental

III.5.3.4.1 Espace des paramètres et configurations

L'ensemble des dix filtres qui contrôle les dix résonances du système offre un grand nombre de paramètres et donc de configurations possibles (10 paramètres de gain et 10 paramètres de delay). L'espace des configurations utilisées dans les expériences de dissemblance a été pré-déterminé avec l'aide d'un clarinettiste professionnel qui a évalué leur pertinence. Plusieurs tentatives ont mis en évidence la nécessité d'imposer un ordre de grandeur suffisamment important pour les gains imposés, au risque qu'aucune différence ne soit perçue, tout en évitant les phénomènes de larsen dus à la variation de la phase en BO. Le choix final a consisté à créer un espace de paramètres à 2 dimensions, l'une concernant le gain appliqué à toutes les résonances (sauf la première, dont la fréquence est trop basse par rapport aux capacités du haut-parleur, et l'amplitude très basse à la position du système de contrôle), et l'autre le déphasage imposé. L'espace des paramètres en question est représenté sur la figure 111.

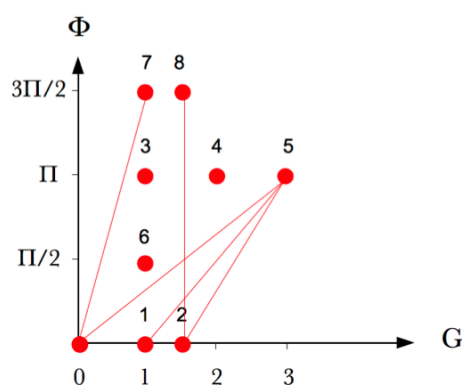


Figure 111. Espace des paramètres pour les configurations testées dans notre expérience.
En abscisse, le gain des filtres, en ordonnée, le déphasage imposé.

Les paires de configurations reliées par des traits rouges sont les paires ayant recueilli en moyenne, par la suite, les plus grands jugements de dissemblance (voir § III.5.3.5.2.2)

III.5.3.4.2 Sujets

Afin de respecter l'homogénéité de l'ensemble des musiciens effectuant le test, ceux-ci ont été choisis parmi des clarinettistes basses classiques et professionnels. En raison de la période où les tests ont été réalisés (juin 2013), seuls 9 musiciens ont été disposés à participer à l'expérience. Ce nombre étant peu élevé, les résultats exposés plus loin n'auront pas de valeur statistique à proprement parlé, bien qu'ils révèlent tout de même des aspects intéressants de l'étude.

III.5.3.4.3 Déroulement de la session

La session a permis de recueillir 2 séries de données : des données numériques de jugement de dissemblance de paires de configurations [(G1; ϕ_1); (G2; ϕ_2)], ainsi que des données verbales concernant la comparaison de ces mêmes paires de

configurations. Le recueillement des données ainsi que le déroulement de l'expérience étaient entièrement pilotés par Max/MSP et Psiexp, une interface de programmation d'expérience développée par B. Smith (1995). Une pédale de pied permettait au sujet de passer d'une configuration à l'autre pour la comparaison des paires. La session s'est déroulée dans un studio de l'IRCAM, et a duré entre une demi-heure et une heure trente en fonction des musiciens. Une phase de familiarisation de 5 minutes a été prévue, afin que le musicien apprécie l'échelle des différences perceptives induites par la variation des deux paramètres. Les paires de configurations ont ensuite été présentées au musicien. Celui-ci était laissé libre de les comparer comme il le souhaitait en passant, pour chaque paire, d'une configuration à l'autre à l'aide de la pédale. Il lui a ensuite été demandé de juger la dissemblance des deux configurations en déplaçant un curseur virtuel continu, ayant pour extrémités les termes "très similaires" (dissemblance de 0) à gauche et "très dissimilaires" (dissemblance de 1) à droite. Enfin, pour chaque paire, il a été invité à faire part oralement des différences qu'il avait entendues. La consigne "Commentez les différences entendues" était précisée oralement en début de séance, ainsi que pour chaque paire sur l'écran de l'ordinateur piloté par Psiexp. Toutes les verbalisations ont été enregistrées afin d'être traitées par la suite.

III.5.3.5 Résultats

Les résultats de deux des sujets (parmi les neuf ayant participé à l'expérience) ont été compromis en raison d'un dysfonctionnement de l'expérience. On présente donc ici les résultats de 7 sujets. Certains résultats des analyses verbales (voir figures 113, 114 et 115) présentent les données de 8 sujets, lorsque les verbalisations qui lui sont associées ont pu être récupérées.

III.5.3.5.1 MDS INDSCAL globale

Les données numériques de dissemblance ont été analysées par analyse multidimensionnelle (MDS). Le modèle utilisé est le modèle INDSCAL [Caroll et al., 1970], que l'on utilise classiquement dans l'analyse des expériences de dissemblance. Ce type de modèle, par rapport au modèle euclidien classique, permet d'ajouter dans l'algorithme MDS des poids ω_{nr} ; ceux-ci permettent de modéliser l'importance particulière que le sujet n accorde à la dimension r de l'espace perceptif représentant

les dissemblances d_{ijn} entre les stimuli i et j :
$$d_{ijn} = \sqrt{\sum_{r=0}^R \omega_{nr} (\delta_{ir} - \delta_{jr})^2} \quad (\text{Eq. 44})$$

On présente sur la figure 112 l'espace perceptif 2D issu de la MDS INDSCAL obtenu à partir de l'ensemble des matrices de dissemblances des musiciens.

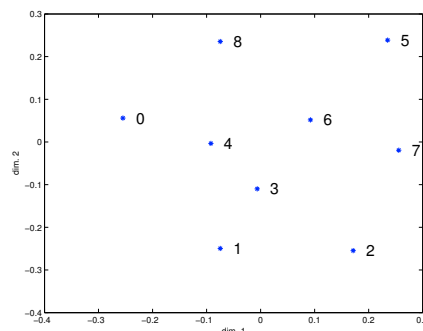


Figure 112. MDS INDSCAL de l'ensemble des matrices de dissemblances des 7 sujets. Les numéros des configurations sont identiques à ceux de la figure 111. Les paires les plus éloignées sont celles qui ont recueilli en moyenne les plus grands jugements de dissemblance de la part des musiciens.

En raison du faible effectif de musiciens recrutés pour l'expérience, il s'est révélé impossible d'associer les dimensions perceptives de cet espace avec des attributs sémantiques déclarés par les musiciens. En revanche, cette MDS a permis d'identifier par la suite les paires de configurations jugées, en moyenne, plus dissemblables que les autres. Ces paires (voir § III.5.3.5.2.2) ont ensuite été utilisées comme données pour l'analyse verbale.

III.5.3.5.2 Analyse des données verbales

III.5.3.5.2.1 Champs lexicaux

Tous les musiciens se sont naturellement placés dans une optique de comparaison des configurations. Les verbalisations enregistrées ont donc naturellement convergé vers des formules du type "La configuration A est plus ... que la configuration B." Il a donc suffi de faire une analyse lexicale pour faire ressortir les aspects musicaux qui avaient été perçus. L'analyse lexicale révèle que les verbalisations tournent autour de trois grands aspects musicaux : l'équilibre des harmoniques (à travers des mots comme "riche", "plein", "métallique", "nasillard", "timbré", "brillant", ...), l'ouverture (au sens de la directivité, comme les termes "ouvert", "fermé", "centré"), et la facilité à l'émission ou au développement (qualifiée par les mots "facile", "dur", "résiste", "développe", ...).

III.5.3.5.2.2 Consensus des musiciens

On s'est intéressé en premier lieu au consensus qui pouvait exister dans les verbalisations des musiciens, pour des paires de configurations données. Les résultats des jugements de dissemblance ayant montré que beaucoup d'entre eux sont faibles, certaines verbalisations peuvent ne pas être considérées comme pertinentes, ou être très variables d'un musicien à l'autre. C'est pourquoi ce consensus a été analysé uniquement sur des paires qui, en moyenne, ont été jugées très dissemblables par les musiciens. Les paires qui ont recueilli les plus grands jugements de dissemblances ont été déterminées à l'aide de la MDS INDSCAL, en prenant les 5 paires les plus éloignées sur la figure 112. Ces paires sont répertoriées sur la figure 111 par des traits continus (0-7, 0-5, 1-5, 2-5, 2-8).

Pour ces 5 paires, les termes verbaux utilisés dans les comparaisons verbales ont été relevés et classés dans les trois catégories relevées ci-dessus (équilibre des harmoniques, ouverture, facilité). Pour la paire constituée des configurations n°1 et n°5, on relève par exemple : "5 est plus timbrée, plus de couleurs que 1", "5 est plus ronde, plus centrée", etc. Au final, concernant cette paire, 43% des instrumentistes perçoivent un changement dans l'équilibre des harmoniques, tandis que 43% perçoivent une facilité de développement ou d'accès aux partiels modifiée. Cette dichotomie dans les jugements se retrouve pour les 5 paires, c'est-à-dire qu'il n'y a pas consensus complet, même sur des paires très dissemblables. Cela peut provenir de stratégies de jugement différentes, ou tout simplement de spécificités d'attention qui diffèrent d'un sujet à l'autre. En général, au moins 40% des sujets se focalisent sur un des trois aspects, tandis que l'autre partie se divise sur les deux autres.

III.5.3.5.3 Corrélation des verbalisations et des paramètres physiques

La deuxième partie de l'analyse a consisté à regarder les corrélations existantes entre les paramètres physiques du système de contrôle (gain et phase) et les verbalisations recueillies.

III.5.3.5.3.1 Effet de l'augmentation de gain à phase constante

Tout d'abord, on s'est intéressé à l'influence de l'augmentation du gain quand la phase est fixée à $\phi = 0$. On compare ainsi les verbalisations faites sur les paires $G=0/G=1$ et $G=0/G=1.5$, c'est à dire les paires 0-1 et 0-2 (voir figure 111). Les tendances sont résumées sur la figure 113. On observe que l'ajout de gain modifie deux aspects majeurs : un enrichissement en harmoniques, ainsi qu'une diminution de la facilité de jeu. En abscisse sont représentées les deux paires 0-1 et 0-2, chaque barre du diagramme représentant le nombre de personnes ayant évoqué l'aspect musical qui correspond à cette barre (c'est-à-dire harmoniques, facilité ou ouverture). Le signe en ordonnée correspond à l'évolution des aspects musicaux de la configuration sans contrôle (configuration 0) à la configuration avec contrôle (configuration 1 ou 2). Si la barre prend des valeurs positives, c'est que l'aspect musical a augmenté en passant de la configuration sans contrôle à la configuration avec contrôle. Si la barre est dans les valeurs négatives, c'est que cet aspect a diminué. Un exemple : pour la paire 0-1, 2 musiciens ont trouvé que la richesse en harmoniques a augmenté avec le contrôle, tandis que 3 musiciens ont trouvé que la facilité a diminué.

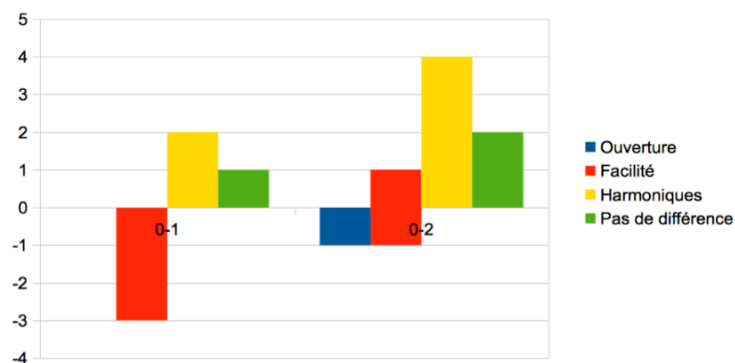


Figure 113. Aspects musicaux (harmoniques, facilité, ouverture) jugés avec variation de gain pour $\phi = 0$.
En abscisse, la paire comparée (voir § III.5.3 4.1) :
la configuration 1 correspond à $G=1$, la configuration 2 correspond à $G=1.5$.

On remarque que les deux aspects majoritairement évoqués sont la facilité et l'enrichissement en harmoniques : il est net que plus le gain augmente, plus l'enrichissement en harmoniques devient évident pour près de 50% des musiciens interrogés ; ces verbalisations qui concernent l'enrichissement en harmoniques dans la configuration n°2 ($G=2$) sont de plus associées à des jugements de dissemblances élevés (> 0.7) ce qui témoigne d'une modification quasi-évidente selon eux. L'aspect de la facilité, évoqué dans le cas de la configuration n°1, et bien qu'associé à des jugements de dissemblance plus faible, est aussi intéressant. Cet aspect semble disparaître avec l'augmentation du gain, cela étant peut être dû à l'augmentation de l'effet des harmoniques au détriment de l'effet de la facilité.

La même analyse est réalisée pour la valeur $\phi = \pi$. On compare ainsi la configuration sans contrôle avec les configurations n° 3, 4 et 5 afin d'évaluer l'influence du gain sur les verbalisations qui ressortent de ces configurations. Ceci est résumé sur la figure 114.

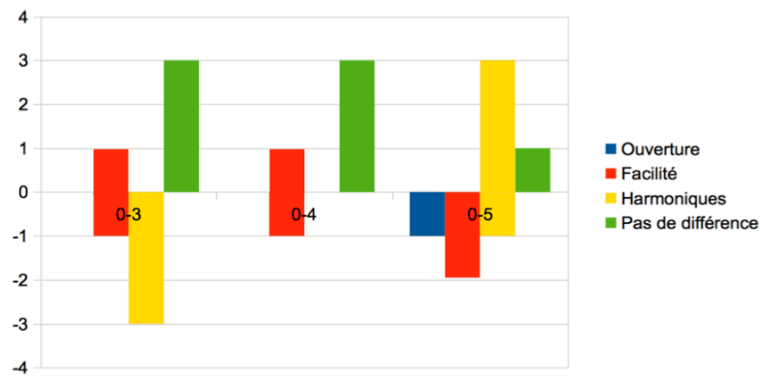


Figure 114. Evolution des aspects musicaux lors de l'augmentation du gain, à phase constante $\phi = \pi$. On compare la configuration 0 (sans contrôle) avec la configuration n°3 ($G=1$), n°4 ($G=2$) et n°5 ($G=3$).

Là aussi, on observe que les deux aspects majoritairement évoqués sont la facilité et l'enrichissement en harmoniques. Les verbalisations concernant les paires 0-3 et 0-4 ($G=1$ et $G=2$) ne semblent pas forcément pertinentes : le nombre de musiciens n'ayant pas perçu de différence s'élève à presque 50%. En revanche, les différences s'avèrent plus nettes pour la paire 0-5 ($G=3$) : de façon générale, la configuration n°5 est perçue plus claire que la configuration sans contrôle.

III.5.3.5.3.2 Effet de la phase à gain maximal

En dernier lieu, l'étude de l'effet de la phase à gain maximal a été étudiée. En effet, après l'étude des corrélations entre verbalisation et augmentation du gain à phase constante, il aurait paru normal d'étudier ensuite les corrélations entre verbalisations et augmentation de la phase à gain constant. Cependant, cela nous semble plus pertinent de regarder l'effet de la phase de manière globale : pour chaque valeur de phase ($\phi = 0$, $\phi = \pi$, $\phi = \pi/2$, $\phi = 3\pi/2$) il s'agit de considérer uniquement la configuration de l'espace des paramètres qui a le gain le plus grand, afin d'étudier l'effet de la phase dans son effet maximal, et de pouvoir ainsi considérer les verbalisations comme "vraisemblables". Les comparaisons des paires 0-2 (phase de 0, gain maximal $G=1.5$), 0-5 (phase de π , gain maximal $G=3$), 0-6 (phase de $\pi/2$, gain maximal $G=1$) et 0-8 (phase de $3\pi/2$, gain maximal $G=1.5$) ont été effectuées. Les verbalisations qui leur sont associées sont résumées sur la figure 115.

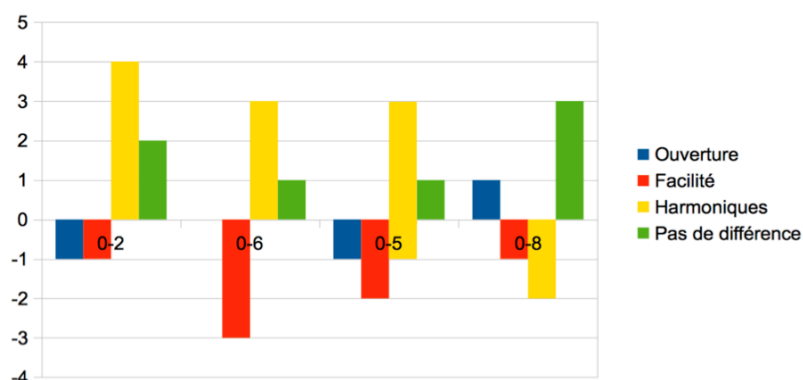


Figure 115. Effet de la phase pour les configurations à gain maximal, en comparaison, avec la configuration sans contrôle.

Configuration n°2 : $\phi=0$ et $G=1.5$. Configuration n°6 : $\phi=\pi/2$ et $G=1$. Configuration n°5 : $\phi=\pi$ et $G=3$. Configuration n°8 : $\phi=3\pi/2$ et $G=1.5$.

On peut noter que : 1/ amener la phase à $\pi/2$ (configuration n°6) entraîne une diminution de la facilité de jeu, ainsi qu'une augmentation de la richesse en harmoniques. La configuration n°6 paraît en effet plus claire que la configuration sans contrôle ; elle est par ailleurs plus difficile à contrôler (multiphoniques, amplification etc.) ; 2/ amener la phase à π (configuration n°5) entraîne aussi une augmentation de la clarté et une diminution dans la facilité de jeu (un musicien note par ailleurs une diminution dans l'ouverture) ; 3/ amener la phase à zéro (configuration n°2) entraîne principalement une augmentation de la richesse en harmoniques ; 4/ amener la phase à $3\pi/2$ (configuration n°8) n'entraîne pas d'effet notable (près de 50% des musiciens interrogés n'ont pas décelé de différence). Cela est peut être dû à la valeur peu élevée du gain ($G=1.5$), limité du fait des phénomènes de *larsen*.

III.5.3 6. Conclusion

Ainsi, nous avons développé un protocole original, permettant de quantifier et qualifier l'effet du contrôle actif sur les dissemblances perçues par un musicien. L'analyse de paires de configurations jugées très différentes (en moyenne) ont permis de dégager trois aspects perceptifs utilisés par les musiciens pour qualifier les différences perçues : richesse (au sens large), ouverture (au sens directivité du son), et facilité (à l'émission ou au développement du son). De plus, il semble bien exister une corrélation entre les modifications physiques apportées par le système de contrôle, et les différences effectivement ressenties par les musiciens. Les analyses de l'effet de la phase à gain maximal et de l'effet du gain à phase constante ont montré que l'augmentation de l'amplitude ainsi que les décalages en fréquence de la fonction de transfert du système étaient corrélés perceptivement et respectivement à des enrichissements harmoniques et à une modification de la facilité de jeu – pour peu que le gain soit assez grand. Cependant, nous restons conscients que l'étendue et l'homogénéité de l'espace des paramètres utilisé a dû répondre à un certain nombre de contraintes, dont celle du *larsen*, dû au retard du système de contrôle numérique qui pourrait être amélioré. Enfin, le nombre de musiciens pour cette étude était relativement restreint, et la possibilité de mener une étude à plus grande échelle permettrait peut être de révéler, à terme, des aspects musicaux communs sous-jacents aux stratégies de jugement des musiciens.

III.6 Timbre des sons environnementaux : meta-analyse

III.6.1 Préambule

Dans la continuité des études sur le timbre musical, et dans l'optique de tendre vers une plus grande "universalité" de la modélisation de la perception des sons environnementaux, nous avons effectué une méta-analyse de cette notion de timbre à partir de plusieurs études en qualité sonore réalisées dans différents domaines industriels.

Conformément aux postulats de McAdams sur le timbre musical (voir § III.2.2), cette analyse révèle une structure à deux niveaux : catégorielle, établissant des classes de sons de même nature, et continue, mettant en évidence des continuum perceptifs à l'intérieur de chaque classe. Cela étant, certaines dimensions perceptives – notamment la brillance – s'avèrent être communes à toutes les classes moyennant quelques éléments de définition spécifiques.

Ce travail s'est inscrit dans le cadre d'une contribution au projet de recherche

SampleORchestrator³² (ANR, programme RIAM 2006) [Vinet et al., 2011] et a fait l'objet d'une publication dans le *Journal of Audio, Speech and Music Processing* (EURASIP JASMP) [Misdariis et al., 2010b].

III.6.2 Publication [Misdariis et al., Journ. Audio, Speech and Music Process., 2010b]

Environmental sound perception: meta-description and modeling based on independent primary studies

Nicolas Misdariis ⁽¹⁾, Antoine Minard ⁽¹⁾, Patrick Susini ⁽¹⁾, Guillaume Lemaitre ⁽¹⁾, Stephen McAdams ⁽²⁾, Etienne Parizet ⁽³⁾

⁽¹⁾ STMS-Ircam-CNRS, 1 place Igor Stravinsky F-75004 Paris, France

⁽²⁾ CIRMMT, Schulich School of Music, McGill University, 555 Sherbrooke St. W. Montreal, QC, Canada H3A 1E3

⁽³⁾ LVA-Insa Lyon, 25bis Avenue Jean Capelle F69621 Villeurbanne Cedex, France

III.6.2.1 Introduction

The purpose of this study is to transpose and extend the timbre description principles of musical sounds to environmental sounds considered, by nature, as non-musical. More precisely, environmental sounds were first defined by Vanderveer [Vanderveer, 1979] as "... any possible audible acoustic event which is caused by motions in the ordinary human environment. (...) Besides 1) having real events as their sources (...), 2) [they] are usually more 'complex' than laboratory sinusoids, (...), 3) [they] are meaningful, in the sense that they specify events in the environment. (...), 4) the sounds to be considered are not part of a communication system, or communication sounds, they are taken in their literal rather than signal or symbolic interpretation."

Within the restricted framework given by the scope of the primary research upon which the present study is based (see § III.6.2.2), the final aim is also to automate indexing and classification of environmental sounds. This goal is actually essential for sound quality measurements, as well as for further sound-content-based searching and browsing methods that use perceptual models of environmental sounds and often require measurements based on perceptually relevant acoustical similarities. Indeed, in the sound-quality field, most studies use acoustical/psychoacoustic descriptors such as loudness or roughness, in order to explain unpleasantness ratings, whereas several studies have shown that no "universal" descriptors exist for all classes of everyday sounds.

The work detailed in this article starts from four primary industrial studies on sound attributes dealing with sounds produced by car engines ([Susini et al., 1997a ; 1997b, 1999], [McAdams et al., 1998]), air-conditioning units ([Susini et al., 2004]), car horns ([Lemaitre et al. 2007; 2009a) and closing car doors ([Parizet et al., 2008]). The aim of these studies was to apply the methodology developed to study the timbre of musical sounds to a specific category of environmental sounds. The standard methodology used in these studies was based on a multidimensional scaling technique (MDS) applied to dissimilarity judgments.

The MDS technique is a fruitful tool for studying perceptual relationships among sounds and for determining the underlying auditory attributes used by participants to rate the perceived similarity among sounds. The term auditory attribute is used to

³² <http://anasynth.ircam.fr/home/projects/sample-orchestrator>

describe the perceived properties or qualities of the sounds. Well-known auditory attributes include loudness, pitch, duration, sharpness, etc. The MDS technique does not require a priori assumptions concerning the number of auditory attributes or their nature, unlike semantic differential methods that use ratings along specific dimensions, such as roughness, for example. The MDS technique represents the perceived similarities in a low-dimensional Euclidean space (referred to as the *perceptual space*), so that the distances among the stimuli reflect the perceived dissimilarities. Each dimension of the space (called a *perceptual dimension*) is assumed to correspond to a perceptual continuum that is common to the whole set of sounds. Thus the main hypothesis with the MDS technique is that the sounds under study can be compared on auditory attributes that are shared by all sounds in the corpus. In other words, this technique is appropriate for characterizing sounds that are comparable along continuous auditory attributes of a homogenous corpus of sounds, i.e. composed of sounds produced by the same type of source (musical instruments, car sounds, vacuum cleaner noises, etc.). Considering musical sounds, the most common timbre space found by several studies (among which [Grey, 1977], [Krumhansl, 1989], [McAdams et al., 1995] and [Marozeau et al., 2003]) consisted of three dimensions correlated with acoustic features in order to associate a measurable sound parameter with each perceptual dimension of timbre. The assumption of this approach rests on the model suggested by McAdams [McAdams, 1993], who postulates that the recognition of the sound sources arises from a process of analysis, computation and extraction of a certain number of auditory features related to the acoustic parameters of the signals. Then, in many of these musical timbre studies, the three dimensions were found to be significantly correlated with a spectral feature that most often represented auditory brightness (energy distribution along the frequency scale), a temporal feature that characterized attack and a spectro-temporal feature corresponding to spectral variations over time. The MDS technique has been shown to be an efficient tool for revealing and describing the unknown auditory attributes underlying the timbre of musical sounds.

In the present context, environmental sound studies, experimental data, analyses and acoustic parameters have been reviewed and compared from the four initial studies. An investigation of these combined data was conducted, and an attempt to model the resulting structures on the basis of the primary results was made using generalized toolboxes (essentially, "Ircamdescriptor" from Peeters [Peeters, 2004] and "Auditory Toolbox" from Slaney [Slaney, 1998]) in order to unify – and in some cases to improve – the description of the initial data. Here we will first introduce and describe all the studies taken into account in this review, their stimulus sets, the experiments performed, the resulting perceptual spaces and the correlated acoustic features. Then, in order to contribute to environmental sound perception, we will first present this overall stimulus set organization in terms of the main environmental sound classes, propose both inter-class and intra-class structure descriptions, and finally initiate an automatic classification modeling approach within the restricted scope of the present study but on the basis of perceptually relevant data and results gathered during its experimental parts.

III.6.2.2 Primary Studies

We present in this section the frameworks, motivations and results of the four experimental studies that represent the starting point of our meta-analysis. These studies focus on the sounds from:

- A.- Car interiors ([Susini et al., 1997a ; 1997b, 1999], [McAdams et al., 1998])
- B.- Interior air-conditioning units ([Susini et al., 2004])
- C.- Car horns [Lemaitre et al. 2007; 2009a]
- D.- Closing car doors ([Parizet et al., 2008])

These four studies all addressed the issue of sound quality and shared a common approach: they studied the timbre of the different types of sounds. More precisely, they use a common methodology and share similar analysis techniques. This procedure relies on the psychoacoustic definition of timbre: "Timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar." (American national standard acoustical terminology" (1994). American National Standards Institute, ANSI S1.1-1994 (R1999); see also [Krumhansl, 1989]). Timbre is thought to be multidimensional, encompassing several perceptual attributes that are collectively referred to by this term. In order to uncover the attributes of timbre, the methodology used in the studies was based on the procedure developed to study the timbre of musical sounds ([McAdams et al., 1995]). It has three main steps, the first one being sometimes preceded by a preliminary step (labeled "0" below) used to reduce the number of sounds to be tested in the first step:

0. Because the following step of the methodology needs a small number of sounds to be experimentally feasible, a preliminary step is sometimes used in order to reduce the original corpus to an acceptable number of stimuli (usually not more than 20 samples). Free-sorting tasks and cluster analyses (see § III.6.2.3.1 for further details) are used to attain this goal. A free-sorting task consists in asking participants to sort the sounds of the set into as many categories as they wish. Thus, they identify the main categories of sounds that are studied and allow for the selection of representative subsets of sounds by homogeneously sampling across the categories.
1. A dissimilarity rating experiment collects the perceived dissimilarities among the sounds, which are then used as proximity data. It consists in asking the participants to rate directly the dissimilarity between both sounds of each possible pair within the set of sounds. The evaluation is made on a continuous scale labelled "Very Similar" at the left end and "Very Dissimilar" at the right end. It has the great advantage that it does not impose any predefined rating criteria on the listener.
2. The proximity data are modeled with a multidimensional scaling (MDS) analysis that fits distances in a geometrical space to the dissimilarity data. The dimensions of this space represent the perceptual dimensions underlying the proximities. Different levels of complexity exist in the MDS approach depending on the model and associated algorithm (see § III.3.1); in the present case, two particular MDS techniques were used in the studies: the INDSCAL (Individual Difference Scaling) and CLASCAL (Latent Class Approach) models.
3. The final step of a timbre study is to give a physical interpretation of the perceptual dimensions revealed by the MDS analysis. This is usually done by submitting the perceptual dimensions to linear regression analyses with relevant acoustic features. Some of them are psychoacoustic descriptors, i.e. acoustic features that have been found to correspond to auditory sensations. Models that compute psychoacoustic descriptors are usually based on a model of the peripheral auditory system.

III.6.2.2.1 Studies A (A1, A2): car interior

Context. The main goal of this study was to analyze the timbre of the sounds of car interiors in a given driving condition from the driver/passenger point of view.

Stimuli. The sounds were recorded in 16 different vehicles at two different engine modes. The engine modes defined two sub-studies: study A1 involved sounds produced when the engine was running in 3rd gear at 4000 RPM (Round Per Minute) and study A2 involved sounds produced when the engine was running in 5th gear at 3500 RPM. A preliminary experiment showed that loudness was the main auditory cue used by the participants to rate the dissimilarity. Thus, in order to let other auditory attributes emerge, loudness was equalized. Both stimulus sets were composed of 16 stereophonic sounds that were 4.1 seconds in duration. Their levels – after loudness equalization – varied between 69 and 80 dB SPL (Sound Pressure Level).

Participants. For each engine mode stimulus set, a dissimilarity rating experiment was conducted with 30 participants.

Analysis and Results for Study A1. A CLASCAL analysis of the data yielded a 1-latent class, 3-dimensional space with specificities. Figures 6-1 to 6-3 in Appendix 6 represent the projections of the space, and Table 6-1 reports the correlation coefficients of the acoustic features best fitting the perceptual dimensions. The first dimension is correlated [$r(14)=-0.81$, $p<0.01$] with a feature corresponding to the relative balance of the harmonic (motor) and noise (air turbulence) components. The second dimension is correlated [$r(14)=-0.70$, $p<0.01$] with a variation of the spectral centroid with the frequency dimension represented in ERB-rate. The third dimension is significantly correlated [$r(14)=-0.83$, $p<0.01$] with an acoustic feature quantifying the spectral decrease of the harmonic part of the sound (see Appendix 6 for more details).

Analysis and Results for Study A2. A CLASCAL analysis yielded a 1-latent class, 2-dimensional space with specificities. Figure 6-4 in Appendix 6 represents the perceptual space and Table 6-2 reports the correlation coefficients of the acoustic features best fitting the perceptual dimensions; the features that are the best correlated with the two dimensions are also reported in Table 6-2. The first dimension is correlated [$r(12)=0.93$, $p<0.01$] with an acoustic feature conveying the relative balance between two groups of partials of the harmonic part of the signal. The second dimension is correlated [$r(12)=0.86$, $p<0.01$] with the spectral centroid computed on the C-weighted version of the signal (see Appendix 6 for more details).

III.6.2.2.2 Study B: interior air-conditioning units

Context. This study focused on the sound quality of interior air-conditioning units.

Stimuli. The initial set consisted of 43 sounds produced by units of different brands. A free-sorting experiment was first conducted to select an homogeneous subset of sounds representative of the existing range for this type of sounds. The results of this experiment also showed that three categories were made mainly by grouping together sounds with similar loudness levels. As in study A, in order to prevent loudness from dominating the ratings (possibly masking more subtle effects), the sounds were selected in the category corresponding to a medium loudness level (average level: 46.5 dB SPL, 2.2 dB standard deviation). An informal experiment was then performed with only 5 participants to get an initial estimate of the perceptual space structure. The outcome of the MDS analysis was that the space was not homogeneously sampled. Therefore synthesized sounds were added and redundant sounds were removed in order to produce a more homogeneously distributed stimulus set. The synthesized sounds were

created on the basis of features of the sounds in the stimulus set, using a geometric interpolation within the space. The resulting stimulus set consisted of 19 sounds: 15 recordings of air-conditioning units and 4 synthesized sounds. They were all 5.9 seconds in duration with levels varying between 44 and 52 dB SPL.

Participants. The dissimilarity rating experiment was conducted with 50 participants.

Analysis and Results. A CLASCAL analysis of the dissimilarity ratings yielded a 5-latent class, 3-dimensional space with specificities. Figures 6-5 to 6-7 in Appendix 6 represent the projections of the 3-dimensional space, and Table 6-3 presents the correlation coefficients of the features that are the best correlated with the perceptual dimensions. The first dimension is correlated [$r(17)=-0.97$, $p<0.01$] with a feature corresponding to the relative balance of the harmonic (motor) and noise (air turbulence) components. The second dimension is correlated with a frequency-weighted variation of the spectral centroid of the noise component [$r(17)=0.73$, $p<0.01$]. The third dimension is correlated with loudness [$r(17)=0.84$, $p<0.01$]. Indeed, even though the selected sounds are in the same range of loudness, they were not equalized in loudness (see Appendix 6 for more details).

III.6.2.2.3 Study C: car horns

Context. This study concerned the timbre of car horns in order to define specifications for the design of new sounds. The initial stimulus set consisted of 43 recordings of current car horn sounds. These sounds can be either monophonic (one note) or polyphonic (two or three notes to make a chord) and are produced by two different mechanisms: a metal plate or a folded horn that acts as a resonator and is attached to the membrane of an electroacoustic driver. Both produce very specific timbres. A preliminary sorting experiment highlighted 9 main categories of sounds connected with these different mechanisms and properties.

Stimuli. A sample of 22 sounds was chosen among the 9 categories. Among these 22 sounds, 13 were monophonic and 9 were polyphonic, 10 were produced by "plate" resonators and 12 by "horn" resonators. All sounds lasted between 0.6 and 2.2 seconds. Their levels varied between 63 and 77 dB SPL.

Participants. A dissimilarity rating experiment using this set of sounds was conducted with 41 participants.

Analysis and Results. The dissimilarity ratings were submitted to a CLASCAL analysis, resulting in a 6-latent class, 3-dimensional space with specificities. Figures 6-8 to 6-10 in Appendix 6 represent the projections of the 3-dimensional space, and Table 6-4 reports the best-correlated features (see [Lemaitre et al., 2007], for more details on the acoustic features). The first dimension is correlated [$r(20)=-0.9$, $p<0.01$] with roughness. The second dimension is correlated [$r(20)=0.9$, $p<0.01$] with a variation of the spectral centroid integrating a perceptual approach to compute this parameter (ERB scale, see [Marozeau et al., 2003]). The third dimension is correlated [$r(20)=-0.8$, $p<0.01$] with an acoustic feature related to the fine structure of the spectral envelope.

III.6.2.2.4 Study D: car door closing

Context. The main goal was to study the timbre of car door closing sounds in the context of evaluating their sound quality.

Stimuli. An initial set of 27 stereophonic recordings (16 sounds from different cars and 11 sounds from two cars with modified seals) was submitted to a sorting experiment with 31 participants in order to select a representative subset of 12 sounds. Among

these 12 sounds, 4 were recorded from cars with modified seals. The durations of the sounds varied between 0.3 and 0.5 seconds, and their levels varied between 66 and 84 dB SPL.

Participants. A dissimilarity rating experiment was conducted with 40 participants.

Analysis and Results. The dissimilarity data were submitted to an INDSCAL analysis. A 3-dimensional space was found. Figures 6-11 to 6-13 in Appendix 6 represent its projections, and Table 6-5 reports the correlation coefficients of the features best correlated with the perceptual dimensions. The first dimension is correlated with a feature corresponding to sharpness, as defined by Aures [36] [$r(10)=-0.90$, $p<0.01$], as well as to the spectral centroid [$r(10)=-0.93$, $p<0.01$]. The second dimension is correlated [$r(10)=0.87$, $p<0.01$] with an indicator related to the temporal evolution of instantaneous loudness, according to Zwicker's model [Zwicker, 1977]. No descriptor was found that correlated significantly with the third dimension.

III.6.2.2.5 Comparisons and discussion

The studies reported in the previous subsections identify the perceptual space of sounds contained in five separate stimulus sets (labelled A1, A2, B, C and D), associated with different kinds of environmental situations, mainly related to car and appliance industries. The results of these studies are summarized in Table 7 below.

	A- Car interior	B- Air-conditionning units	C- Car horns	D- Car door closing
Corpus	A1: 16 snds 3 rd gear, 4000 rpm A2: 14 snds, 5 th gear, 3500 rpm	19 sounds (4 synthesized)	22 sounds	12 sounds
Analysis	CLASCAL	CLASCAL	CLASCAL	INDSCAL
Results	A1: 3 dim., specif., 1 lat. class. A2: 2 dim., specif., 1 lat. class.	3 dim., specif., 5 lat. class.	3 dim., specif., 6 lat. class.	3 dim.
Descriptors	A1, dim.1: RAPmv-A A1, dim.2: CGg-ERB A1, dim.3: Dec A2, dim.1: rad_2N/0.5N A2, dim.2: CGg-C	dim.1: NHR-A dim.2: Sc _n -B dim.3: N (Loudness)	dim.1: Roughness dim.2: Spectral centroid dim.3: Spectral deviation	dim.1: Spectral centroid dim.2: Cleanness indicator dim.3: ...

Table 7. Table recapitulating methodological context and main results for studies A1, A2, B, C, and D (see corresponding sections above and Appendix 6 for further details)

A comparison of the acoustic features correlated with the dimensions of these four perceptual spaces yields some interesting facts:

- In every study, at least one dimension in the perceptual space resulting from the MDS analysis is found to be related to a spectral centroid feature, usually describing the "brightness" of a sound. This "brightness" attribute seems therefore incontrovertible when trying to compare two sounds of any of these kinds of sources. However, this attribute seems to take different forms across the studies:
 - It can be computed with a frequency weighting representing the variation in sensitivity of the human ear over the audible frequency range at different presentation levels (A-, B- and C-weightings).
 - It can introduce a much more sophisticated model of the hearing process (ERB filters).
 - It is sometimes only computed on a particular part of the signal (noise part).

The subsidiary questions are: will all these "brightness" predictors be as efficient for all studies? If not, is there a particular calculation that fits all of the spaces equally well?

- In 3 studies, relevant acoustic features appear to include separate calculations for the harmonic and noise parts of the signals. The signal processing needed for this separation is quite complex and often includes the setting of several initial algorithm parameters. Again, the question raised is: will a common set of these parameters result in the same efficiency of separation for the correlation scores in the 3 studies ?
- For the 2 other studies, the correlation results exhibit specific relevant acoustic features. This fact confirms that a universal low-dimensional perceptual space describing all sounds does not exist. It would also tend to agree with [McAdams et al, 1998] and [McAdams, 1993] who observed that when sounds are produced by too different kinds of sources, the dissimilarity judgments may be based on cognitive factors rather than on perceptual signal-related ones, which results in a strongly categorical description.

III.6.2.3 Meta-processing: complementary experimental investigations

The MDS technique is appropriate to characterize a set of sounds caused by very similar sound sources, but not for different and obviously identified sources. For instance, McAdams et al [McAdams et al., 1998] applied an MDS analysis to an extremely heterogeneous set of environmental sounds (trains, cars and planes). The analysis yielded a strongly categorical perceptual structure: listeners identified the sound sources rather than comparing them along continuous perceptual dimensions. In that case, participants based their perception on a predominant cognitive factor: recognition, classification, and identification of the sound source (see [McAdams, 1993]). In other words, when the sounds under consideration are similar, which means that they are provided by the same type of sources, listeners are able to compare them on continuous perceptual dimensions, otherwise they are categorized by association with the type of source. As a consequence, the perceptual organization of the five groups of stimuli may be based on a 2-level structure displaying both categorical and continuous levels (see figure 116, for illustration):

- a categorical (discrete) level corresponding to the main sound event categories, each of them being related to a distinct physical cause or source;
- a continuous level that will associate each of these categories with a perceptual space with salient dimensions that can be either specific to the given category or shared with the others.

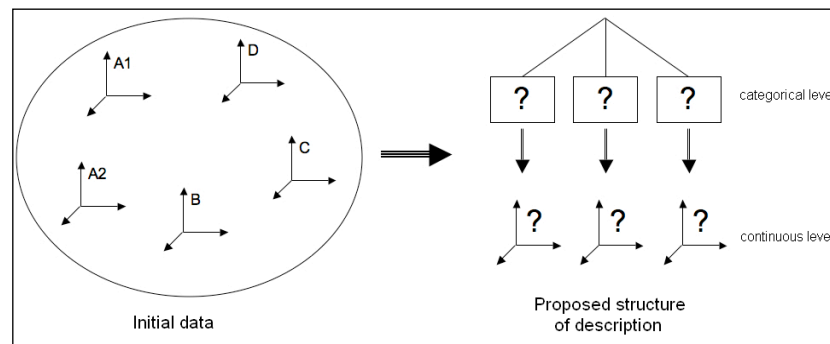


Figure 116. Schematic representation of the proposed 2-level organisation structure.

In order to evaluate the consistency of this structure and to validate it within the scope of the present research, an additional experimental investigation was conducted consisting of two successive experiments:

1. a free-sorting task to identify the main sound event categories composing the overall sound corpus, combined from the initial corpora presented in § III.6.2.2;
2. a forced-choice sorting task on more heterogeneous sounds (extracted from commercial sound libraries) in order to extend and determine more precisely the boundaries of these categories.

The results of these experiments will then be used in the last part of the study (see § III.6.2.4) in order to define new ways of modeling the structure on both discrete and continuous levels, as defined in our hypothesis, that describes the main sound event categories and perceptual dimensions attached to each of those categories, respectively.

III.6.2.3.1 Experiment 1: Free sorting task on the initial corpora

In order to identify the main perceptual categories among the sounds under consideration in this study, a free-sorting experiment with this complete stimulus set was conducted.

III.6.2.3.1.1 Method

Participants. Twenty participants (8 women and 12 men) volunteered as listeners for this experiment and were paid for their participation. All reported having normal hearing.

Stimuli. The resulting unified stimulus set is a collection of 83 sounds distributed as follows: 16, 14, 19, 22 and 12 sounds from studies A1, A2, B, C and D, respectively. See the *Stimuli* subsections of § III.6.2.2 for more details on the sounds. In order to prevent the listeners from sorting the sounds according to their loudnesses, a preliminary loudness-equalization experiment was conducted with 7 participants working at IRCAM, resulting in an 83-sound loudness-equalized corpus.

Apparatus. Testing took place in a double-walled IAC sound-isolation booth. The sounds were played over Sennheiser HD 520 II headphones through a RME Fireface 400 audio card plugged into a Macintosh Mac Pro (Mac OS X v10.4 Tiger) workstation. The test was run using a Graphical User Interface (GUI) specifically developed in Matlab (v7.0.4) including stimulus control, data recording and sound play-back³³.

Procedure. At the beginning of the procedure the participants were given written instructions briefly presenting the context of the study and detailing the task to be performed. The task was to classify the 83 sounds of the corpus in as many categories as they wished according to their own criteria and, in a second step, to select the most representative sound – the prototype – for each of the classes (see § III.6.2.3.2 for the definition of a prototype by Rosch [Rosch, 1978]). In the GUI, the sounds were represented as dots that could be either played (double-click) or moved (drag and drop) in the dedicated area of the screen in order to graphically compose the categories (see figure 7-1 in Appendix 7 for an illustration of the interface).

III.6.2.3.1.2 Results

Analysis. The experimental data consist in individual incidence matrices (coding the set partitions of each subject) that are summed to form a co-occurrence matrix. The co-

³³ This GUI was developed by Vincent Rioux.

occurrence matrix represents how many subjects have placed each pair of sounds in the same category. This can also be interpreted as a proximity matrix ([Kruskal et al., 1978]). In the present case, we derived a hierarchical tree representation from these data using an unweighted arithmetic average clustering (UPGMA) analysis algorithm (see [Legendre et al., 1998] for computational details). In such a representation, the distance between two sounds is represented by the height of the node which links them. Among the 91,881 triplets that can be formed out of 83 sounds, 94% follow the ultrametric inequality, which shows the adequacy of the tree representation for these data. The tree representation is shown in figure 117. It can be clearly seen that 3 main categories constitute the unified corpus. Looking in detail at the items inside each of them, we can observe that these 3 categories correspond respectively to studies A and B (right part of figure 117), study C (left part of figure 117) and study D (middle part of figure 117). Moreover, listening to these items led us to propose a semantic labeling for each of these 3 categories: "motor", "instrument-like", and "impact", respectively.

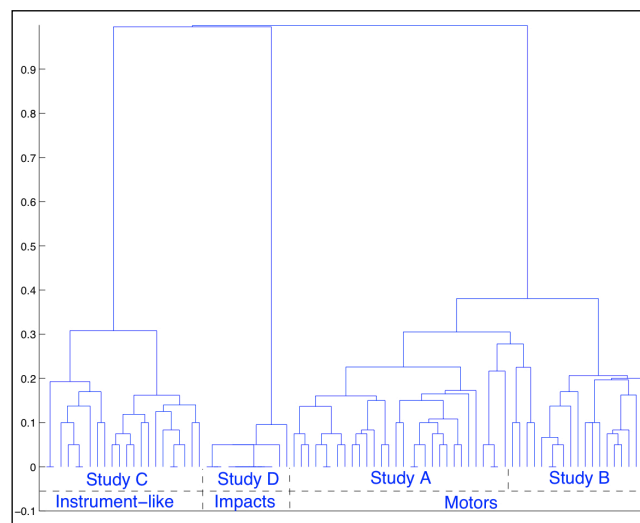


Figure 117. Experiment 1, dendrogram resulting from the cluster analysis.
Representation of the 3 main categories: *motor* (right), *instrument-like* (left) and *impact* (middle).

We subsequently extracted the prototypic sound for each of the 3 categories by a specifically developed algorithm. Each listener selected prototypes with regard to her/his own categories, which are not necessarily the 3 categories extracted from the cluster analysis. Consequently, we had to consider the prototype selection for each pair of sounds, as follows:

- In an individual 83 X 83 matrix, for each pair of sounds, if the sound j was selected by a listener as prototype of a category that contains a sound i , then the cell (i,j) is incremented (but not the cell (j,i)).
- After summing the matrix over the panel of listeners, a sub-matrix is extracted, for each of the 3 final categories with the rows and lines indexed by the sounds constituting the category.
- Each obtained sub-matrix is averaged over its rows, and the highest score gives the index of the prototype.

With this method, the selection by a listener of a sound as prototype for sounds that do not belong to the same final category does not influence the final selection of prototypes. The 3 selected sounds will be used in Experiment 2 (§ III.6.2.3.2) as a definition of the 3 categories.

Discussion. As a result, the five initial corpora can be reorganized into 3 main categories on the basis of the perceptual results (Experiment 1). Obviously, these categories are strongly defined by the initial studies from which they were drawn. In other words, there is no overlap between the initial corpora and the final structure: studies A1, A2, B belong to a first category, study C to a second category and study D to a third one. However, this fact was intuited before the experiment just by listening to the sounds. According to the sound production type, we semantically defined these categories as:

- "**motor**" (first category): sounds from both car interior (studies A1 and A2) and air-conditioning unit (study B) corpora. These sounds have two discriminable components: a harmonic part with a quite low fundamental frequency produced by a "motor" and a noisy part produced by air turbulence.
- "**instrument-like**" (second category): sounds that correspond to the car horn corpus (study C), which are defined by one or several higher tones, closer to those produced by musical instruments than those generated by motors.
- "**impact**" (third category): sounds of the car door closing corpus (study D). Actually, one can easily discriminate these sounds from the others because of their temporal structure. This idea is consistent with the discrimination of percussive and sustained sounds among musical timbres. Indeed, impact sounds of the environment are quite close to musical percussive sounds in terms of sound production.

This categorization is consistent with the product sound classification proposed by Özcan et al. [Ozcan et al., 2007] defined by 6 sound categories: *air*, *alarm*, *cyclic*, *impact*, *liquid* and *mechanical*. Even though these product sounds were from a domestic context, Özcan et al. found an **impact** category; they also found an *alarm* category that can correspond to the present **instrument-like** category with regard to basic similarities in pitch, harmonic structure or stationary aspects of the sounds; and finally, the present **motor** category can be linked to both their *air* and *mechanical* categories.

III.6.2.3.2 Experiment 2: Forced-choice sorting task on an extended corpus

On the basis of the previous results (3 main categories of environmental sounds within the scope of the unified corpus, with an associated prototype for each), a second experiment was conducted in order to generate a more heterogeneous corpus that would better represent the range of variation of each category. This was done by means of a forced-choice procedure, the main choices being the categories found in Experiment 1. These 3 categories were each identified by their respective prototypic sound extracted from Experiment 1, instead of being verbally defined as it is usually done in this kind of procedure. The notion of prototype is based on psychological principles related to the way one organizes knowledge of the surrounding world. For Rosch [Rosch, 1978], a prototype is the element of a group that is the most similar to all items inside the group and, at the same time, that is on average the most different from all items of all the other groups. The notion of prototype used in the present study is directly derived from Rosch's concept.

Furthermore, the outcome of this second experiment will also provide perceptually validated data for the modeling part of the present study in order to implement an automatic classifier (§ III.6.2.4.2).

III.6.2.3.2.1 Method

Participants. Twenty-one participants (8 women and 13 men) volunteered as listeners for this experiment and were paid for their participation. All reported having normal hearing.

Stimuli. A new extended corpus was created on the basis of the main categories found in the previous experiment. Several sounds were added to each category in order to make the stimulus set more complete and heterogeneous. We therefore chose various new sounds with quite extreme cases for each category from commercial sound libraries (Hollywood Edge Premiere Edition I, II and III, Sound Ideas General Series 6000 and Blue Box Audio Wav). Here are some examples of sounds added in each category:

- for *motor* sounds: truck, aircraft, motorbike, helicopter, crane, vacuum cleaner, fridge, blender, electric shaver, lawn mower;
- for *instrument-like* sounds: phone ringing, dishes squeak, door creak, alarm, bell;
- for *impact* sounds: glass shock, various doors closing (fridge door, house door, etc.), computer keyboard, water drop, tennis ball.

Again, the sounds needed to be equalized in loudness so that the judgements would not be based on this auditory attribute. However, considering the high number of sounds, a preliminary experiment of loudness equalization would have been quite long. As a consequence, the sounds' loudnesses were equalized with regard to the value given by the loudness model of Zwicker et al. [Zwicker et al., 1990].

The final corpus was composed of 150 loudness-equalized sounds with an equal distribution of 50 sounds in each category.

Apparatus. The same technical equipment as in Experiment 1 was used. However, the study was run using a GUI specifically developed in the PsiExp v3.4 experimentation environment including stimulus control and data recording ([Smith, 1995]). The sounds were played with Cycling '74's Max/MSP software (v4.6).

Procedure. At the beginning of the experiment, the participants were given written instructions briefly presenting the context of the study and detailing the task to be performed. They were asked to classify the 150 sounds of the new corpus into 3 unnamed categories associated with their respective prototypical sounds by clicking on the corresponding button. A fourth button labeled "other" allowed participants to not choose any of the 3 main categories (see figure 7-2 in Appendix 7 for an illustration of the interface). The specificity of the present paradigm was to make the categories explicit with the prototype sounds found in Experiment 1 – with the obvious exception of the class "other" – instead of naming them directly. This implementation was chosen in order to avoid any ambiguity in the understanding of the arbitrary semantic attributes that did not result from verbalization analyses.

III.6.2.3.2.2 Results

Analysis. Table 8 presents the sound distribution, i.e. mean and standard deviation of the number of sounds placed by the participants in each category. Note that these data are strongly influenced by the choices of sounds added by the experimenter, and that these numbers mainly show the adequacy or inadequacy of these choices. However, the following points may be emphasized:

- The high standard deviation of the number of rejected sounds ("other") might be related to differences in strategy among the participants who did not use the same selectivity threshold, or the same granularity, to group the sounds.

- The high mean number of sounds combined with a relatively low standard deviation for the motors shows a consensus among the participants that proves the adequacy of the selection of sounds for this category.
- On the contrary, the relatively high standard deviations for the two other categories show some variability in the listeners' judgements, which is probably due to the quite large variety of chosen sounds for these categories. For the "instrument-like" category, the variability seems to be related to the difficulty of theoretically defining this type of sound, whereas for the impacts, it could be explained by the too-general character of this category.

Nevertheless, after computing a percentage of belonging to the categories for each sound of the 150-item corpus, we observed that these disparities in classification were concentrated on certain sounds only, which were then rejected (26 sounds under a threshold of about 65% of belonging to a category). We thus obtained a selection of the extended corpus leading to a 124-sound stimulus set: 50 "motor", 27 "instrument-like", 47 "impact". Note that this final distribution corresponds roughly to that of Table 8.

	Prototype #1 (<i>motor</i>)	Prototype #2 (<i>instrum-like</i>)	Prototype #3 (<i>impact</i>)	"other"
mean	48.7	32.6	45.9	22.7
std	5.5	8.3	13.2	15.5

Table 8. Experiment 2 – distribution of the sounds in the experimental categories.

Discussion. The partitioning of the data across the 3 categories shows a good consensus on a certain number of sounds for each class. With this result, we are then able to make a selection of sounds that are clearly associated with one of the 3 categories revealed in Experiment 1. This leads to the constitution of a perceptually validated sound corpus with regard to the *motor*, *instrument-like* and *impact* categories, which is now large and representative enough to consider the conception and validation of a predictive tool for automatic classification of environmental sounds in these three categories.

III.6.2.3.3 Discussion

Within the restricted scope of environmental sounds studied here (industrial sounds from cars and machines), we are now faced with the following structure:

- a *motor* category including 49 sounds from 3 different corpora (A1, A2, B), each of them being described by a perceptual space and augmented with 50 perceptually validated new sounds, for a total of 99 items,
- an *instrument-like* category including 22 sounds from corpus C described by a perceptual space and augmented with 27 perceptually validated new sounds, for a total of 49,
- an *impact* category including 12 sounds from corpus D described by a perceptual space and augmented with 47 perceptually validated new sounds, for a total of 59.

In the next step, this corpus will serve as input for the implementation of the automatic classifier detailed in § III.6.2.4.2.

III.6.2.4 Meta-processing: modeling the description structure

This section was designed to confirm the second part of the starting hypothesis, which stipulates that both inter-category and intra-category properties exist, i.e. dimensions shared by the all categories and specific dimensions related to their mutual discriminating differences. Furthermore, the knowledge of these discriminating features

could facilitate the implementation of a predictive tool capable of automatically recognizing whether a new item belongs to one of the 3 meta-categories. Note that every acoustic feature mentioned in this section is extracted either from the Ircamdescriptor toolbox (CUIDADO project, [Peeters, 2004]) or from the Auditory Toolbox ([Slaney, 1998]).

III.6.2.4.1 Continuous level: unifying the perceptual space dimensions

In this first part, we investigate the shared and specific properties across the categories by considering the data coming from the corpus described by perceptual space dimensions (corpuses A to D). The main idea is to unify these data by recomputing the acoustic features explaining the different perceptual dimensions in a more systematic manner, in order to point out regularities and singularities among the given spaces. The implementation of these acoustic features is detailed in Appendix 8.

Note that some of the stimulus sets contain only monophonic sounds, whereas others contain only stereophonic sounds, and, although the acoustic features are calculated on both channels in the latter case, the salience of an indicator in one channel compared to the other depends on the recording context. For example, if a car interior sound has been recorded from the driver's seat, the most relevant channel for a given sound feature will probably not be the same as if it had been recorded from the passenger's seat. Accordingly, the features in the correlation tables can be either from the left or the right channel, or from the mean of both channels.

III.6.2.4.1.1 MDS analyses compatibility

The two models giving rise to the perceptual spaces that will be unified in this section are INDSCAL and CLASCAL (see § III.3.1). As both models remove the rotational invariance of the obtained spaces, one could assume that both models would result in similar main perceptual dimensions (even if possible slight differences on items' position or axes's orientation may be due to the precision of the model). However the presence of specificities in the latter can modify the psychological meaning of the dimensions. Indeed, the fact that a part of the Euclidean distances is explained by those specificities leads to a modification of the proportion explained by the dimensions. Thus the dimensions obtained by both models will not necessarily be the same.

All the same, the only sound corpus for which the INDSCAL method was used (study D) corresponded to a different sound category than those of the other corpora (see § III.6.2.3.1). As a consequence, the fact that the dimensions were obtained differently from the other studies is not a problem. Indeed, this perceptual space will be studied separately from the others.

III.6.2.4.1.2 Motor category

One of the main characteristic of this kind of sound is that it contains two different simultaneous parts. The first one corresponds to a harmonic pattern that can be easily modeled by a sum of sinusoids, and the second one corresponds to the noise resulting from the air turbulence. Perceptually, these two parts are highly discriminable. Consequently, unlike the other two categories, both parts need to be taken into account independently when estimating the acoustic features. This is the reason why harmonic separation methods were tested and used in order to describe both parts, as well as their mutual interaction.

This meta-category regroups stimulus sets A1, A2 and B, presented in § III.6.2.2.1 and § III.6.2.2.2. Those stimulus sets' MDS analyses resulted in 3-dimensional perceptual

spaces, except for that of study A2, which gave a 2-dimensional space. Because of the relative proximity of the sounds coming from these 3 stimulus sets, two shared dimensions were found. The first one is related to the harmonic/noise ratio, while the second is related to the spectral centroids of both parts with some interactions. Finally, the stimulus sets of studies A1 and B differ in their third dimension (the stimulus set of study A2 only gave a 2-dimensional space), most likely because of a practical particularity of the experimental protocol: the sounds of set A1 were first loudness-equalized, unlike those of set B. The correlation scores between dimensions of the motor sound stimulus sets and the best-fitting acoustic features (see Appendix 8) are presented in Table 9.

	Dimension 1	Dimension 2	Dimension 3
-- study A1 --			
HNR	-0.93**	0.12	-0.22
<i>Complex brightness</i>	0.09	0.86**	-0.17
PSS	0.07	-0.15	0.83**
-- study A2 --			
HNR	0.83**	-0.17	
<i>Complex brightness</i>	-0.34	0.90**	
-- study B ---			
HNR	0.91**	-0.07	0.47
<i>Complex brightness</i>	-0.52	0.81**	0.00
Loudness	0.42	-0.07	0.84**

Table 9. Correlations between acoustic features and dimensions of the *motor* category / studies A1, A2, B ($df = 14, 12, 17$, respectively, ** $p < 0.01$).

- Dimension 1: Harmonic emergence (**HNR**)

For all three stimulus sets, several acoustic features correlate highly with this shared dimension, but they were of quite different types, and not all of them were significant. Furthermore, only one feature correlated well with this first dimension for the three stimulus sets: the Harmonic-to-Noise Ratio (HNR). Perceptual differences in the sounds along this dimension are related to the amount of harmonic (or pseudo-harmonic) energy in the signal. The HNR linear regressions with the first dimension of every motor stimulus set are shown in figure 118 and 119. The other features that correlated highly with this dimension were usually spectral envelope features. Actually, those high correlation scores are consequences of the HNR correlation. Indeed, the spectral envelopes of both parts of the sounds have quite different behaviors, and when the proportion of both parts is modified, the overall spectral aspect of the sound is also modified.

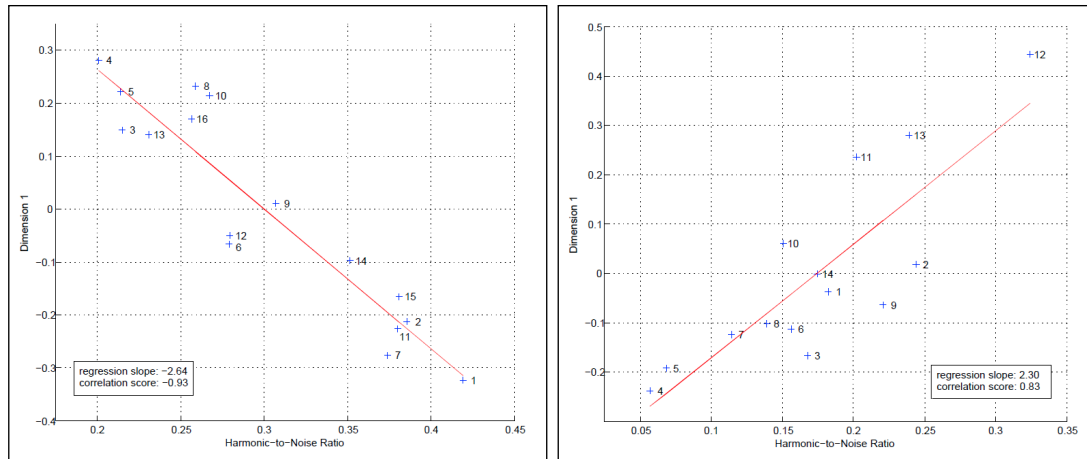


Figure 118. Linear regression between dim. 1 and HNR
(left) *motor class / study A1*, (right) *motor class / study A2*.

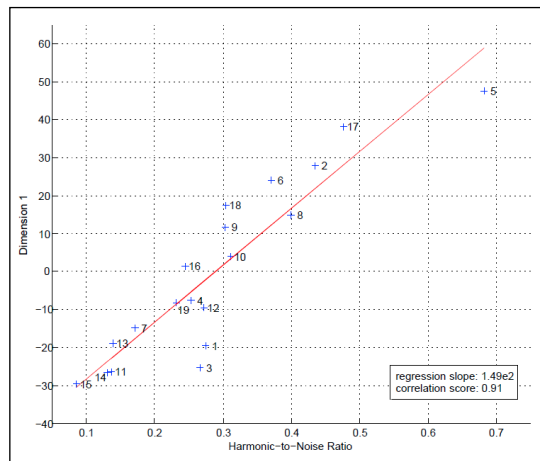


Figure 119. Linear regression between dim. 1 and HNR
motor class / study B.

- Dimension 2: **Complex brightness**

For the three stimulus sets, when listening to the sounds along this scale, brightness features, such as spectral centroid or sharpness, seem to explain the dimension. However, for the two stimulus sets in which the harmonic part is most prevalent, i.e. sets A1 and B, the perception of brightness seems to depend on the harmonic proportion. Indeed, the brightness perception of a predominantly noisy sound is not the same as that of a predominantly harmonic sound, all the more because both parts have quite different spectral behaviors: the energy of the harmonic part is quite concentrated in the low frequencies for this type of sounds. It is thus essential to take into account both the harmonic and noise parts in the brightness estimation. That is the reason why multidimensional linear regression theory (see [Legendre et al., 1998]) is applied in order to characterize that dimension with a unique feature depending on the brightnesses of both parts. Therefore, for each of the three stimulus sets, a linear combination of 3 components is found to be significantly correlated: Complex brightness = $\alpha.x_1 + \beta.x_2 + \gamma.x_3$, where x_1 is the Perceptual Spectral Centroid of the harmonic part, x_2 is that of the noise part and x_3 is the overall Perceptual Spectral Spread (see Appendix 8). Both harmonic and noise parts were separated with the method and MATLAB code taken from Ellis [Ellis, 2003]. The linear regressions of the

obtained "complex brightness" with the second dimensions of the *motor* meta-category are shown in figure 120 and 121. However, no common combination was found to be correlated for every stimulus set. Table 10 shows the coefficients of this "Complex brightness" for each stimulus set.

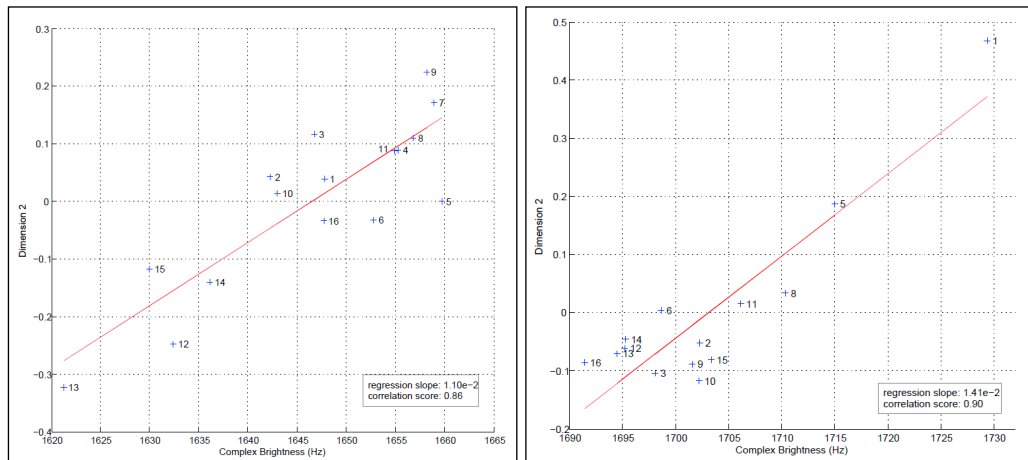


Figure 120. Linear regression between dim. 2 and Complex Brightness (left) *motor* class / study A1, (right) *motor* class / study A2.

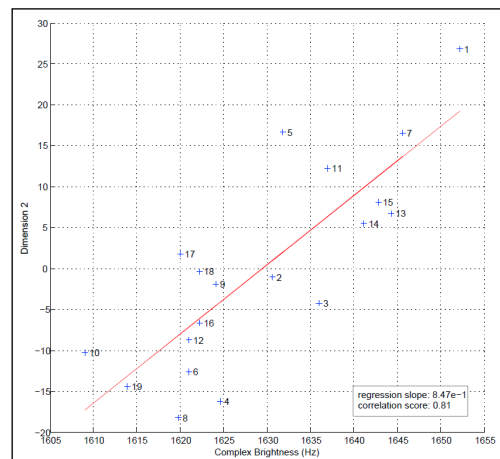


Figure 121. Linear regression between dim. 2 and Complex Brightness *motor* class / study B.

Study	α	β	γ
A1	+3.82e-3	1	-89.7
A2	+1.15e-2	1	-25.8
B	-1.08e-2	1	-63.0

Table 10. Coefficients of the linear combination defining Complex brightness for studies A1, A2 and B.

- Dimension 3

Study A1. This dimension seems to be well correlated with the **Perceptual Spectral Spread** – PSS (see Appendix 8) calculated with logarithmic scales for both magnitude (level) and frequency. The linear regression of this feature with the third dimension of the perceptual space of this study is shown in figure 122.

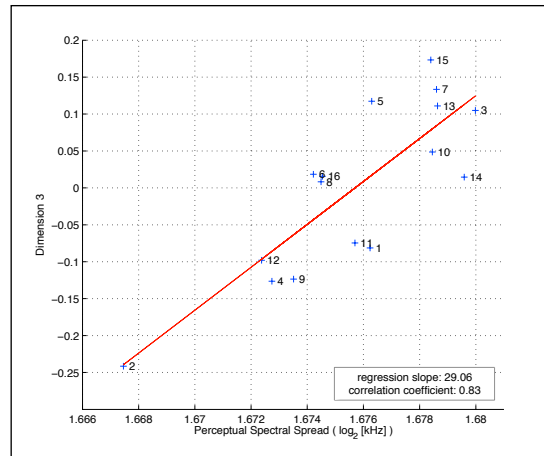


Figure 122. Linear regression between dim. 3 and Perceptual Spectral Spread
motor class / study A1.

Study B. Unlike study A, the sounds were not initially loudness-equalized in study B. Quite logically, the last dimension of this MDS analysis result is found to be significantly correlated with **Loudness** (see Appendix 8). The linear regression between Loudness and the third dimension of the study B perceptual space is shown in figure 123.

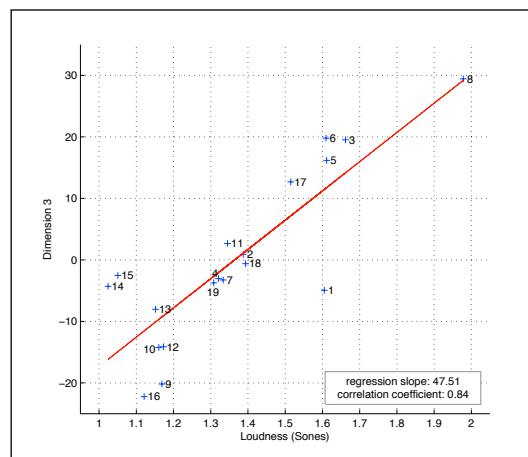


Figure 123. Linear regression between dim. 3 and Loudness,
motor class / study B.

Loudness is a perceptually strong characteristic that can easily prevent slight variations of other features from emerging. Moreover, the fact that no third perceptual dimension was obtained for stimulus set A2 can be related to the predominance of the noisy part, which can mask some variations of other features. On the contrary, when the sounds are loudness-equalized and when the harmonic part is not entirely masked by the noise, such as in stimulus set A1, a third perceptual dimension (PSS, Perceptual Spectral Spread) seems to emerge and matches that of the perceptual space of (pseudo-)harmonic instrument-like sounds (see § III.6.2.4.1.2 – Dimension 3). For these reasons, we were not able to unify this third dimension along the three corpora (A1, A2 and B).

III.6.2.4.1.3 *Instrument-like category*

This sound category corresponds to the stimulus set of study C. Its MDS analysis resulted in a 3-dimensional perceptual space presented in § III.6.2.2.3. According to

the correlation scores in Table 11, those 3 dimensions are related to three different acoustic features presented below:

	Dimension 1	Dimension 2	Dimension 3
Roughness	-0.93**	-0.06	0.37
<i>Simple brightness</i> (PSC)	0.04	0.97**	0.05
PSS	0.05	-0.11	-0.90**

Table 11. Correlations between acoustic features and dimensions of the *instrument-like* category / study C ($df = 20$, ** $p < 0.01$).

- Dimension 1: **Roughness**

Study C. This dimension seems to discriminate the monophonic from the polyphonic sounds. When listening to the sounds along this scale, one goes from perfectly harmonic tones to successively pseudo-harmonic tones (tones with inharmonicity relationships between their partials) and polyphonic sounds (with several tones). Consistently, roughness correlates significantly with this dimension (see Appendix 8). The linear regression of roughness onto the first dimension is shown in figure 124.

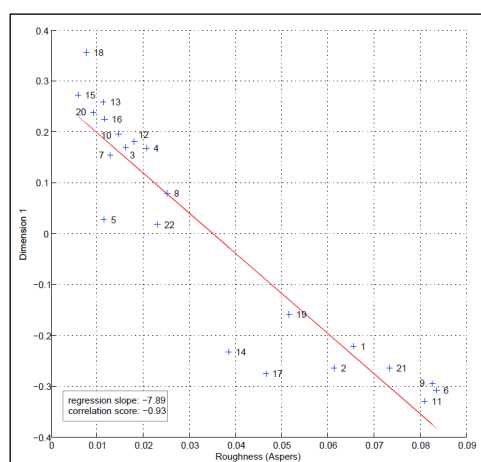


Figure 124. Linear regression between dim. 1 and Roughness *instrument-like* class / study C.

- Dimension 2: Perceptual Spectral Centroid (**PSC**) – *Simple brightness*

Study C. When listening to the sounds along this scale, the relation to the brightness of the sounds seems quiet obvious. This brightness is well quantified by the spectral centroid all the more when a perceptual model is used. Consistently, the Perceptual Spectral Centroid gives the best correlation score (see Appendix 8). We call it *Simple brightness* because it can be formally seen as the degenerated form of the *Complex brightness* defined in the previous section, when harmonic and noise part of the signal are not separated. The PSC linear regression with the second dimension is shown in figure 125.

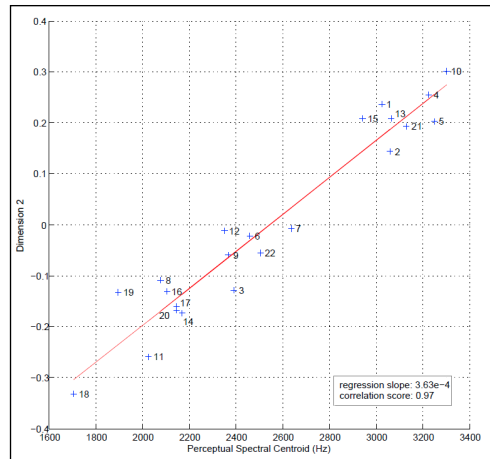


Figure 125. Linear regression between dim. 2 and Perceptual Spectral Centroid
instrument-like class / study C.

- Dimension 3: Perceptual Spectral Spread (**PSS**)

Study C. This dimension is the one whose interpretation is the most difficult just by listening to the sounds along the scale. However, it could be associated with their "richness". It correlates quite well with the Perceptual Spectral Spread (see Appendix 8). The linear regression of PSS onto the third dimension is shown in figure 126.

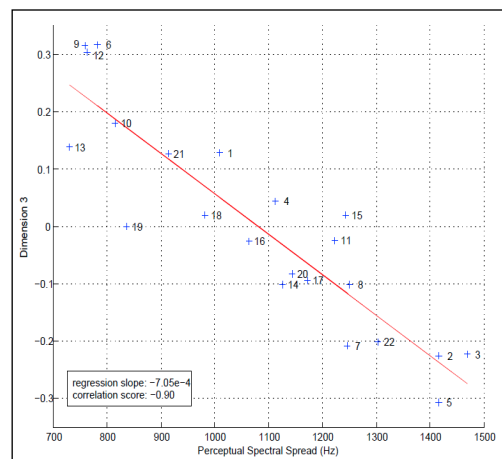


Figure 126. Linear regression between dim. 3 and Perceptual Spectral Spread
instrument-like class / study C.

III.6.2.4.1.4 *Impact* category

This sound category corresponds to the stimulus set of study D. Its MDS analysis resulted in a 3-dimension perceptual space (§ III.6.2.2.4). According to the correlation scores in Table 12, those 3 dimensions are related to three different acoustic features:

	Dimension 1	Dimension 2	Dimension 3
<i>Simple brightness</i> (PSC)	-0.89**	0.05	0.08
Cleanness indicator	-0.18	0.90**	0.24
RMS value	-0.27	0.18	0.88**

Table 12. Correlations between acoustic features and dimensions of the *impact* category / study D ($df = 10$, ** $p < 0.01$).

- Dimension 1: Perceptual Spectral Centroid (**PSC**) – *Simple brightness*

Study D. The feature that best suits this dimension is the Perceptual Spectral Centroid (PSC) that includes a hearing model (see Appendix 8). Indeed, this dimension describes the sounds' brightness. We call it *Simple brightness* for the same reasons presented in § III.6.2.4.1.2, regarding the second dimension of the *instrument-like* category. The linear regression between the PSC feature and the first perceptual dimension is shown in figure 127. However, it is noticeable that there is a categorization phenomenon along this dimension, as the sounds labeled 9, 11 and 12 are much lower on that dimension than the other ones. This phenomenon comes from the MDS analysis results and is not only related to the tested features. Nonetheless, it tends to improve the correlation score.

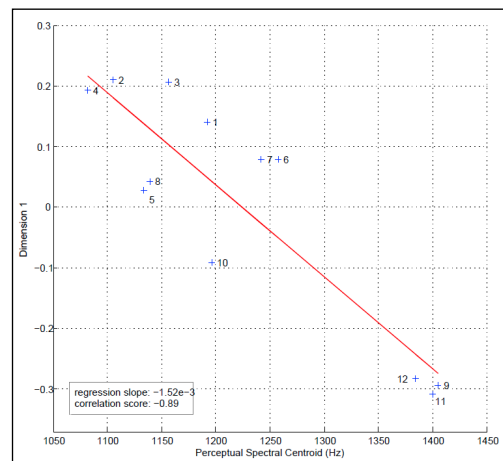


Figure 127. Linear regression between dim. 1 and Perceptual Spectral Centroid – *impact* class / study D.

- Dimension 2: **Cleanness indicator**

Study D. It seems, when listening to the sounds along this scale, that this dimension is linked with the *cleanness* of the sounds. More precisely, it discriminates sounds containing only one impulse such as the sounds numbered 1, 2 and 3, from those in which one or more impulses follow the main one (rattle, bounce...), such as the sounds numbered 10, 8 and 7. The acoustic feature (*Cleanness indicator*) that best suits this dimension is an estimator of the short-term loudness variability of the sounds (see Appendix 8). This linear regression of the Cleanness indicator onto the second dimension is shown in figure 128.

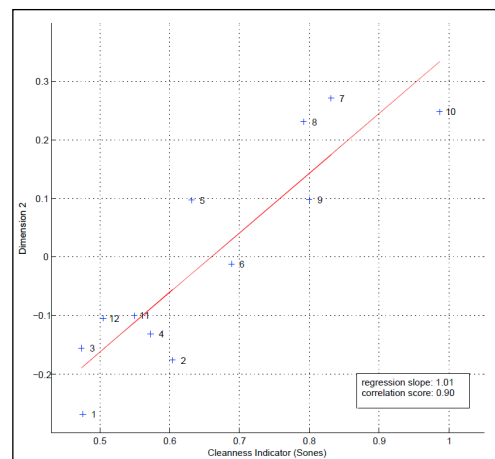


Figure 128. Linear regression between dim. 2 and Cleanness indicator – *impact* class / study D.

- Dimension 3: **Sound level**

Study D. The RMS value is correlated with this dimension. Indeed, the dimension seems to be somehow related to pulse amplitude. The linear regression of this feature onto the perceptual dimension is shown in figure 129

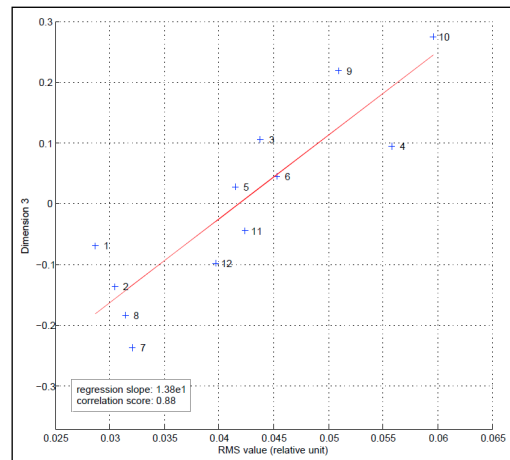


Figure 129. Linear regression between dim. 3 and Sound level
impact class / study D.

III.6.2.4.1.5 Discussion

Looking for regularities and singularities among the 3 important categories of environmental sounds derived from the first part of this study, we finally identified:

- One feature, *Brightness*, that is preponderant for the description of all sound categories (i.e., 1 dimension of the 5 perceptual spaces)., This feature is actually a combination of different spectral envelope features: the perceptual spectral centroid of both harmonic and noise parts of the signal (PSC_h and PSC_n) – or perceptual spectral centroid of the whole signal (PSC) – and perceptual spectral spread (PSS). And no unique combination has been found to describe uniformly this dimension. So this feature still remains a generic notion of brightness and cannot be transformed into a real metric for quantifying this dimension.
- One or two features, in each category, that are related to specificities of the corresponding sounds:
 - motor sound perception is largely characterized by the mixture of two highly discriminable parts, in terms of either energy or spectral content;
 - instrument-like sounds present timbre features that have been found previously for musical sounds (essentially, roughness);
 - an important part of the perceptual discriminability of impact sounds is related to a temporal behavior feature, describing the sounds' cleanness.

III.6.2.4.2 Categorical level: building an automatic classifier

Now that we have identified the inter-category specificities, we must address the development of a predictive tool able to automatically classify the sounds on the basis of a perceptually validated corpus. In other words, the aim here is to use the results presented in § III.6.2.4.1 as relevant cues in order to find a limited number of acoustic features that would be efficient for the implementation of an automatic perceptual classifier.

III.6.2.4.2.1 Specificities of the categories

Before considering the implementation of such a tool, it is essential to identify which features are used when listening to the sounds in order to discriminate the three categories. As partially concluded in § III.6.2.4.1.5, we can assume that:

- Impact sounds differ from the other ones in their temporal structure: they are quite short because they are damped, while the other sounds are as long as desired because they are sustained;
- Instrument-like sounds differ from the other ones in their spectral structure: their spectrum energy is usually localized in the middle frequencies and their spread is quite low, because they are harmonic sounds whose degree of spectral envelope decrease is high. To the contrary, the spectrum energy of the other sounds is localized in much lower frequencies with a much higher spread and a lower degree of spectral envelope decrease.

Thus, it seems obvious that the cue that discriminates motor sounds from impact sounds, for instance, is very different from the one that discriminates motor sounds from instrument-like sounds. As a consequence, it is quite certain that a unique feature will not be enough to describe the categories, and it is more likely that we will have to use a pair of temporal and spectral features.

According to these preliminary observations, a large set of temporal/spectral feature pairs could be used in order to discriminate the category to which a given sound belongs. Spectral and temporal features that seem to be good candidates for dealing with this problem are listed below. Their terminology and computing techniques are taken from [Peeters, 2004]:

- Temporal features: Log-Attack-Time (LAT), Temporal Increase (TI), Temporal Decrease (TD), Temporal Centroid (TC), Effective Duration (ED), Energy Modulation Frequency (EMF) and Energy Modulation Amplitude (EMA);
- Spectral features: *mean* component of Spectral Centroid (SC), Spectral Spread (SSp), Spectral Skewness (SSk), Spectral Kurtosis (SK), Spectral Slope (SSl), Spectral Decrease (SD), Spectral RollOff (SR) and Spectral Variation (SV).

III.6.2.4.2.2 Classification modeling tool: the multinomial logistic regression

Now that we have identified the feature combinations that are likely to discriminate the three sound categories, we need a regression modeling tool able to predict the values of a qualitative and polytomous dependent variable Y (i.e., the sound category) by a combination of quantitative independent variables X_1, \dots, X_k (i.e., acoustic features). This tool is the multinomial logistic regression (see [Legendre et al., 1998] and [Woodcock, 2003]). In its basic definition, logistic regression is used to discriminate only two different attributes (or values) of a binary dependent variable Y (with values 0 and 1). With the probability notation $\pi(x) = P(Y = 1 | X = x)$ of the event where the Y variable has the value 1, given the $x = (x_1, \dots, x_k)$ value of the $X = (X_1, \dots, X_k)$ set of variables, both event probabilities are related to each other by Eq. 45:

$$\pi(x) = P(Y = 1 | X = x) = 1 - P(Y = 0 | X = x) \quad (\text{Eq. 45})$$

A logistic regression tool models the $\pi(x)$ probability by a logistic function, formulated in Eq. 46. This function, which exhibits a sigmoid curve (“S-shaped” curve), is defined as the cumulative distribution function of a logistic probability distribution (similar to the normal distribution).

$$\pi(x) = \frac{1}{1 + e^{-u}} = \frac{e^u}{1 + e^u} \quad (\text{Eq. 46})$$

where u is a linear combination of the values of x : $u = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$

Its inverse function, the “logit” function, corresponds to the natural logarithm of the odds' ratio in Eq. 47:

$$\text{logit}(\pi(x)) = \log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (\text{Eq. 47})$$

When the dependent variable Y corresponds to a polytomous nominal response (i.e., that has more than two different unordered values), the generalized logit models are used. In our case, the dependent variable Y corresponds to a 3-valued response: ‘0’ for *impact*, ‘1’ for *motor* and ‘2’ for *instrument-like*. With the notation $\pi_i(x) = P(Y = i | X = x)$, the multinomial logistic regression consists in modeling the relationship between the set of independent variables $X = (X_1, \dots, X_k)$ and the generalized logits, $\log(\pi_1(x)/\pi_0(x))$ and $\log(\pi_2(x)/\pi_0(x))$. The model assumes a linear relationship for each logit as in Eq. 48:

$$\begin{aligned} \log \frac{\pi_1(x)}{\pi_0(x)} &= \beta_{10} + \beta_{11} x_1 + \dots + \beta_{1k} x_k \\ \log \frac{\pi_2(x)}{\pi_0(x)} &= \beta_{20} + \beta_{21} x_1 + \dots + \beta_{2k} x_k \end{aligned} \quad (\text{Eq. 48})$$

The regression tool searches iteratively for the best-fitting solution (β_{ik} coefficients) using the Newton-Raphson method and maximum log-likelihood as a convergence criterion. The predicted probabilities are then given by Eqs 49 to 51:

$$\pi_1(x) = \frac{e^{u_1}}{1 + e^{u_1} + e^{u_2}} \quad (\text{Eq. 49})$$

$$\pi_2(x) = \frac{e^{u_2}}{1 + e^{u_1} + e^{u_2}} \quad (\text{Eq. 50})$$

$$\pi_0(x) = 1 - (\pi_1(x) + \pi_2(x)) \quad (\text{Eq. 51})$$

where $u_1 = \beta_{10} + \beta_{11} x_1 + \dots + \beta_{1k} x_k$ and $u_2 = \beta_{20} + \beta_{21} x_1 + \dots + \beta_{2k} x_k$

III.6.2.4.2.3 Model selection

This tool is applied to the perceptually validated sound corpus established at the end of § III.6.2.3, in order to predict the belonging of a sound to one of the 3 identified categories. This corpus is large enough (207 sounds) to make the results of such a procedure relevant. According to the set of acoustic features selected in § III.6.2.4.1, we can compute a classification model for each pair of spectral/temporal features. The best model's selection is made on the basis of their respective log-likelihoods. The log-likelihood LL is a statistical feature that corresponds to the sum of each natural logarithm of the predicted probability $\pi(x)$ that a sound belongs to its supposed category, as described in Eq. 52:

$$LL = \sum_x \log(\pi(x)) \quad (\text{Eq. 52})$$

However, the log-likelihood value depends on the number of elements within the stimulus set, and having the same value with stimulus sets of different size is not as relevant. A way to take this into account is to calculate a likelihood ratio that quantifies

the gain in correct prediction of the model compared to the "intercept only" model, where only the β_0 constant coefficients are used³⁴. The likelihood ratio feature LR is obtained with the relation defined in Eq. 53:

$$LR = -2 * (LL_n - LL) \quad (\text{Eq. 53})$$

where LL_n is the "intercept only" model log-likelihood. This statistical feature allows us to compare the effectiveness of each model (i.e., each feature pair) in predicting the category to which a given sound belongs. The higher the LL and LR values are, the more efficient the model is (see [Legendre et al., 1998]). The LR value for each feature pair is shown in Table 13, where we can see that the SSp / ED model seems to best suit the data.

	LAT	TI	TD	TC	ED	EMF	EMA
SC	167.1	252.7	333.5	373.0	380.5	289.8	101.8
SSp	187.0	257.1	354.6	399.2	407.3	317.5	122.7
SSk	106.1	180.2	314.5	373.1	385.7	235.5	35.7
SK	110.2	188.3	318.9	374.6	386.9	245.1	43.3
SSI	167.1	252.7	333.5	373.0	380.5	289.8	101.8
SD	108.7	187.0	295.9	355.9	369.9	215.8	22.1
SR	98.3	177.9	295.2	353.9	368.3	223.5	43.3
SV	115.6	171.1	306.8	360.9	376.1	224.6	74.0

Table 13. LR value for each spectral/temporal feature pair.
Spectral features are in rows and temporal features are in columns.

III.6.2.4.2.4 Model validation

In order to test the robustness of the selected model using SSp and ED features (see § III.6.2.4.2.3), a usual method consists of:

- i/ re-estimating the model on a randomly selected reduced part of the stimulus set, 70% of it for instance (144 sounds with respect to the distribution in the 3 categories),
- ii/ calculating the estimated probabilities on the remaining 30% (63 sounds),
- iii/ evaluating the error percentage³⁵.

This procedure was performed 100 times with a different random selection of sounds each time. This method tests whether the effectiveness of the model prediction will hold when applied to other sounds than those used to estimate its coefficients.

When estimated on the whole 207-sound stimulus set, the best-fitting model makes 7 errors, which corresponds to an error percentage of 3.3%. Over the 100 times we performed the procedure explained above, we obtained the results presented in Table 14, calculated on the recall number (total number minus number of errors) of every remaining 30% selection of the stimulus set. One may observe that the mean recall percentage (95.9%) is rather high, not even much smaller than when obtained on the whole stimulus set (96.7%), which proves the model's adequacy for this dataset.

³⁴ This means that the "intercept only" model will give the same probabilities whatever the data. In the present case, it will give a $99/207=0.48$ probability of belonging to the *motor* category, a $49/207=0.24$ probability of belonging to the *instrument-like* category and a $59/207=0.28$ probability of belonging to the *impact* category.

³⁵ We consider as an error the case of a sound for which the probability of belonging to its supposed category is smaller than one of the two other probabilities. This means that if the model has to choose the category to which the sound belongs, it will choose a wrong one.

Minimum recall number	57
Minimum recall percentage	90.5%
Maximum recall number	63
Maximum recall percentage	100%
Recall number standard deviation	1.3
Mean recall number	60.4
Mean recall percentage	95.9%
Mean recall percentage interval	93.8% — 97.9%

Table 14. Results of the predicting tool based on SSp/ED features, after 100 runs of a 70%-learning/30%-predicting loop on the 207-sound perceptually extended corpus (Experiment 2)

III.6.2.4.2.5 Discussion

The selected model tested on a 207-sound stimulus set (augmented corpus established in Experiment 2, § III.6.2.3.2) gives significant stable results in terms of automatic classification with only around 4% mean error in the prediction, with only 2 predicting acoustic features. This is a rather encouraging result, even if this tool is built with only 3 main sound categories of quite different kinds (*motor*, *instrument-like* and *impact*). It could be extended to other categories in order to cover a larger scope of environmental sounds.

Other automatic classification methods exist that are much more complex and that use much more input information about the sounds. But considering the significant results of this relatively simple method, exploring these algorithms further is quite pointless. However, with more than 3 categories, these methods may outperform the one presented here and could therefore be useful for efficient automatic classification.

From a larger point of view, other classification approaches also exist that are less time consuming with regard to the available data needed for performing them: they usually consist in defining sound classes, collecting training examples for each class, computing a large set of spectral and temporal features on sounds and letting a machine learning method pick features that are efficient in discriminating the classes. But, the main difference between this approach and the one proposed in the present paper relies on the fact that in the former, the classes are arbitrarily defined (or at least, are the result of a single expert's analysis), whereas in the present paper the classes are deduced from an experimental procedure, which is more time consuming but allows them to be considered as perceptually relevant. This is one of the original contributions of this study with regard to traditional methods based on *a priori* sound categories and powerful learning techniques (e.g., like the ones used in the Music Information Retrieval research³⁶)

III.6.2.4.3 Summary

We have built a 2-level description structure of environmental sounds that consists of:

- a categorical level that considers the different sound categories corresponding to particular sound production mechanisms,
- a continuous level that defines, within each of these categories, the perceptual space of the sounds allowing the representation of the perceptual dissimilarity between two sounds of the same kind.

³⁶ <http://www.ismir.net/>

This description is associated with automatic processing of acoustic features. When considering a new sound of one of these kinds, this processing allows: i) the identification of the sound category to which it belongs, with regard to the probabilities estimated by the logistic regression model, and ii) its correct placement along several perceptual dimensions.

III.6.2.5 Conclusion

This work originally aimed to extend timbre description principles, usually used for musical sounds, to environmental sounds and to apply them in a more systematic manner to this class of sounds. It is based on a first step of re-examination and comparison of four primary studies mainly dealing with industrial (car and machine) sounds. An inventory of their respective contexts, motivations, procedures and results gave us input data consisting of 5 coherent stimulus sets with their associated low-dimensional perceptual spaces. It also allowed us to intuit some regularities and singularities among the different kinds of sounds under consideration. Within the restricted scope of these 5 stimulus sets, a 2-part experimental approach revealed 3 meta-categories (*motor*, *instrument-like* and *impact*) and precisely defined them in a larger scale by extending their respective contents. This categorical description structure is also coherent with the categories of product sounds that Özcan et al. found [Ozcan et al., 2007]. Finally, a modeling approach was designed to describe more precisely the intuited regularities and singularities of these 3 categories. This includes comparing the initial perceptual spaces by means of systematically correlated acoustic features, which can be summarized by two important facts:

- One feature is preponderant for the description of all sound categories, i.e., the *brightness* feature, usually based on spectral envelope features. Therefore, this perceptual feature appears to describe musical sounds as well as environmental sounds.
- One or two features, in each category, are related to a specificity of the corresponding sounds:
 - motor sound perception is largely characterized by the mixture of two highly discriminable parts, in terms of either energy or spectral content,
 - instrument-like sounds present timbre features, originally derived for the description of musical sounds,
 - an important part of the perceptual discriminability of impact sounds is related to a temporal behavior feature, describing the sounds' *cleanness*.

This modeling approach also includes the building of a predictive tool based on logistic regression able to classify automatically and rather efficiently (with only a 4% mean error) this meta-structure with regard to the 3 categories under consideration.

Note that contrary to musical timbre for which attack time is an important cue of the perceptual space, the studies revealed no temporal features corresponding to the two first categories. This may be mainly due to the quasi-stationary nature of these sounds. Nonetheless, a temporal parameter, associated with a spectral one, appeared to be fairly efficient in automatically discriminating impulsive environmental sounds (car door closing) from non-impulsive ones.

However, according to Özcan et al. [Ozcan et al., 2007], other major sound categories, such as liquid or cyclic sounds, exist and need a definition as well, and their main perceptual features must be investigated. Furthermore, they focused their study on domestic "product sounds", while we were more interested in industrial (cars and

machine) sounds. Considering environmental sounds in a more general sense may again reveal other categories that would also need to be taken into consideration when building an overall environmental sound description structure, in terms of either definition or automatic description.

From an application point of view, the relevant acoustic features obtained for the three categories of sounds will allow us to conceive a perceptually relevant organisation structures of large environmental sound collections and to propose retrieval systems using an intuitive query process by searching for sounds that are similar to a target sound in that kind of database. The search will be based on similarity metrics computed from the acoustics features, stored with the sounds in the database as proposed by previous studies for musical sounds [Blum et al., 1995; Misdariis et al., 1998a; Qi et al., 2002]. In a larger perspective, these results should also contribute to the elaboration of a functional Computer-Aided Sound Design framework as they will help users to describe, associate, compare, share and finally manipulate sounds that can be considered as prototypes or initial ideas of concepts that the designer has in mind and tries to materialize in the framework of a specific project.

III.7 Conclusion

Cette troisième partie relative à la thématique de la **perception du son** a donc permis de dresser le cadre théorique de la perception et de la cognition en recensant les principales théories (behaviorisme, psychologie de la forme, psychophysique, psychologie cognitive, etc.) et concepts (processus ascendant/descendant, représentations mentales, catégorisation et catégories, etc.) associés à ces notions.

En se plaçant ensuite dans le cadre plus précis de la perception auditive, un état de l'art des études relatives au timbre musical et à la reconnaissance des sources sonores a permis, en premier lieu, de faire la distinction entre son musical et environnemental et a également mis en évidence la formalisation de plusieurs niveaux d'écoute et d'appréhension du monde sonore (écoute réduite, causale ou sémantique), la notion d'espace perceptif et de descripteurs acoustiques ainsi que la notion d'invariant (structurel ou informationnel) qui permet d'expliquer en partie les mécanismes relevant de la psychomécanique. Les questions méthodologiques rattachées à cette thématique ont finalement été abordées sous la forme d'une présentation des principaux paradigmes et protocoles expérimentaux utilisés pour "mesurer" la perception et les sensations d'un sujet confronté à un stimulus sonore.

Dans ce contexte, les contributions personnelles à cette thématique de recherche se sont traduites par plusieurs travaux portant à la fois sur l'intégration des connaissances du timbre musical et de la modélisation de la similarité perceptive pour le développement d'un moteur de recherche par le contenu dans une base de données d'échantillons sonores, mais aussi par un apport en connaissance théorique concernant la structure hybride (catégorielle / continue) qui sous-tend la perception de sons environnementaux de natures relativement variées – démontrant ainsi une relative universalité de certains attributs perceptifs –, ou bien encore par des résultats expérimentaux relatifs à la question fondamentale de l'évaluation subjective de la qualité d'un instrument de musique ; ce dernier point abordant aussi bien la facture instrumentale traditionnelle (trompette) que certaines approches innovantes dans le domaine (contrôle actif).

Nous verrons, par la suite, que cette thématique relative à la perception sonore – au même titre, d’ailleurs, que les deux précédentes respectivement liées à la synthèse et la reproduction des sons – alimentera de manière induite le champ d’application visé, le design sonore, et que l’expertise scientifique acquise dans ces différents domaines, à la fois d’un point de vue théorique, méthodologique et technologique, contribuera à terme à le constituer en véritable sujet de recherche.

Quatrième partie

Domaine d'application

PARTIE IV. Domaines d'application

IV.1 Design sonore

Même si le terme "design sonore" est né – semble-t-il – au milieu des années 70, dans l'industrie du cinéma américain, avec la figure emblématique de Walter Murch, réalisateur de la bande-son de nombreuses productions de Georges Lucas ou Francis Ford Coppola (dont le fameux "THX 1138"³⁷ de 1971), beaucoup de monde s'accordent pour dire que les bases de cette nouvelle discipline ont réellement été posées avec l'ouvrage fondateur de Robert Murray Schafer « The tuning of the world », paru en 1977 [Schafer, 1977], traduit en français 2 ans plus tard sous le titre, « Le paysage sonore » [Schafer, 1979] et récemment réédité, en 2010 [Schafer, 2010].

Avant cela, c'est-à-dire avant les années 80, la recherche et le développement dans le domaine sonore se résument à des travaux d'ingénierie et de mesure acoustique, le son étant principalement perçu comme un phénomène négatif – on parle alors plutôt de "bruit" – et les industriels, notamment dans l'automobile, s'intéressant à cette dimension de leurs produits quasi-exclusivement en termes de réduction, voire d'annulation : le silence absolu comme idéal.

A partir de cette période, et donc de l'intégration des grands concepts de Schafer, deux mouvements successifs vont s'opérer :

- la prise en compte que le son intrinsèque – ou non-intentionnel – d'un objet peut véhiculer une notion de **qualité** et, à ce titre, peut être caractérisé, contrôlé et amélioré afin de s'adapter davantage au ressenti des individus / utilisateurs (voir, par exemple, l'étude sur le claquement des portières de voiture [Bezatz, 2007], celle sur les sonorités de systèmes de climatisation [Susini et al., 2004; Minard, 2013] ou bien encore les travaux généralistes de Richard Lyon [Lyon, 2000] ou Jens Blauert et Ute Jekosh [Blauert et al., 1997]).
- la prise en compte que, au même titre que le médium visuel, le son peut aussi être porteur de **sens**, d'information, ou d'**esthétique**. Dans ce cas, et notamment grâce à des avancées technologiques en termes de captation et de miniaturisation, l'objet en question peut être "augmenté" de nouveaux sons, cette fois-ci intentionnels, qui lui procurent une fonction ou une forme nouvelle (voir, par exemple, les travaux de Lemaitre et coll. sur un objet prototype appelé "Spinotron" [Lemaitre et al., 2009b]).

Cette dernière approche définit dans les grandes lignes le champ du **design sonore** qui pourrait se résumer de la manière suivante : créer des sons nouveaux avec l'intention qu'ils soient entendus dans un contexte donné [Susini et al, 2014]. Elle part du constat que l'activité humaine – et la vie en général – est, par essence, bruyante et que le son peut y jouer un rôle fonctionnel, esthétique voire social en influençant de manière profonde notre relation au monde : l'idéal n'est plus complètement silencieux comme le prouve, par exemple, le champ d'investigation ouvert par la problématique du véhicule électrique, objet silencieux par nature mais auquel il semble falloir donner une signalétique / signature sonore pour des raisons de sécurité des piétons et d'ergonomie de conduite (pour plus d'éléments sur cette question, voir entre autres [Cocron et al.,

³⁷ http://fr.wikipedia.org/wiki/THX_1138

2012 ; Dubois et al., 2012 ; Misdariis et al., 2013]).

Mais, toujours dans la lignée de Schafer, la démarche de design sonore milite également pour l'utilisation harmonieuse et parcimonieuse du son comme nouveau medium au risque de se retrouver dans une situation de « schizophonie » aboutissant à « un paysage sonore synthétique dans lequel les sons naturels sont de plus en plus remplacés par des sons artificiels, et où les signaux qui ponctuent la vie moderne ne sont plus que des substituts fabriqués par les machines ». Autrement dit, l'idéal n'est plus le silence absolu mais plutôt le son "intelligent", maîtrisé, utilisé à bon escient et quoiqu'il en soit, situé à l'opposé d'une utilisation aveugle – ou plutôt sourde ! – et systématique aboutissant inmanquablement à un état de pollution sonore ou de paysage « lo-fi » [Schafer, 2010].

Pour autant, et bien que de plus en plus établi dans les esprits et les pratiques, le design sonore reste malgré tout difficile à définir et à baliser. Initialement, pour Schafer, le design sonore – pensé dans ce cas par rapport à la notion de paysage sonore – repose sur une série de principes permettant de « juger et de corriger » : respect de l'oreille et de la voix, conscience du symbolisme des sons, connaissances des rythmes naturels et compréhension des mécanismes d'équilibre ; ainsi, le « véritable designer sonore » doit posséder « des connaissances en acoustique, en psychologie, en sociologie, en musique et en bien d'autres matières encore suivant les cas » [Schafer, 2010]. Depuis ces postulats, le cadre formel de la discipline s'est précisé et s'est ouvert à la recherche scientifique et technologique par l'intermédiaire de disciplines comme la synthèse sonore numérique, la psychoacoustique, les sciences cognitives ou l'ergonomie. Ce cadre reste néanmoins très large, et dans une certaine mesure, relativement variable pour ne pas dire flou : la sonification d'une interface numérique, la production de sons pour le cinéma, la composition d'une charte musicale pour un lieu ou un événement, l'étude du son d'un produit manufacturé ou bien la conception de la signature sonore d'une marque sont, parmi d'autres, autant d'exemples d'activités qui peuvent légitimement se prévaloir de rentrer dans le champ du design sonore.

On trouve, dans la littérature quelques tentatives de définition du domaine. En premier lieu – et déjà cité plus haut – Susini pour qui le design sonore consiste à « rendre une intention audible » [Susini et al., 2014] ou, plus précisément, à « penser la composante sonore dès la phase de conception d'un objet afin de faire entendre une intention » [Susini, 2011] ; cette intention pouvant porter soit sur la forme (qualité globale, esthétique), soit sur le fond (fonctionnalité, transmission d'information). Pour Robare, cette composante sonore *designée* constitue un élément de l'expérience-utilisateur d'un produit et un moyen efficace pour favoriser et développer les interactions [Robare, 2009] ; ce passage d'une simple transmission de l'information à un paradigme interactif complexe s'effectuant selon trois dimensions principales : l'utilité (« *usefulness* ») – pour que l'expérience soit facile (« *easy* ») –, l'utilisabilité (« *usability* ») – pour que l'expérience soit efficace (« *efficient* ») – et la désirabilité (« *desirability* ») – pour que l'expérience soit agréable (« *enjoyable* »). Pour Dandrel cité par Rodriguez [Rodriguez, 2003], le design sonore représente « l'un des attributs sensoriels du design » ; faire du design sonore consistant ainsi simplement à faire du design en intégrant la dimension – ou le matériau – sonore. Enfin, Rodriguez lui-même, conçoit le design sonore d'un point de vue purement professionnel comme un « travail du son appliqué » [Rodriguez, 2003].

Afin d'apporter des éléments de réflexion à ce sujet, nous nous contenterons ici – à la manière de Dandrel – de considérer le "parent" du design sonore, à savoir le **design**, et, sur la base de données extraites du « Court traité du design » de Stéphane Vial [Vial, 2011], d'en considérer quelques aspects étymologiques et historiques.

Les origines latines du design donnent deux sources : un nom (*signum*) qui indique le signe, la marque, l'empreinte et un verbe (*designare*) qui indique le fait de marquer d'un signe, signaler. La version anglaise du verbe (*to design*) prend, quant à elle, deux acceptions : le fait de dessiner, former des signes, tracer une figure ou un dessin, et le fait de concevoir, projeter en fonction d'un concept, d'une intention ou d'un dessein. A la lumière de l'étymologie, il apparaît donc deux notions associées au design :

- le **dessin** qui consiste à marquer quelque chose d'un signe ;
- le **dessein** qui consiste à concevoir – ou donner un sens à – un projet s'incarnant dans ce signe.

D'un point de vue historique, la première utilisation du terme "design" remonte à 1849 et la publication par Henry Cole du 1^{er} numéro du Journal du Design et des Manufactures (« *Journal of Design and Manufacture* »). Le design y est alors associé au principe d'une production industrielle combinant fonction, intelligence et décoration et affiche son ambition de faire converger les arts et l'industrie. Cette ambition reste à l'état d'utopie pendant de longues années et le développement de la Révolution Industrielle qui, outre la deshumanisation du travail, produit, selon certains, une laideur et une mauvaise qualité des produits manufacturés. C'est en réaction à cette tendance, et à l'industrialisation en général, que le design – étiqueté d'abord comme art décoratif ou art appliqué – s'affirme et se développe pour rendre « plaisant, beau et généreux le cadre matériel de [la] vie » (William Morris). Du statut d'art mineur, les arts décoratifs (ou appliqués) deviennent donc des arts majeurs et donneront naissance, en premier lieu, à l'Art Nouveau du début du XX^{ème} siècle.

Ainsi, après un premier mouvement de rejet de l'industrialisation, le design va progressivement réalisé cette alliance entre art et industrie imaginée par Cole à la fin du XIX^{ème} siècle et prendre tout son sens en tant qu'art appliqué « à l'industrie » (d'après la définition d'Etienne Souriau). Sous l'impulsion d'une union d'artistes allemands, préfigurant le Bauhaus, un artiste (Peter Behrens) prend la direction artistique d'un grand groupe de l'industrie électrotechnique (AEG) dont il conçoit aussi bien le design des produits que l'image de marque ou l'architecture des bâtiments. Dans la continuité de ce mouvement, se crée donc en 1919 l'école du Bauhaus qui se veut, selon les propres mots de son fondateur (Walter Gropius) « un établissement d'enseignement, conseiller artistique de l'industrie, des métiers d'art et de l'artisanat ». et qui comptera parmi ses "maîtres de formes", entre autres, Vassily Kandinsky, Paul Klee ou Laszlo Moholy-Nagy. Pour Vial, le véritable acte de naissance du design se situe donc à ce moment précis où les artistes décident « d'assumer » l'industrie » en travaillant « non plus contre elle et à cause d'elle, mais avec elle et grâce à elle ».

Ces idées s'exportent ensuite aux Etats-Unis à partir des années 30 – pour revenir en Europe 20 ans plus tard – avec l'avènement de l'esthétique industrielle (traduction de *industrial design*) dont Raymond Loewy est l'un des représentants les plus connus. En France, l'esthétique industrielle est notamment promue par Roger Tallon (directeur artistique de la 1^{ère} agence de design, le bureau d'études techniques et esthétiques, en 1949) et Jacques Vienot (fondateur de l'Institut d'esthétique industrielle, en 1951) sur les bases de l'idéologie fonctionnaliste moderne postulant que « la beauté d'un objet provient de son adaptation à sa fonction ».. Cette vision effectuera progressivement sa mutation vers le concept de design, moins « callo-centriste », c'est-à-dire « la tendance

contemplative de la pensée à privilégier le Beau et le ravissement artistique dans l'étude de la culture matérielle » ; le terme "design" étant officialisé par l'Académie Française en 1971.

Ainsi, par héritage des gènes du design, apparaît un possible cadre conceptuel à la discipline du design sonore dont une nouvelle tentative de définition pourrait être : marquer un objet d'un signe – ou signature sonore – (dessin) qui incarne un projet, un sens (dessein) ; le terme "objet" étant ici générique et pouvant être pris au sens d'objet tangible (produit manufacturé), d'objet numérique (interface homme-machine) ou d'objet spatial (environnement).

Le cadre conceptuel en question est celui d'un art sonore appliqué qui se déploie sur des thématiques propres en relation avec celles des arts appliqués communément admises (notamment, design d'espace, de communication ou de produit) :

- l'espace, l'urbanisme ;
- le multimédia, les jeux interactifs ;
- l'industrie des objets ;
- l'identité de marque ;
- le cinéma, le spectacle vivant.

En outre, dans sa dimension de recherche telle qu'elle se déploie depuis plusieurs années – notamment à l'Université Technique de Delft (TU Delft/Product Sound Design group³⁸), à l'Université d'Aalborg (Department of Architecture, Design and Media Technology³⁹), à l'institut d'Architecture de Venise (IUAV⁴⁰), à l'Ircam (équipe PDS⁴¹, Perception et Design Sonores) ou bien encore dans d'autres centres de recherche et d'enseignement –, le design sonore apparaît comme une instance privilégiée de cette relation art/science/industrie, en partie formalisée par Cole au XIX^{ème} siècle. En effet, en reprenant des termes cités par Schafer (en référence à O. Laske – [Schafer, 2010]), il permet de combiner le « savoir » du scientifique apportant des connaissances dans les disciplines connexes (acoustique, psychologie expérimentale, perception, etc.), avec le « savoir-faire » de l'artiste – compositeur / designer sonore – qui conçoit des sons nouveaux à partir de ses intuitions et de sa singularité créative, et ceci, le plus souvent, dans le cadre d'une collaboration industrielle qui impose ses règles fonctionnelles ou esthétiques.

En résumé, à partir des différents éléments explicités ci-dessus, il est possible d'essayer de qualifier la pratique du design sonore – aussi variée soit elle – par différents attributs.

Premièrement, en accord avec le point de vue de Ozcan et van Egmond [Ozcan et al., 2009], on peut raisonnablement affirmer que le design sonore est le résultat d'une pratique collective où le designer sonore – au centre de la sphère – interagit avec un ensemble de disciplines et de protagonistes connexes : scientifiques (acousticiens, psychologues, sociologues), ingénieurs techniciens voire consommateurs (au travers d'études marketing, par exemple). Ces interactions sont nécessaires pour déployer le processus de design sonore dans son ensemble et, de fait, caractérisent la discipline comme une **création interdisciplinaire**.

Par ailleurs, la seconde caractéristique du design sonore peut se nommer **création contrainte**, en contradiction avec une approche purement artistique. En effet, même si

³⁸ <http://www.io.tudelft.nl/en/organisation/departments/industrial-design/design-aesthetics/research/sounddesign/>

³⁹ http://vbn.aau.dk/en/organisations/pp_f8719ab3-f92e-4e56-9538-49ba29ac52d6.html

⁴⁰ <http://www.iuav.it/>

⁴¹ <http://pds.ircam.fr>

l'art n'est pas complètement détaché de toutes contraintes, il paraît évident que le design sonore – et par extension, le design – présente des contraintes spécifiques plus nombreuses et plus fortes, notamment en termes de spécifications générales (le plus souvent induites par des considérations marketing ou commerciales) mais aussi relatives à l'ergonomie, la technique ou encore le format singulier que peut prendre une œuvre de design sonore (durée, instrumentation, support ou lieu de "représentation"). En outre, ces contraintes peuvent parfois être exprimées de manière très précise par la définition de paramètres objectifs (hauteur, rythme ou autres paramètres de description du timbre) forçant ainsi le compositeur à s'adapter à des règles fortes qui transforment, dans certains cas, la pratique du design sonore en un processus de **création objectivement informée**. Ces composantes liées aux contraintes du design sonore sont à mettre en relation avec les propos de Vial à propos du design : « faire de l'art et faire du design sont deux choses complètement distinctes » [Vial, 2011].

Enfin, une dernière propriété relative à la discipline pourrait être la **création ciblée**. En effet, pour le designer sonore – et, à nouveau, par extension pour le designer – la contrainte est aussi placée à la fin du processus si l'on considère que l'objectif ultime est de concevoir la "meilleure œuvre" (ou le "meilleur son"). Dans ce cas, le qualificatif "meilleur" peut être compris comme le plus beau, plaisant, fonctionnel, ou globalement, le plus adapté aux règles établies initialement. En outre, cette notion de "meilleur" est souvent déterminée par des procédures de sélection ou de test, à nouveau, assez éloignées des mécanismes d'acceptation subjective ayant cours dans le cas d'une œuvre purement artistique.

IV.2 Etude et conception de sons de clignotant

IV.2.1 Préambule

L'intérêt de l'industrie automobile pour la pratique du design sonore est avéré depuis un certain nombre d'années, aussi bien en recherche – donc très en amont de la conception des véhicules – qu'en développement, pour implémenter des solutions viables sur des productions de série (voir [Bijsterveld et al., 2013] pour une revue historique sur la relation son/automobile).

Cette étude se place dans le cadre d'une collaboration avec le département Recherche du constructeur Renault et s'intéresse aux sons de clignotant. En reprenant la terminologie de Schafer [Schafer, 2010], il s'agit d'un son « naturel » (le bruit provoqué par les contacts d'un relais électromécanique) qui s'est transformé en son « artificiel » (son généré par un buzzer électrostatique – voir § II.1) et qui, répondant à un standard assez strict, a très peu évolué au cours du temps. Il fait par ailleurs, partie de la famille dite des "interfaces homme-machine (IHM) sonores" qui matérialisent, *via* le canal auditif, les informations transmises dans l'habitacle de la voiture, soit en complément d'une information visuelle (p. ex., signal de ceinture de sécurité) ou bien de manière autonome (p. ex., le signal d'aide au parking).

Cette étude part d'un test d'écoute effectué auprès d'utilisateurs mettant en évidence l'intérêt d'une innovation en termes de sonorité du clignotant. La convergence des savoirs s'est principalement opérée sur un plan méthodologique et technologique avec l'utilisation du moteur de synthèse par modélisation physique Modalys comme outil de formalisation et d'exploration d'environnements sonores inouïs.

En outre, les solutions finalement implémentées sur véhicule ont permis de mettre en évidence certaines capacités des technologies employées, et jusqu'alors inexploitées.

Dans le cas de cette étude, le design sonore a donc joué le rôle particulier de générateur d'idées à la fois créatives et techniques.

Cette collaboration a donné lieu à une communication scientifique au 4^{ème} congrès de la société européenne d'acoustique (Forum Acusticum) [Langlois et al., 2005].

IV.2.2 Publication [Langlois et al., Forum Acusticum, 2005]

Man-Machine Interfaces in Cars: Technological Impact of Sound Design

Sabine Langlois ^(a), Bénédicte Le Nindre ^(a), Claire Peteul ^(a), Nicolas Misdariis ^(b)

^(a) Renault, Research Department, 1 Avenue du Golf, 78288 Guyancourt, France

^(b) Ircam, Sound Design team, 1 place Igor Stravinsky, 75004 Paris, France

IV.2.2.1. Introduction

Apart from the radio, the sound environment inside a car consists of the mechanical noises of the running car and the electronic sounds added to help the driver. We focus here on the turn signal.

Whereas the turn signal was made before by a mechanical relay, it is now "played" electronically. This technological change allows us to modify the characteristics of the sound made by the turn signal and so offers the possibility to create sounds never heard before.

We first conducted a listening test to understand which turn signals the customers like and dislike. We thus collected ideas to design new sounds, which is what we did with IRCAM without taking into account the technology used in our cars. The next step was naturally for us (researchers in psychoacoustics) to learn how to transpose these sounds to this technology. The experience had an unexpected side effect: electronic engineers and suppliers discovered some unknown capabilities of the electronic devices.

IV.2.2.2. Sound Quality Study

Our first objective was to understand which turn signals the customers like to hear and why. We decided to work on sounds made by actual cars. We realized that many kinds exist, and that turn signals of Asian cars have a specific sonic signature. We thus decided to concentrate on European turn signals, and chose 12 cars representative of the market.

We recorded the sounds inside the cars, engine off, in a semi-anechoic chamber with a Head Acoustics HMS III dummy head. The sounds were played back in a quiet room through a Head Acoustics amplifier and HALL (Sennheiser) headphone; this apparatus ensures that sounds are played back at real levels.

Forty-three persons participated in the listening test. The test consisted of three steps:

1. Free categorization: the subjects were asked to group together the sounds they heard as sharing common characteristics, and to define the representative of each group. They were allowed to form as many groups as they liked, and to put as many sounds in each group as they wished.
2. Description of the group: the subjects were asked to write down how they would describe each group.
3. Projection along the "sound quality" axis: on the test sheet, we drew a line with "0" marked at one hand and "10" at the other end; the subjects were asked to place each representative along this axis.

During the test, the subjects could listen to the sounds as often as they needed.

We often use this type of test for preliminary studies [Saint-Loubry, 2000]. The theory underlying the categorization step can be found in [Guyot, 1996] or [Guénoche et al., 1993].

IV.2.2.2.1 Results

To analyze the free categorization, we used an algorithm [Guénoche et al., 1993] that checks which sounds are most often placed together in the same group, pair by pair. The result can be represented by a tree, whose leaves are the sounds and whose nodes are the consensual groups. The algorithm indicates the quality of the tree, i.e. the degree of agreement between the subjects (here 15%, which is under the 20% upper limit we accept).

We obtained here six groups, each containing between one and three sounds, as shown in figure 130. Note that we rejected sound 8 from group 6 because it sounded too differently from sounds 7 and 10.

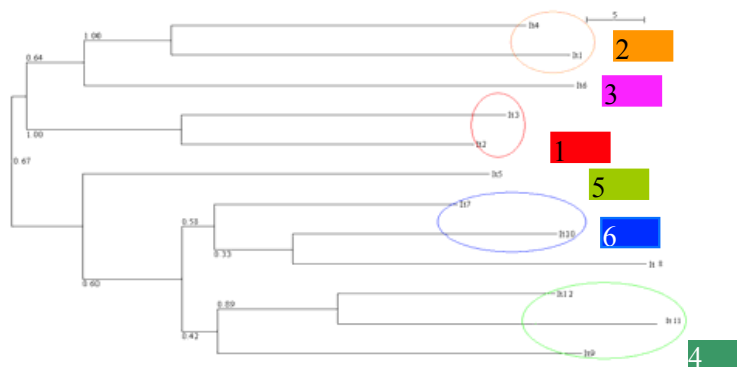


Figure 130. Tree computed from the categorization data

We conducted a semantic analysis of the verbatim to define which subjective criteria the persons used to describe each group. The verbatim dealt with sound description (regularity, speed, level, pitch difference between tick and tock, low / high-pitched, attack, decay, timbre), interpretation of similarity between turn signal and sound sources (material, animal, clock, watch, mechanical / electronic), and judgment (old / modern, pleasant).

The first group of sounds (the most consensual one) is described as “regular, large pitch difference between tick and tock, electronic”; the second as “irregular and slow”; the third as “irregular, quick, low level, no pitch difference between tick and tock); the fourth as “regular, quick attack (“sharp”), like a clock, metallic”; the fifth as “high level, large pitch and decay difference between tick and tock, aggressive”; and the sixth as “low-pitched, large decay difference between tick and tock, like a clock, old”.

To analyze the projection along the “sound quality” axis, we used software [Favre, 1999] based on multi-dimensional scaling, that optimizes the distances between sounds to get the most consensual representation, in one or more axes. If we use only one axis, we get a large error (27%), a poor result that indicates that the customers do not share the same preference. The representation along two axis is much better (error is reduced to 17%).

When we go back to the “sound quality” axis of each subject, we observe that nobody uses the full range of the scale: the best mark is only 6/10, which allows us to conclude that the customers are not very satisfied with the actual turn signal sounds.

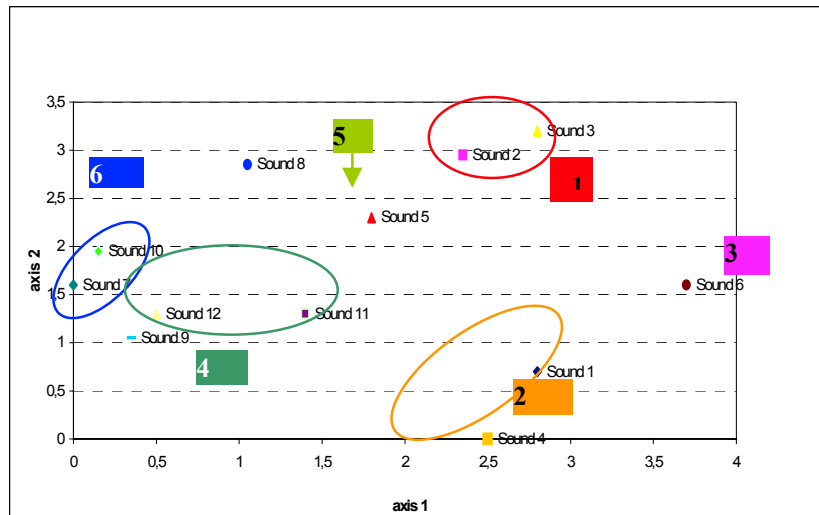


Figure 131. Two-axis preference mapping computed from the "sound quality" projection data

Note that on the preference mapping, the software puts the most appreciated sound at zero. Figure 131 shows that this two-axis representation allows us to find again the categorization groups, which was not possible with one axis. This result guarantees the reliability of the two-axis mapping. In order to understand why people like or dislike sounds, we can thus refer to the verbatim analysis and get the following results:

- the first part of the population of customers appreciates regular sounds with high level, even rhythm (duration between tick and tock equal to the one between tock and tick), and large pitch difference between tick and tock. They like turn signals that refer to the mechanical world of clock and watch, i.e. turn signals that sound like the mechanical relay used before.
- the second population prefers slow, low-pitched sounds, with low level and quick decay.

IV.2.2.2.2 Conclusion

Whereas for the first population, the function of the sound ("my turn signal is on") should be the priority, the second population appreciates that the turn signal becomes part of vehicle comfort, and expects that car manufacturers explore the possibilities offered by electronics to propose turn signals with new tones. Everybody agrees on the fact that actual sounds are not so nice.

This test has also brought to light the evocative power of sounds: some people associated these sounds with clocks, animals and materials. This result offers us a new domain to investigate.

IV.2.2.3. Sound Design

We already had the idea that interactions exist between sound and other senses. The associations suggested during the test confirmed that this is an interesting direction for sound design of turn signals. We thus decided to create sounds of turn signals that would be in harmony with the ambience inside the car, and especially with the materials used for seats, door cladding, dashboard... As such, hearing will match with touch and vision. We also wanted to design different sounds for the different cars made by Renault: family, sporty, city or luxury cars, where a specific ambience defines each type of car.

IV.2.2.3.1 Method

To create sounds in direct connection with material composition (wood, metal, plastic, tissue...) and touch (knocking, friction...), we decided to work with IRCAM Sound design team and the sound synthesis tool Modalys, developed at IRCAM for about fifteen years. This tool is a sound synthesis engine based on a physical modeling that gives access to a quite large collection of elementary physical structures, either mechanical (string, plate, ...) or acoustical (closed-open tube, ...) and a series of physical interaction (hit, plucked, bowed, blown or even glued, ...). Besides, thanks to its modular property, one can combine elementary models as mentioned above to create a complex structure. For instance, a musical instrument model, or any other "super-structure" that could be, of course, much more unrealistic than the former. The sound computed by Modalys is then the vibrating data of one or several points of the model, as if one would put a contact microphone on the surface of the given structure (see [Eckel et al., 1995]).

In order to create turn signal sounds, IRCAM Sound Design team explored different types of structures, with different mechanical parameters (density, Young's modulus ...), excited in different ways, and got thirty different sounds. On the one hand, the sound quality study and further objective analysis (not presented in this paper) have shown that there are fundamental properties that the sound has to respect in order to fulfill its function of information. These properties are regularity (duration between two ticks perfectly constant), even rhythm (duration between tick and tock equal to the one between tock and tick), speed and overall level. Note that we also took into account the bandwidth of frequency response of the buzzer (roughly between 1 and 5 kHz). On the other hand, these studies have brought to light other directions as yet unexplored, in particular timbre, pitch, attack and decay, with eventually differences (including level differences) between tick and tock.

IV.2.2.3.2 Conclusion

Even if some parameters are fixed, there is a large variety of sounds that can be matched with the desired ambiances. The sound design approach, where we do not care about the technology that will be used to reproduce the sounds inside the cars (except buzzer typical frequency bandwidth), allows us to create sounds never heard before.

IV.2.2.4. Technological Impact on Sound Design

Once we got a database of thirty "ideal" sounds, we wanted to install them on a car, simply to know what to do to make it possible. We will discuss here the case where the sound of the turn signal is played by the electronics of the dashboard. Note that some car manufacturers already use the audio system to play that sound.

IV.2.2.4.1 Method

From the thirty sounds, we chose the three, which sound the most different:

- one with low pitch and slow decay,
- one with high pitch and quick decay,
- one with a sharp attack, and an increased overall level.

We gave these sounds to different dashboard suppliers, asking them to transpose the three target sounds into their technology and to let us listen to the transposed sounds.

We spoke together to explain why some transposed sounds were close to the target sounds, why some were far away. As such, we were able to define a way to create sounds for our future cars.

IV.2.2.4.2 Results

Through the discussions with suppliers, we could get a better understanding about how the technology makes the sound. Two parts should be separately looked at: sound generation from electrical tension, and sound emission.

When a sound is played by the electronics of the dashboard, the sound emitter can be either a buzzer (piezoelectric or magnetic) or a loudspeaker. Whereas the frequency response of the buzzer presents strong peaks (fundamental, usually at 2 kHz, and harmonics), the response of the loudspeaker is flatter, especially at high frequencies, as illustrated in figure 132. The sound emission performances depend also on the fixation of the sound emitter and the available air volume. Note that some research deals with increasing efficiency of piezoelectric sound emitters (see [Woodard, 2005]).

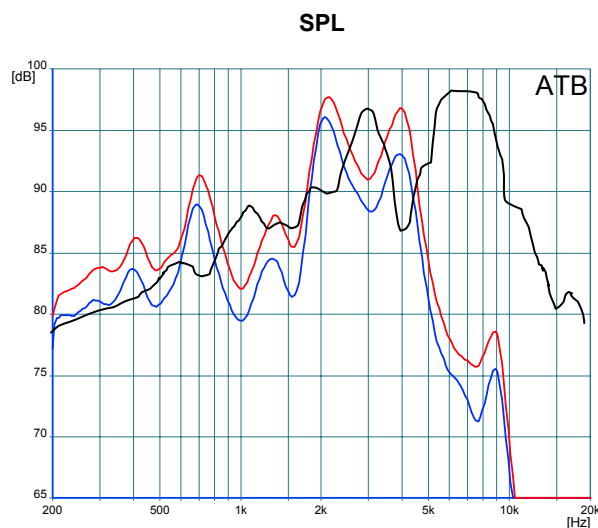


Figure 132. Frequency response of examples of buzzers (red and blue lines) and loudspeaker (black line)

We will discuss here the case where a sound generator is installed in the microprocessor (sometimes there is none!). The sound generator is able to play only monophonic sounds. In that case, the characteristics of each note are the following. They are illustrated by figure 133:

- Total duration: defined by the duration of the pulse width modulation (PWM) signal;
- Frequency: the inverse of the duration between two square peaks of the PWM signal;
- Level: when a buzzer is used, level is adjusted by changing the duration of the peaks; when a loudspeaker is used, level can be also modified by changing peak amplitude;
- Decay: it is possible to add decay within the total duration, which gives a so-called “gong” effect. Note that the duration of the decay depends on the frequency, which makes sound creation not so easy.

Note that it is not possible to change the note attack, whereas this parameter is very important for timbre.

The different sound generators do not offer the same degrees of adjustment for each characteristic. For example, the number of frequencies can be limited to 100, which does not always allow forming a perfect fifth between two notes.

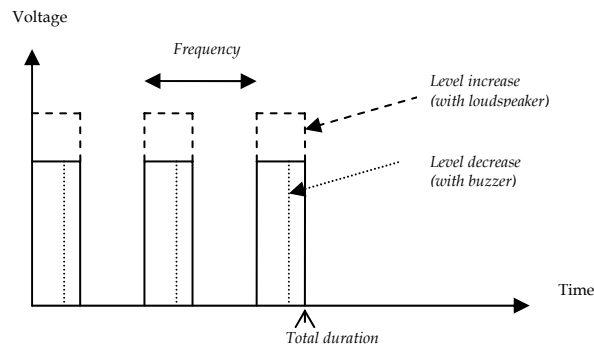


Figure 133. Sound characteristics (in italics) produced by the PWM signal

Extending the capacities of the sound generator (more degrees of adjustment), one supplier was able to create sounds close to the targets.

IV.2.2.4.3 Conclusion

Our first intention was simply to install our new sounds on cars, but finally, the method, that consists in proposing target sounds to dashboard suppliers and inviting them to transpose them to their technology, had two consequences. As expected, it gave us (researchers in psychoacoustics) the possibility of learning about the capacities of the system. In addition, both Renault electronic engineers and dashboard suppliers discovered unknown capacities of the electronic components. In today's Renault, electronic sounds are composed of few notes, where only duration and frequency are adjusted (level stays at maximum and decay at none), we are now able to tune sounds with four parameters (duration, frequency but also level and decay) with numerous degrees of adjustment. We can expect to get nicer sounding turn signals and other sonic man machine interfaces, without any cost effect. The next step will be to use the audio system to play the sounds, which will offer us new possibilities: larger frequency bandwidth, especially at low frequencies, possible change of attack, and polyphony.

IV.2.2.5 Conclusion

The sound quality study on turn signals provided us with information about customers' expectations and the acoustic features needed to fulfill these expectations. On this basis, and especially the idea to create sounds in harmony with material used inside the car, we worked with IRCAM Sound Design team to create a database of turn signal sounds never heard before.

Working together with dashboard suppliers, we discovered that Renault has created very poor sounds compared to the possibilities offered by the electronic components: only two parameters were used whereas four are available.

The sound design approach, that consists here of creating sounds according to our intentions and then transposing these target sounds to the technology of Renault cars, changes our habits: instead of designing sounds with normally used parameters, we pushed the known limits of the technology in order to better use it, but without changing it, so without cost effect.

IV.3 Stratégies de sonification d'un clavier d'ATM

IV.3.1 Préambule

Cette étude de cas s'intègre doublement dans le paradigme du design sonore tel que décrit précédemment (voir § IV.1) dans la mesure où elle concerne un objet tangible (clavier) servant d'interface à un objet numérique (écran de distributeur automatique de billets – ATM, *Automatic Teller Machine*). On étudie ici la capacité du son à favoriser la manipulation de cette interface, à la fois en situation normale mais aussi en situation dégradée – c'est-à-dire résultant, par exemple, d'une surexposition ou d'un dysfonctionnement de l'écran.

Par un procédé technique simple, le clavier est sonifié, c'est-à-dire "augmenté" d'un retour sonore accompagnant les actions qui lui sont associées. L'objectif du travail est d'étudier l'influence de la nature de ces retours sonores sur l'utilisation du dispositif et le ressenti des utilisateurs, au moyen de la construction de scénarios réalistes de mise en contexte (consultation de compte bancaire, retrait d'argent, etc.).

La nature des sons est principalement définie par la notion de caractère naturel (*naturalness*), les mesures expérimentales effectuées concernant la performance des tâches scénarisées ainsi qu'un ressenti relatif à l'aspect fonctionnel (*usability*) et agréable du dispositif (*pleasantness*).

Enfin, par transposition d'une hypothèse étudiée – et en partie vérifiée – par Tractinsky sur la relation entre esthétique et fonctionnalité [Tractinsky et al., 2000], ce travail cherche à tester une hypothèse dérivée qui peut se formuler de la manière suivante : *what is natural is usable*.

Cette étude a été réalisée dans le cadre du projet européen CLOSED⁴² et a fait l'objet d'une communication scientifique à la 6^{ème} conférence *Sound and Music Computing* (SMC) [Susini et al., 2009] ainsi que d'une publication dans le *Journal of Multimodal User Interfaces* [Susini et al., 2012].

IV.3.2 Publication [Susini et al., Journ. Multimodal User Interfaces, 2012]

Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback

Patrick Susini, Nicolas Misdariis, Guillaume Lemaitre, Olivier Houix
STMS Ircam-CNRS-UPMC. 1, place Igor Stravinsky – 75004 Paris

IV.3.2.1 Introduction

Sound is widely used in variety of products, ranging from desktop computers to mobile phones, and from coffee makers to electric cars. One of the most important issue for the sound designers is that of the relationships between the sounds and the information to be conveyed to the users.

Two kinds of information are usually considered in the design process. The first one is related to the function of the sound, as regards the use of an object or an interface. The aim in this case is to design sound characteristics that will establish an efficient interaction between the user and an object or an interface: for example, reaction to an alarm sound is faster when the alarm sound has a rapidly repeating pulse. The second

⁴² <http://www.closed.ircam.fr>

kind of information is related to the global coherence between a sound and an object or the identity of a brand. In this case, the aim can be for example to design a pleasant sound.

IV.3.2.1.1 Auditory displays: mapping sounds and information

A variety of interfaces commonly use sounds to display information to a user. Some applications have used sounds to let users explore large amounts of complex data (e.g. seismic data [Fröhlich et al., 1999]). More commonly, sounds in interfaces provide a feedback to a user's action (e.g. the deletion of a computer file), or warn the person that something is happening (e.g. an alarm clock) [Hermann, 2008].

In fact, the design of warning signals has become increasingly sophisticated, in particular for applications that need to display different warnings with different meanings and different urgency levels to the users. Such interfaces are now relatively common in applications like hospital or car equipment, or high performance aircraft [Edworthy et al., 1991 ; Patterson et al., 1986 ; Stanton et al., 1999]. The design of warning signals provides the designers with an interesting framework to analyze how sounds can display information. Three types of relationships are usually considered: symbolic (arbitrary relationship), iconic (representational relationship), or causal [Jekosch, 1999]. Symbolic relationships (e.g. different tone frequencies coding different levels of urgencies, or the Morse code) allow the designer to map any kind of information to any kind of sound parameter, but requires the users to learn the mapping. Iconic (e.g. a downward pitch to signal the loss of altitude to the pilot of an aircraft) and causal relationships (e.g. the squeaking sound of a car braking to signal a potential accident to car driver) limit the design possibilities to existing referential sounds, but relies on relationships already learned by the users.

Sounds in human-computer interfaces can be analyzed in the same framework. Historically, designers of humancomputer interfaces have focused on sounds in the form of short abstract static signals, typically warning or feedback sounds. The iconic and arbitrary relationships have been used to respectively design two types of sounds for humancomputer interactions: auditory icons and earcons. Auditory icons are iconic caricatures of sound occurring in everyday life situations that are transposed in a virtual environment [Gaver, 1986]. Earcons are arbitrary simple sounds that are related to elementary actions, in a virtual environment, using hierarchical rules for creating more complex messages [Blattner et al., 1989].

The most preeminent example of an auditory icon is probably the sound of a sheet of paper being crumpled and thrown down to a garbage can, used as a feedback to file deletion. This sound was originally designed as a part for Sonic Finder , a suite of sounds developed for Apple Computers, Inc. [Gaver, 1989 ; 1994]. Another example of a feedback created in Sonic Finder was a scraping sound when the user dragged a file on the desktop. In this latter example, the causal relationship between the sound and its meaning was created by using the sound naturally resulting from a user's gesture (at least metaphorically): dragging an object on a surface.

In fact, sonic feedback, both natural and created, are quite common in interactions with everyday objects. They are important because they tell the user that his or her action has resulted in the expected consequence. For instance, door latches make a characteristic impact sound when they correctly catch in the socket. This sound indicates that the door is securely locked, and a silent door lock would be utterly inconvenient. The design of auditory icons, or more generally sounds based on a causal

relationship between a sound and its meaning, therefore relies on the fact that users have already integrated a number of relationships between a mechanical event and its sonic consequences: listeners spontaneously identify the cause of a sound [Lemaitre et al., 2010]. The meaning of interfaces based on causal relationships are expected to be intuitively understood by users⁴³

IV.3.2.1.2 Causality and Naturalness

The perception of causality is closely tied to the synchrony between a user's gesture and the resulting sound [Guski et al., 2003]. For example, whether two moving discs with crossing trajectories are perceived as bouncing or overlapping is heavily affected by the presence, timing and nature of a sound occurring at the contact instant [Grassi et al., 2010].

But even more important are the expectations formed by an individual interacting with his or her environment. The sounds that a user "commonly" expects as a result of his or her gesture (i.e. the sound of an impact resulting from striking an object) are here referred as "natural". The natural relationships between a sound and a gesture are those driven by the laws of physics [McIntyre et al., 2002].

IV.3.2.1.3 Sonic Interactions: mapping sounds and actions

Causal relationships have been mainly studied for continuous sonic interactions [Rochesso et al., 2008]. In fact, the tight coupling between action and sounds has lead designers to create sonic interfaces that involve the active and continuous manipulation by the users [Hermann et al., 2005]. Such interfaces are, for instance, used to let users explore complex sonified data by actively manipulating sound sources and processings [Hermann et al., 2007]. Other examples are the sonification of an athlete's movements to help him or her achieve better performance [Eriksson et al., 2010 ; Schaffert et al., 2010], or the design of pseudo-haptic interfaces in wearable computers [Fernström et al., 2005]. Even though our research was concerned with discrete interactions⁴⁴ it is first important to examine the knowledge that can be gathered from such studies.

Two recent examples are worth detailing. In the first example, Lemaitre and colleagues [Lemaitre et al., 2009b] designed a tangible interface (the Spinotron) based on the metaphor of a child's spinning top: when the users pumped the Spinotron, they drove a physical model of a ratcheted wheel that produced a characteristic clickety-clack sound. The participants were required to pump the interface and to reach and maintain a precise and constant pace (indicated by a visual target). Half of the participants were provided with the continuous auditory feedback, half with only the visual target. Only the participants who were provided with the auditory feedback were able to improve their performance across trials. However, when asked to describe their appraisal of the sonic feedback, the subjects reported two interesting comments: first, they were not aware that the sound actually helped them improve their performance. Second, they found the sound irritating.

The methodology developed by Rath and colleagues [Rath et al., 2005 ; 2008] provides another interesting example of a sonic interaction. The Ballancer is a tangible interface consisting of a wooden plank that can be tilted by its user, in order to roll a virtual ball

⁴³ Note that using a natural or causal relationship may have its own drawbacks – e.g. users having an overly deterministic vision of the feedback model based on prior expectations from the "natural" situation at play

⁴⁴ A *discrete* feedback occurs for a finite and short amount of time as the result of a user's action (think of the beeps of a microwave oven); on the contrary, a continuous feedback follows the dynamics of a gesture sustained by the user. The sound produced by a musician bowing a string is a typical example of a continuous interaction.

along a plank. A user tilting the latter heard the rolling sound produced by the virtual ball. The authors used this interface to study subjects' abilities to use the auditory feedback in a task involving guiding the ball to a target region along the length of the plank. In a comparison using the same task with the rolling ball sound and with a "synthetic" sound (i.e. one that did not mimic any physical system) that preserved the same information, subjects were found to perform better, early in training, with the causal sound than with the "synthetic" one, whereas the latter provided better performance after training. Participants also reported having preferred causal sounds. These studies therefore suggest that using a causal mapping between the user's action and the resulting sonic feedback has two advantages. First, as in the case of passive auditory displays, causal relationships are already integrated by the subjects, and the interfaces are easier to learn. Second, users seem to generally prefer natural over synthetic sounds. It seems likely that sounds causally mapped to the user's actions would be perceived more natural than arbitrarily associated sounds. However, these studies all focused on continuous interactions. The current research studied a type of interaction that is more commonly found in interfaces: a discrete feedback.

IV.3.2.1.4 How to evaluate an interface ?

These studies also point us towards two aspects that are important for the evaluation of interfaces: the usability and the aesthetics of the design.

The usability of an interface is generally indexed by measuring the performances of users required to fulfill a task with the interface. Aesthetics is most of the times considered along two aspects: sensory pleasantness and annoyance. Sensory pleasantness, a notion that directly comes from the psychoacoustical tradition, is considered as an auditory attribute dependent on other auditory attributes such as roughness, sharpness, tonality and loudness [Fastl, 1997 ; , Zwicker et al., 1990]. Annoyance is a concept used to describe the nuisance caused by noises, particularly in the case of urban noises. Noise annoyance can be related to acoustic variables, but acoustic characteristics do not play the most important role. Psychosociological variables are important determinants [Guski et al., 2003].

In fact, both the objective usability of an interface and the pleasantness of the sound feedback can potentially influence the user's emotional reactions. According to Scherer's appraisal theory of emotions, both the appraisal of how well the interface allows the user reaching a goal and the pleasantness of the sounds may influence the valence of the user's feelings [Scherer et al., 2006]. Lemaitre and colleagues [Lemaitre et al., 2012] showed for instance that in the case of an interface that they designed (the Flops), the valence of the user's feelings were independently influenced by the difficulty of the manipulation of the interface and the pleasantness of the sounds (in this case, natural sounds were found more pleasant than artificially irritating ones).

But the usability of an interface is not only a matter of objective performance measurement: the usability of an interface is also perceived by the user. The perceived usability of an interface depends on the context, the goal and the skills of the user. It is also influenced by the aesthetics of the interface. Tractinsky et al. [Tractinsky et al., 2000] showed for instance that the user's initial aesthetic perception of an interface (an Automatic Teller Machine) influenced the perception of the usability, after its manipulation. The objective usability did not have such an effect. To put it in a few words, attractive products are perceived easier to use, or to borrow Tractinsky's words, "What is beautiful is usable".

IV.3.2.1.5 Goals of the study

Our goal was to evaluate the influence of sonic feedback on the appraisals of the interface. Specifically, we have investigated how naturalness interplays with the user ratings on pleasantness and perceived usability of the interface.

We used an interface that required one of the simplest gestural interactions possible: pressing a key. Natural sounds could therefore be easily defined as those naturally caused by a keystroke.

Similarly to Tractinsky et al. [Tractinsky et al., 2000], the keyboard interface was used to interact with an Automatic Teller Machine Interface (ATM). This is a commonly used interface. Moreover, sounds are in this case not a simple embellishment: they help the users correctly enter a Personal Identification Number (PIN). In fact, it is rather common that the sun's reflections on the visual displays makes the visual feedback (a star displayed after each key press) illegible, and the interaction particularly difficult and frustrating. A sonic feedback solves this issue.

More precisely, we were interested in three questions:

- Do the different mappings between a keystroke and the resulting sound change the perceived naturalness of the sound?
- Does the naturalness of the sonic feedback influence the perceived usability and pleasantness of the interface?
- Is the influence of the naturalness affected by the active manipulation of the interface?

This latter question is particularly relevant from a design point of view. Usually, auditory displays are evaluated by having listeners passively listening and evaluating the sounds. However, this underestimates the influence of the manipulation. Lemaitre and colleagues [Lemaitre et al., 2012] have shown for instance that evaluating sounds alone gives an overemphasized view of the actual influence of sounds on users' feelings when the sounds are embedded in an interactive interface.

Therefore, the study consisted of several experiments in which the subjects interacted with the interface, and the naturalness of the sound feedback was manipulated. The naturalness was manipulated by using three different mappings between the action of pressing a key and the resulting sounds. The highest degree of naturalness was achieved by using a causal mapping: the actual sounds of a keystroke. The lowest degree of naturalness was achieved by using sounds that did not bear any semantic relationships with a keystroke (e.g. a bicycle bell): an arbitrary mapping. In between, the medium level of naturalness was created by using sounds that had the temporal envelope of keystroke sounds, but a different timbre: an iconic mapping.

IV.3.2.1.6 Outline

The experimental design was inspired by Tractinsky et al. [Tractinsky et al., 2000]. It consisted of five steps. In the first step, the participants were presented with the interface and required to rate the naturalness, usability and pleasantness of the sounds. In a second step, the participants were required to perform a number of bank operations by manipulating the interface. In the third and fourth step, they had to indicate how the manipulation of the interface had changed their appraisal of the sounds. The fifth step was used to validate some results observed in the previous steps. Before reporting on this experiment, the next section will first detail the design of the interface and the sounds.

IV.3.2.2. Design of the sonified ATM interface

IV.3.2.2.1 The ATM interface

We designed a simplified interface for a bank Automatic Teller Machine (ATM). The interface consisted of a numerical keypad (Mobile Numeric USB Keyboard) connected to a laptop computer. The interface was sonified: when a user hit a key, a sonic feedback was played. The interface was controlled by Cycling 74's Max/MSP v.4.6. The users could use the ATM interface to execute different bank operations.

There were two scenarios:

- Withdrawing cash from an bank account. This scenario required the user to enter a bank account number and the amount of money to be withdrawn. The account number consisted of the same four-digit number for all participants.
- Transferring cash from an account to another. This scenario required the user to enter two bank account numbers and the amount of money to be transferred.

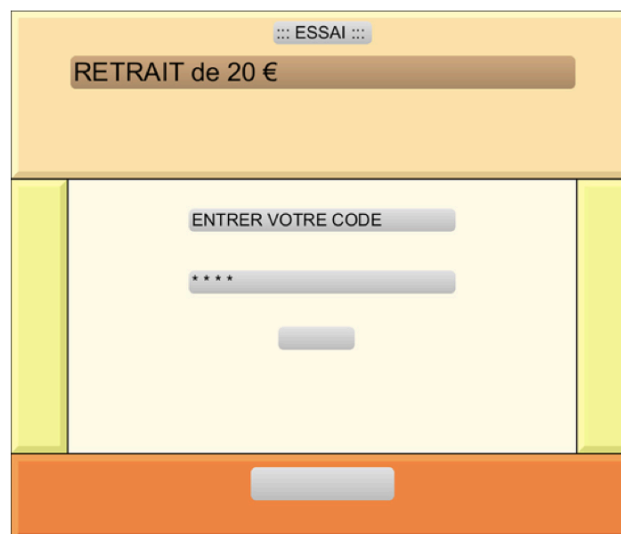


Figure 134. The graphical interface of the ATM used in the study.
The interface was programmed in Max/MSP

Importantly, the users had to hit the “I” key several times to indicate an amount of money. For instance, they had to press six times the key to indicate an amount of 60 €. A star displayed after each keystroke was used as a visual feedback. This rather inconvenient procedure made the sonic feedback particularly useful for the participants⁴⁵.

The interface had two modes of operation. In the normal mode the sonic feedback was systematically played immediately after a key was pressed (normal). In the defective mode of operation keystrokes would sometimes fail to trigger the sonic feedback. The probability that the keystroke triggered a feedback was 70%. The graphical of the ATM is represented in figure 134.

IV.3.2.2.2 Creation of the sonic feedback types

The first experimental step (this step will be labeled "step 0" in the following of the article) was the selection of different sonic feedback types to be associated to a keystroke. Eightyone feedback sounds were first chosen or created according to three

⁴⁵ Note that this interface was only developed for the sake of the experiment. We do not advise to use such an interface in a real ATM.

different mappings between the action of striking a key and the resulting sounds (respectively: arbitrary, iconic, and causal). These mappings were intended to create three levels of naturalness (respectively: low, medium, and high).

– **Causal mapping:** Thirty-nine keyboard sounds were selected to create a high level of naturalness. These sounds were sampled from The SoundIdeas' General 6000 library⁴⁶.

– **Iconic mapping:** To create a medium level of naturalness, 21 sounds were created by cross-synthesis⁴⁷ [Roads et al., 1985]: an impulse envelope was applied to non-natural timbres. This procedure allowed us to create non-natural sounds with the same morphology as the sound of an impact.

– **Arbitrary mapping:** To create a low level of naturalness, 21 sounds were selected, which bore no morphological similarity nor semantic relationship with an impact or keystroke. These sounds didn't have impulsive envelope, and couldn't be associate.

Label	Origin	Description	Duration (ms)
H1	Recording	Computer keystroke (LF)	203
H2	Recording	Computer keystroke (MF)	241
H3	Recording	Computer keystroke (HF)	315
M1	Synthesis	Similar to a metallic ball	176
M2	Synthesis	Similar to a wooden clave	237
M3	Synthesis	Synthetic HF impulse	499
L1	Recording	Bicycle bell	527
L2	Recording	Spring-like sound (Foley)	737
L3	Recording	Fast ascending movement on a piano keyboard	998

Table 15. Different sounds used to sonify ATM keyboard – HF: high-freq; MF: medium-freq; LF: low-freq.

IV.3.2.2.3 Selection of the feedback sounds

To select the sounds that best conveyed the three levels of naturalness, a set of participants was required to rate the naturalness of the sounds.

Before rating the sounds, they saw a video illustrating the three different levels of naturalness. The video displayed a hand holding a hammer and striking a floor made of ceramic tiles, once. The video was presented three times, each time with a different soundtrack. The first sound was the real sound of the impact on the floor (causal mapping); the second was an artificial impulse sound (iconic mapping); the third one was a synthetic chord (arbitrary mapping). The participants were instructed that these three videos illustrated examples of a high, medium and low level of naturalness.

Twenty participants were required to rate the naturalness of the 81 sounds. They used a three-point scale with the following labels: "Not natural at all", "In between", and "Very natural".

From these ratings, we selected three sounds for each of the three levels of naturalness (making a total of nine sounds). For each level, we selected three sounds that more than 90% of the participants had rated at this level. In following of the article, the three levels of naturalness will be labelled L (low), M (medium) and H (high), and the sounds will be labeled L1, L2, L3 (lowest level of naturalness), M1, M2, M3 (medium level of naturalness), and H1, H2, and H3 (highest level of naturalness). The nine sounds are listed in Table 15 and are available at http://pds.ircam.fr/atm_sounds.html.

⁴⁶ <http://www.sound-ideas.com/6000.html>.

⁴⁷ We used the cross-synthesis module for Max/MSP developed at Ircam http://imtr.ircam.fr/imtr/Max/MSP_externals.

IV.3.2.3. Rating the sonic feedback

To assess how the different levels of naturalness influenced the user's evaluation of the pleasantness and usability of the feedback sounds, we conducted an experiment. The procedure of this experiment had several steps that are detailed in the next paragraphs. At each of these steps, the participants were required to rate different feedback sounds on different scales: naturalness, usability, and pleasantness. We manipulated two experimental factors expected to influence the ratings: the *naturalness* of the sonic feedback (factor 1, with three levels: low, medium and high), and the *mode of operation* of the interface (factor 2, with two levels: normal and defective).

IV.3.2.3.1 Three scales of evaluation

The participants were required to rate the different feedback sounds associated to the keys of the ATM, before (step 1: pre-experimental ratings) and after using the ATM (step 2: manipulation; step 3: post experimental ratings). For each of these scales, they were presented with an assertion (see below), and had to qualify the assertion by using a nine-point scale ranging from "I do not agree at all" (1) to "I completely Agree" (9). These scales were:

Naturalness (N): The scale of naturalness was defined similarly as in step 0: the participants had to assess whether they found that the sound was a natural consequence of the action of striking a key. This scale was described by the assertion: "I find that the sound is well associated with the keys".

Usability (U): For the scale of usability, the participants had to rate how useful they found the sounds for typing in the different operations of the ATM. This scale was defined by the assertion: "I find that the sound helps me type in information". Note that in step 1 (see below), the users did not manipulate the interface. Their ratings of usability were therefore only based on their listening to the sounds.

Pleasantness (P): For this scale, the participants had to rate how pleasant they found the sounds. They had to qualify the assertion: "I find this sound pleasant".

IV.3.2.3.2 Method

Participants. There were 90 participants (60 females and 30 males) between the ages of 15 and 58. All participants reported normal hearing and were native French speakers. None of them had participated in step 0.

Apparatus. The experiment was run on an AppleMacBook Pro laptop computer (Mac OSX 10.5), using Cycling 74's Max/MSP 5. The sounds were played through the computer built-in loudspeakers.

Stimuli. The nine stimuli described above were used in the experiment.

Procedure. The participants were split in two groups. Half of them used the normal mode of operation of the ATM (Group 1), and the other half used the defective mode of operation (Group 2). The procedure had five steps that are represented in figure 135. These steps were:

- Step 1: Pre-experimental ratings. The participants were first presented with the interface and the nine sounds, but could not use the interface. They had to rate the sounds on the three nine-points scale previously described (N, U, and P). Each sound was rated twice on each scale (test-retest).
- Step 2: Manipulating the interface Each participant was randomly assigned to one of three sub-groups (L, M, and H), each corresponding to one of the three levels of naturalness (L, M, and H). In each sub-group, the subjects manipulated the ATM

sonified with one sound of the corresponding naturalness level. The sounds were selected on the basis of each participant's judgement in step 1: for example, if a participant was assigned to the group L, he or she used a sound that he or she had rated with a low level of naturalness in step 1. This procedure made certain that the participant interacted with the system using a sound that he or she had evaluated at the corresponding level of naturalness. There were 15 participants in each sub-group. The participants performed the previously described tasks (withdrawing cash and transferring money between accounts) several times.

- Step 3: Post-experimental ratings Each subject rated the sound used in step 2 on the same three scales as in step 1.
- Step 4: Direct comparisons The participants were finally asked to directly rate if the sound was worse or better than what they initially thought. This is, therefore, a direct comparison of their pre- and post-experimental assessments. Step 4 was used to prevent a potential ceiling effect in step 3 (for instance, a sound already maximally rated in step 1 could not received a higher rating in step 3). The participants used a numerical scale ranging from – 4 to 4 and were presented with an assertion of the type: "I find that the sound was more natural/helpful/pleasant after experimenting with the interface". A negative rating meant that the participant did not agree, and a positive one that he or she agreed. In the following analyses, only the ratings measured in step 4 will be analyzed in detail.
- Step 5: Validation The last step was used to validate some of the results observed in step 3 and 4 for the high and medium levels of naturalness by having participants rate the usability and pleasantness of the interface at both levels of naturalness in the same session.

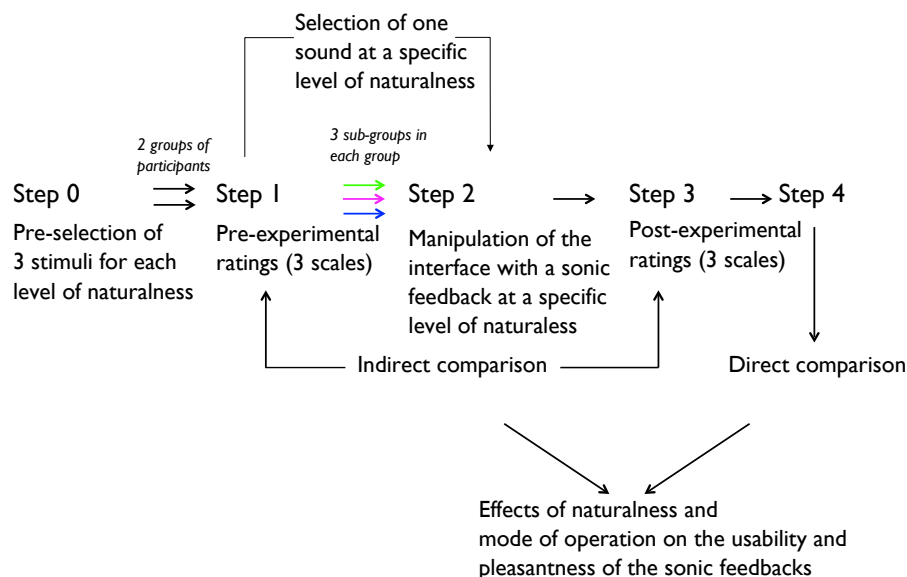


Figure 135. Different steps of the experimental procedure

IV.3.2.3.3 Results

IV.3.2.3.3.1 Ratings of naturalness in steps 0 and 1

The pre-experimental ratings of naturalness (N) allowed us to verify that the sounds selected in step 0 created the appropriate levels of naturalness for the participants in this experiment (steps 1 to 4). Figure 136 represents these ratings averaged across the 45 subjects in Groups 1 (upper panel) and 2 (lower panel). In both these panels, the

horizontal axis represents the naturalness measured in step 0 for the nine sounds, and the vertical axis represents the naturalness measured in step 1. This figure shows that the ratings of naturalness in step 1 preserved the rank ordering of naturalness measured in step 0: the sounds H1 to H3 were rated as more natural than M1 to M3, which were rated in turn as more natural than L1 to L3.

These observations were confirmed by submitting the data to a repeated-measure analysis of variance (ANOVA) with one within-subject factor (the nine sounds), and one between-subject factor (the two groups). The analysis revealed a strong effect of the sounds ($F(8,704)=167.8$, $p < 0.001$), no significant effect of the groups, as well as no interaction between the groups and the sounds: the ratings were similar in the two groups. An analysis of contrasts showed that the ratings of naturalness for the sounds H1, H2, and H3 were significantly larger than the ratings of the six other sounds ($p < 0.001$), and that the ratings of M1, M2, and M3 were significantly larger than the sounds L1, L2, and L3 ($p < 0.001$).

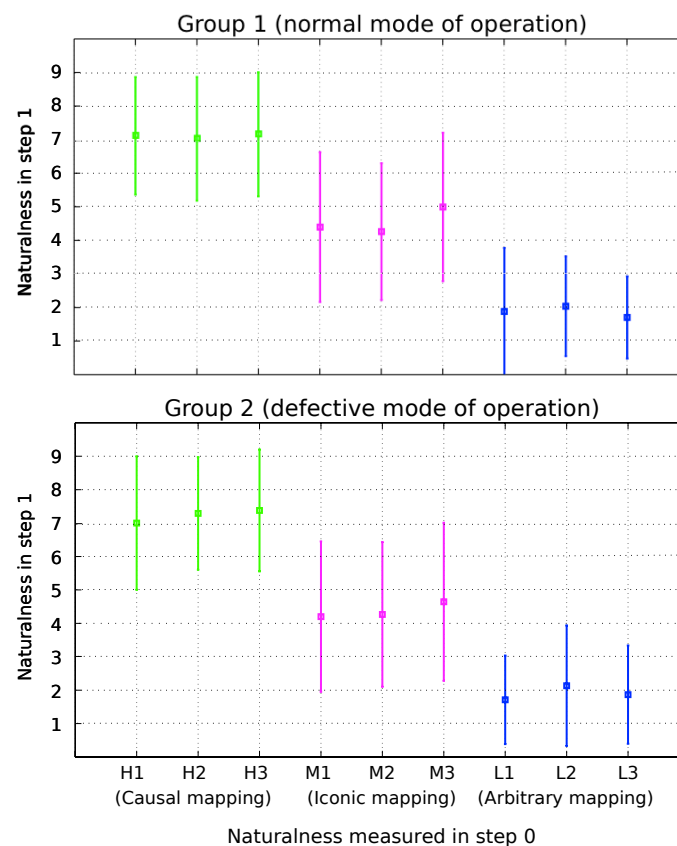


Figure 136. Pre-experimental ratings of naturalness (N, step 1), averaged across the 45 participants. Group 1 (upper panel, normal mode of operation) and Group 2 (lower panel, defective mode of operation). The vertical bars represent the standard deviation

IV.3.2.3.3.2 Analysis of the pre-experimental ratings (step1)

Test-retest reliability For the two groups of participants, we examined the test-retest reliability by computing the coefficient of correlation between the two measures for each scale. The coefficient of correlations were respectively $r = 0.83$, $r = 0.71$, and $r = 0.69$ ($N=180$, $p < 0.001$ in the three cases) for the three scales (N, U, and P).

The coefficients of correlation were statistically different from 0, but not very large. Furthermore, the participants reported that they were more confident in their second ratings. Therefore, we only used the second ratings (retest) in the subsequent analyses.

Raw results. Figure 137 displays the pre-experimental ratings on the three scales, averaged across the participants in Groups 1 and 2. The average values for the three scales are also presented in Table 16 for each level of naturalness. The figure and the table first show that the naturalness of the sounds had an influence on the three rating scales. It clearly shows that the ratings of naturalness (N) decreased (between the high and low levels of naturalness) with approximately the same amount (5.42 in Group 1, 6.24 in Group 2) as the ratings of usability (U; 4.94 in Group 1, 4.04 in Group 2), whereas the decrease of the ratings of pleasantness (P) was slightly less important, especially in Group 1 (2.09 in Group 1, 3.29 in Group 2).

Naturalness Level	Pre-exp. scale	Mode of operation	
		Normal (G1)	Defective (G2)
High	Naturalness	7.08 (1.76)	7.97 (1.01)
	Usability	7.02 (1.80)	7.37 (1.35)
	Pleasantness	4.86 (2.45)	6.73 (1.60)
	N	15	15
Medium	Naturalness	5.15 (1.77)	4.73 (1.82)
	Usability	5.24 (1.93)	5.73 (1.77)
	Pleasantness	4.82 (2.34)	4.86 (1.54)
	N	15	15
Low	Naturalness	1.66 (1.27)	1.73 (1.54)
	Usability	2.08 (1.53)	3.33 (2.53)
	Pleasantness	2.77 (2.10)	3.44 (2.55)
	N	15	15

Table 16. Pre-experimental ratings for the subjects in the two groups (G1: normal mode of operation; G2: defective mode of operation) on the three scales: average and standard deviation (in brackets)

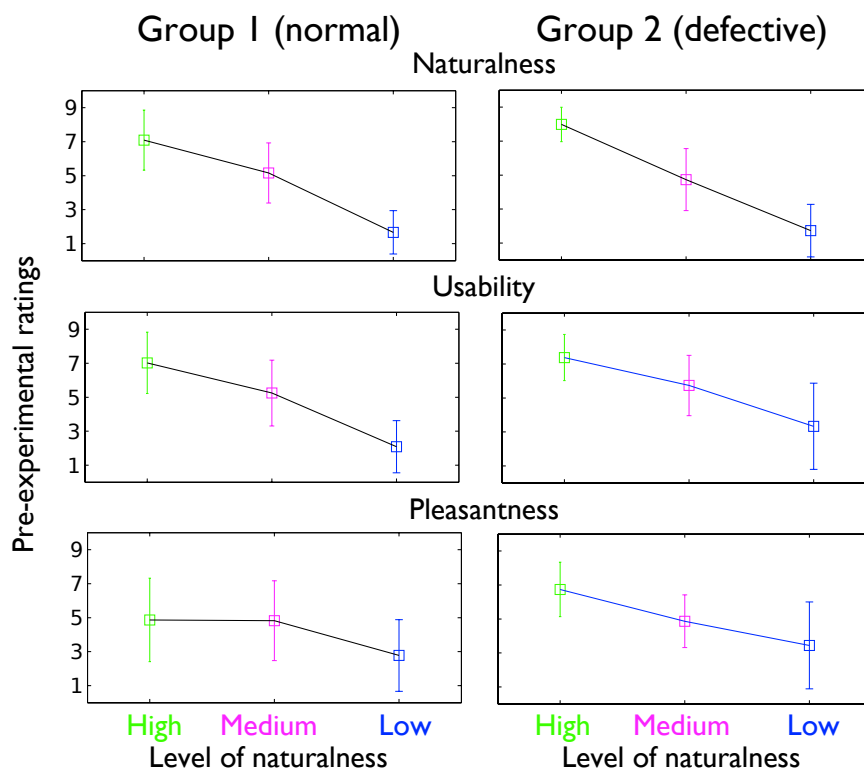


Figure 137. Pre-experimental ratings on the three scales for the sounds at the three level of naturalness. (High, Medium, Low), for the normal mode of operation (Group 1, left panel), and for the defective mode of operation (Group 2, right panel)

Analysis of variance. A multivariate analysis of variance (MANOVA) with repeated measures was conducted on the three dependent variables (N, U and P), using a one between-subject (the two groups), and one within-subject (the naturalness of the sonic feedback with 3 levels) full factorial design. A MANOVA was performed to take into account the potential correlation between the rating scales. The main interest here was to determine whether the naturalness factor globally affected on the ratings on the three scales.

The results of the MANOVA showed that the naturalness of the sounds was the only significant effect (Wilks' lambda value, $F=32.0$, $p<0.001$). Three two-way ANOVAs showed that the effect of the level of naturalness was significant for each scale. The percentage of total variance accounted for by each effect was indicated by the R^2 coefficients. The main effect of naturalness accounted for 58 and 32% of the total variance respectively for the scales U and P. Thus the strongest effect of the naturalness factor was obtained for the ratings on U.

IV.3.2.3.3.3 comparison between the pre- and post- experimental ratings (step 3)

The correlation between the pre- and post-experimental (Table 17) ratings of naturalness was relatively high ($R = 0.78$). This indicates that the ratings of naturalness were rather similar before and after the manipulation of the interface. The correlation between the pre- and post-experimental ratings of usability was lower ($R = 0.55$), as well as the correlation between the pre- and post-experimental ratings of pleasantness ($R = 0.34$). There was a slight difference between the two groups of participants for the ratings of pleasantness: the correlations between the pre- and post-experimental ratings were respectively 0.39 and 0.29 for Groups 1 (normal mode of operation) and 2 (defective mode of operation). Overall, these correlations suggest that the manipulation of the interface modified the user's appraisal of the sounds' usability and pleasantness. More details are provided by the analysis of the direct comparisons in step 4.

	Post-exp. N	Post-exp. U	Post-exp. P
Pre-exp. N	.78***	.54***	.54***
Pre-exp. U	.65***	.55***	.44***
Pre-exp. P	.47***	.37***	.34**

Table 17. Correlation matrix of pre- and post-experimental ratings (N = 90) for the naturalness (N), the usability (U), and pleasantness (P) scales. ***: $p < 0.0001$, **: $p < 0.01$

IV.3.2.3.3.4 Direct comparisons (step 4)

Results presentation of figure 138 represents the average ratings in step 4 (direct comparisons) for the two groups of participants (the two modes of operation). This figure shows how the participants assessed that the manipulation of the interface had changed their assessments of the sounds on the three scales. A positive value indicates an increase on a particular scale after the manipulation.

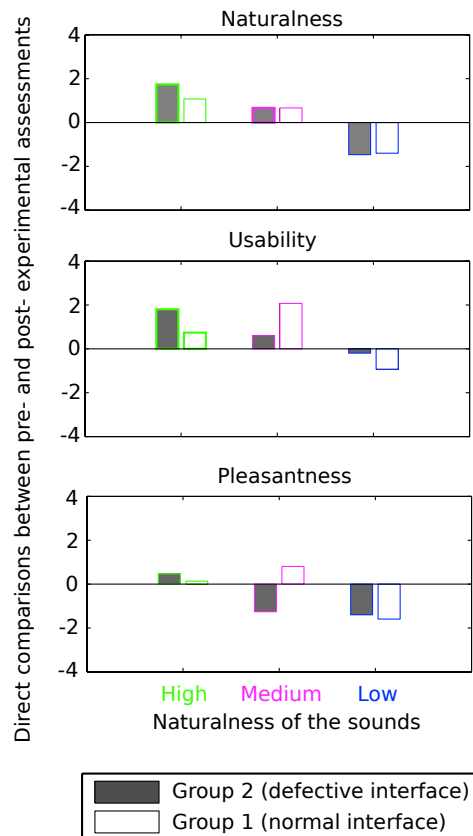


Figure 138. Direct comparisons of pre- and post-experimental assessments on 3 scales for the 2 groups

Analysis of variance. A multivariate analysis of variance (MANOVA) was conducted on the direct estimation of the difference between pre- and post-experimental assessment of the scales. The main interest was to compare the ratings obtained for the two groups, to examine the effect of the mode of operation (Factor 2) on the ratings of usability. The MANOVA revealed an overall significant effect of the naturalness (Wilks' lambda value, $F=4.42$, $p<0.001$) but no effect of the mode of operation. However, the interaction between the two factors was significant: the influence of the naturalness of the sounds on the ratings was modulated by the mode of operation of the interface.

Three ANOVAs revealed that the naturalness of the sonic feedback had a significant effect on the three scales. First, the manipulation of the interface exaggerated the assessment of the naturalness of the sounds (upper panel of figure 138): the sounds at the highest level of naturalness were judged more natural after the experiment than before, and the sounds at the lowest level were judged less natural after the experiment. Second, the sounds at the highest and medium level of naturalness were overall judged more usable after the experiment than before the experiment (middle panel of figure 138). The sounds at the lowest level of naturalness were judged equivalently usable before and after the experiment. Third, the sounds at the lowest level of naturalness were judged less pleasant after the experiment (lower panel of figure 138).

In addition, the analyses revealed a significant interaction between the naturalness of the sonic feedback and the mode of operation of the interface for the scales U ($F(2,84)=7.8$, $p<0.001$) and P ($F(2,84)=6.64$, $p<0.01$). A set of contrast analyses showed that the interactions were only statistically significant for the medium level (M) of naturalness (there was no significant difference between the two other levels of naturalness): the mode of operation of the interface affected the ratings of usability and

pleasantness only for the sounds at the medium level of naturalness (iconic mapping between the gesture and the sonic feedback). When the interface operated normally, the sounds with a medium level of naturalness were considered much more usable after the manipulation than before. This was not the case when the system was defective (see figure 138). At the medium level of naturalness, the sounds were considered more pleasant after the experiment than before for the participants using the normally functioning interface. However, the participants using the defective interface judged the sounds much less pleasant after the manipulation.

IV.3.2.3.3.5 Validation (step 5)

A final step was done in order to confirm, within the same session, previous results concerning the difference obtained for the high and the medium levels of naturalness. In fact, during the previous steps, results for the high and the medium levels of naturalness were obtained in different sessions in which participants manipulated the ATM with only one sound; each participant was assigned to a group corresponding to one of the three levels of naturalness, and differences were compared among the group of participants.

Results clearly showed that the sounds at the lowest level of naturalness (arbitrary mapping) were judged less pleasant and usable after the manipulation of the ATM. More surprisingly, results showed that ratings of usability and pleasantness were affected for the medium and high levels of naturalness by the mode of operation of the ATM; for the normal mode of the ATM, both levels were judged with more or less the same level of pleasantness, but for the defective mode of operation the medium level was judged significantly less pleasant. For the normal mode of the ATM, ratings of usability were higher for the medium level of naturalness than for the high one, and the opposite result was obtained for the defective mode. Thus, in order to confirm the effect of the mode of operation on the medium and high levels of naturalness, one sound at the highest level of naturalness (respectively causal mapping) and a second sound at the medium level of naturalness (respectively iconic mapping) were compared during the same session by each participant. This is the major difference with previous steps.

Thus, in this final step, twenty participants performed the same task as in step 2. Ten started with one sound corresponding to the high level of naturalness and were asked to evaluate the sound on two scales (respectively U and P), then they repeated the same procedure with the sound corresponding to the medium level of naturalness. Among the ten participants, five did the experiment with the normal mode of operation first, and then with the defective one; five did the opposite. The ten other participants started with the medium level of naturalness first.

The results are represented in figure 139. For the normal mode, results did not show any difference between the two levels of naturalness for both scales (resp. U and P), whereas, for the defective mode, results reveal a difference for both scales, but the effect was not significant. Only the interaction between the mode of operation and the level of naturalness for the scale of usability (U) was significant ($F(1,19)= 8.19, p<0.05$). This highlights that both levels of naturalness were rated with a similar level of usability for the normal mode, but not for the defective mode; the sound with the higher level of naturalness was judged more usable and the sound with medium level of naturalness is judged less usable. Even if the effect was not strong, this validation experiment confirmed the effect of the mode of operation on the medium and high levels of naturalness for the scale of usability.

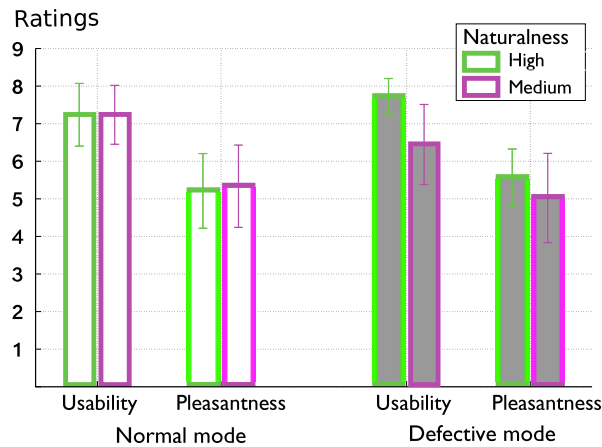


Figure 139. Step 5. Ratings on the two scales (Usability, Pleasantness) for the sounds at the two levels of naturalness (High, Medium), and for the normal and the defective mode of operations. The vertical bars represent the 95% confidence interval

IV.3.2.3.4 Discussion

The first conclusion of the experiment is that level of naturalness of the sounds had an important effect on the appraisal of the sounds. In step 1, before the users had a chance to manipulate the ATM interface (pre-experimental ratings), the sounds at the highest level of naturalness were rated as being more usable and more pleasant than the sounds at the lowest level of naturalness. This indicates that the participants a priori better appraised the causal than the arbitrary sounds.

The ratings measured after the participants had manipulated the interface in step 2 (post-experimental ratings in step 3 and direct comparisons in step 4) further qualified this interpretation. An interesting remark is that there was overall no significant difference between the two groups of participants. The mode of operation of the interface had no principal effect on the ratings of the sounds. This suggests that the participants were actually rating the sounds, and not the interface.

However, the manipulation of interface changed the user's appraisal of the sounds. The participants consistently indicated in step 4 that their appraisal had changed, but the changes depended on the levels of naturalness of the sounds. Overall, the sounds at the highest level of naturalness were rated as more natural and more usable after the manipulation than before the manipulation. The sounds at the lowest level of naturalness were rated as less usable and less pleasant after the manipulation than before the manipulation. Therefore, it can be concluded that the ratings in step 4 tended to exaggerate the pre-experimental ratings.

Several explanations can be proposed. First, the participants' longer exposition to the sounds during the manipulation of the interface might have emphasized their initial judgements. However, a closer look at the interaction between the mode of operation of the interface and the level of naturalness of the sounds shed light on an interesting phenomenon, and allows us to propose a second explanation. For the sounds with a medium level of naturalness (the sounds that bore an iconic relationship the action of striking a key), the effect of the manipulation was different for the two groups of participants. The participants in group 1, who used a normally operating interface rated these sounds as more usable and more pleasant after the experiment (similarly to the sounds with high level of naturalness). But the participants in group 2, who used a defective interface rated these sounds almost equivalently usable after and before the

experiment, and much less pleasant than what they initially thought. Therefore, this suggests that the exaggeration of the appraisal of the sounds was caused by the experience of the sounds in situation. For the sounds with the highest and lowest level of naturalness (i.e. the causal and arbitrary sounds), the manipulation of the interface confirmed the participants' initial appraisal. But the sounds with a medium level of naturalness (the iconic sounds) became less pleasant when the interface was defective. In other words, the appraisal of the sounds was contaminated by the negative appraisal of the interface: the iconic sounds were only tolerated when the interface worked properly. When the frustrating manipulation of the defective interface emotionally upset the users, they found the iconic sounds irritating and less pleasant. This effect did not occur for the causal (which were consistently positively rated) and the abstract sounds (consistently negatively rated).

IV.3.2.4. Conclusion

This study examined three questions. The first question focused on the influence of different kinds of mappings between a keystroke and the resulting sound on the perceived naturalness of the sounds. The results of the experiments clearly indicate that a causal relationship resulted in sounds perceived as natural, whereas an arbitrary relationship resulted in sounds perceived as not natural. By manipulating the relationship between the user gesture required during the interaction and the sounds played as feedback, one can control the perceived naturalness of the feedback sounds. The second question examined whether the naturalness of the sonic feedback had an effect on the pleasantness and perceived usability of the sounds. The results of the experiments also clearly show that this was the case: natural sounds were perceived as being more pleasant and more helpful than non-natural sounds.

Overall, the durations of the sounds (see Table 15) with a high and medium level of naturalness were similar (because the sounds at the medium level were created with the same temporal morphology as the sounds at the high level). In contrast, the sounds at the low level of naturalness were selected so as to be semantically and morphologically different from a keystroke. Accordingly, the durations for the low level of naturalness were longer. Therefore, the semantic and temporal mismatches are confounded in this case, and the negative effect of the low level of naturalness may be explained by either one or both of these aspects.

The third question examined the effect of manipulating the interface on the ratings of naturalness, pleasantness, and usability. Overall, manipulating the interface exaggerated the user's appraisal of the interface. After the manipulation, the users found the natural sounds more natural and more usable than before the manipulation. They perceived the natural sounds as more natural and usable than they had before the manipulation.

Two modes of operation were used during the experiment: a normal mode, and a defective mode. Comparing the results for these two modes of operation offers an interesting conclusion for the sounds with a medium level of naturalness sounds. These sounds were iconic. They were synthetic sounds that had the temporal envelope of an actual keystroke sound, but a different timbre. Overall, these sounds were rated halfway between the natural and the nonnatural sounds in terms of usability and pleasantness. However, the manipulation of the interface had a different effect on the ratings of these sounds, depending on whether the interface was properly working or not. In particular, the iconic sounds were rated as being slightly more useful after the manipulation of the

normal interface, as was the case with the natural sounds. But after the manipulation of the defective interface, they were rated as less usable, an effect opposite to the effect of the natural sounds.

These results provide sound designers with some interesting remarks. First, manipulating the mapping between the user's gesture and the sounds changes the perception of the naturalness of the sounds. It is in particular important to note that naturalness was not influenced by whether the sounds were synthetic or recorded sounds: in fact, the less natural sounds were recordings. Sounds that are perceived as the natural consequence of the user's gesture are especially desirable in user interfaces; they are judged as pleasant and useful. The synthetic iconic sounds used in the experiment were rated as more pleasant and useful than the non-natural recordings (even though less pleasant and useful than the non-natural sounds). As such, they offer an interesting alternative when technical reasons prevent the use of recordings. However, our results show they are more susceptible to be affected by the difficulty of the manipulation. The sounds were contaminated by the negative feelings resulting from the manipulation of a defective interface. In other words, these sounds received the blame for the non-working interface, which was not the case when using natural sounds as feedback.

Finally, these results emphasize that designers must test the quality of their sound design by having users actively manipulate the interface. Even with our simplified interface, our results make it clear that the manipulation of the interface greatly influences the appraisal of the sounds.

IV.4 Navigation dans un menu hiérarchique

IV.4.1 Préambule

L'objet d'étude est ici entièrement numérique et de type interface "homme-machine" (IHM) entre un individu (conducteur) et un dispositif informatique (système centralisé d'informations et de données d'un habitacle automobile). Dans ce cas, le contexte d'utilisation détermine l'enjeu, à savoir la navigation en aveugle – ou à l'oreille – dans la structure du système. En effet, la conduite automobile requérant une attention visuelle quasi-continue, l'usage de la modalité auditive pour la manipulation d'objets à l'intérieur de l'habitacle, et pendant la conduite, s'avère parfaitement pertinent et approprié. Ici, l'objet se présente donc sous la forme d'une interface permettant d'accéder à un ensemble de données organisées suivant un principe hiérarchique, c'est-à-dire avec des inter-relations de dépendance et d'héritage. L'hypothèse de travail consiste à postuler qu'un design sonore de cette structure arborescente va faciliter l'accès aux données et permettre une navigation dans le système sans altérer l'intégrité et la qualité de la tâche principale de conduite. Dans cet objectif, plusieurs stratégies de design sonore sont expérimentées et testées par rapport à des performances de navigation ou de reconnaissance d'objet.

Cette étude présente un caractère relativement emblématique vis-à-vis des concepts posés pour tenter de définir la discipline du design sonore (voir § IV.1). Elle est, en effet, le fruit d'une collaboration entre un constructeur automobile (Renault), un laboratoire de recherche (Ircam/équipe Perception et Design) et un compositeur,

Andrea Cera⁴⁸ et, à ce titre, illustre parfaitement le tryptique art/science/industrie évoqué précédemment. En outre, elle explore (et exploite) le champ scientifique et technologique de la sonification et de la synthèse sonore numérique. Enfin, en termes de création, elle présente à la fois des contraintes fortes de production sonore mais également des axes d'innovation qui ont permis au compositeur / designer sonore associé au projet (A. Cera) d'exprimer une partie de son inventivité et de sa créativité.

Les résultats de ce travail ont fait l'objet d'un chapitre dans l'ouvrage *Human-Computer Interactions in Transport* édité par Christophe Kolski (Université de Valenciennes) [Misdariis et al., 2011] et d'une communication scientifique à la 22^{ème} conférence francophone sur l'interaction homme-machine (IHM) [Langlois et al., 2010].

IV.4.2 Publication [Langlois et al., IHM conférence, 2010]

Evaluation de la Sonification d'un Système Multimedia Automobile

Sabine Langlois ^(a), Severine Loiseau ^(a), Julien Tardieu ^(b), Andrea Cera ^(b), Nicolas Misdariis ^(b)

^(a) Renault / Ergonomie cognitive – 1, av du Golf – 78288 Guyancourt, France

^(b) Ircam-CNRS STMS – 1, place Igor Stravinsky – 75004 Paris, France

IV.4.2.1 Introduction

Renault a souhaité évaluer l'intérêt de l'utilisation du son pour faciliter la navigation et la recherche de contenu dans les grandes bases de données des systèmes multimédia embarqués sur véhicule. On s'intéresse ici aux systèmes pilotés par une commande à distance, et présentés sous forme d'arborescence hiérarchique.

Quatre catégories de retours sonores ont été définies et ont fait l'objet d'un maquettage puis d'une évaluation auprès d'utilisateurs. Ces retours sonores ont pour objectif de faciliter l'apprentissage de la structure d'un menu, le repérage du positionnement et l'identification des items dans les longues listes de données, et de confirmer les actions réalisées par le conducteur, en limitant le recours à l'attention visuelle.

Ce document décrit le guidage sonore défini et maqueté, la méthodologie adoptée pour évaluer son efficacité et son acceptabilité ainsi que les résultats de cette évaluation et les perspectives du projet.

IV.4.2.2 Description du guidage sonore

Nous avons réalisé une maquette interactive, pilotable au moyen d'un pad gestuel. Elle permet d'accéder en théorie à différents menus d'un système multimédia : Communication, Navigation (au sens de fonction GPS), Véhicule (au sens d'informations sur le véhicule), Musique, Images et Vidéos. Dans le cadre de l'étude, seuls les menus Communication et Musique ont été développés, et rendus disponibles. L'arborescence est la suivante :

⁴⁸ <http://brahms.ircam.fr/andrea-cera>

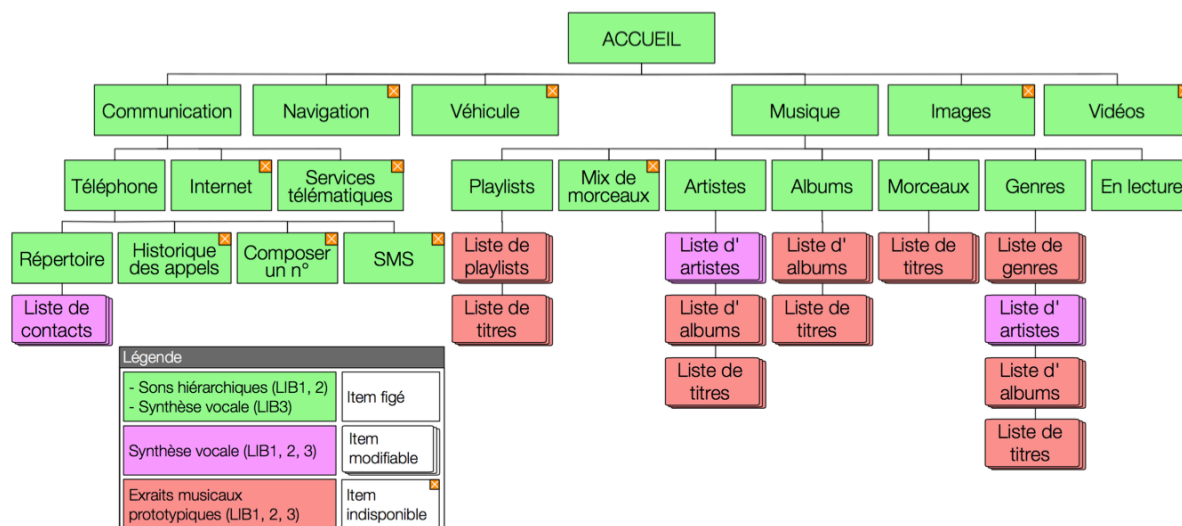


Figure 140. Arborescence des menus et principe d'utilisation des 3 librairies (LIB1, 2, 3) de sons

Trois librairies de sons sont activables sur la maquette. Nous les nommerons dans l'article LIB1, LIB2 et LIB3. Quatre catégories de sons sont utilisés dans les 3 librairies : (1) sons de confirmation d'action, (2) sons hiérarchiques (sons associés aux items du menu), (3) synthèse vocale et (4) échantillons prototypiques (extrait musical représentatif). Les sons de synthèse vocale ont trois variantes : une simple lettre (« Lettre »), le groupe de mots complet (par exemple « Services télématiques ») dit à vitesse normale (« Synthèse vocale »), ou dit à vitesse accélérée (« Spearcon »). Les 4 catégories de sons sont utilisées dans les 3 librairies comme indiqué dans la légende de la figure 140. Nous entendons par « item modifiable », un item que l'utilisateur de la version série de l'interface pourrait modifier. Ces items sont organisés en liste, dont la longueur nous a incités à prévoir les trois vitesses d'exploration évoquées ci-dessus : normale, accélérée (seulement dans les menus sonifiés avec la synthèse vocale) et rapide, sonifiée par la simple lettre dans toutes les listes.

Les sons de confirmation d'action ainsi que les sons hiérarchiques ont été créés dans le cadre d'une collaboration avec l'équipe Perception et Design Sonores de l'Ircam, en association avec le compositeur/designer sonore A. Cera.

Intégrée dans une démarche scientifique, la phase de création a été précédée d'une étude bibliographique qui a permis de faire un état de l'art des travaux relatifs à la sonification des menus hiérarchiques. D'un point de vue fonctionnel, ce problème soulève deux questions fondamentales par rapport à la modalité auditive : la représentation du niveau hiérarchique de l'arborescence et la représentation du contenu des items du menu. L'étude de la littérature permet de distinguer trois principales approches :

- « *earcon* », son synthétique pour lesquels la relation son/signification est arbitraire et doit donc être apprise par l'utilisateur ; cette notion a été introduite par Blattner [Blattner et al., 1989] puis, par la suite, étudiée et expérimentée notamment par Brewster [Brewster et al., 1995], Lepître [Lepître et al., 2000] et Helle [Helle et al., 2001]. L'intérêt des earcons est de pouvoir facilement représenter une structure hiérarchique en modulant les caractéristiques sonores (hauteur, timbre, rythme, durée, etc.).
- « *icône sonore* » (en anglais, « *auditory icon* »), introduite par Gaver [Gaver, 1986], qui établit un lien direct ou métaphorique avec l'objet ou le concept qu'il représente,

en faisant référence à un son de notre environnement, et donc de fait, quasi-directement compris par l'utilisateur. On peut trouver une mise en oeuvre de cette approche notamment dans les travaux de Barrass [Barrass, 1997] ou Conversy [Conversy, 2000].

- « *spearcon* » (contraction de « *speech* » et « *earcon* »), basé sur une synthèse vocale compressée dans le temps tout en maintenant une hauteur constante. La reconnaissance du mot ou groupe de mots n'est pas nécessairement assurée, mais le *spearcon* joue le rôle d'une empreinte univoque. Cette notion a été introduite par Walker [Walker et al., 2006] afin de pallier les limitations d'une sonification reposant sur des *earcons* ou des icônes auditives (difficulté de création et d'apprentissage, en particulier pour les *earcons*). Les travaux de Walker, entre autres ceux menés avec Palladino [Palladino et al., 2007] sur un menu de téléphone portable mettent en avant les performances d'apprentissage de *spearcons* (identiques à celles de la synthèse vocale ; 1,1 présentation suffit), comparées à celles des *earcon* (6,5 présentations sont nécessaires). Dingler [Dingler et al., 2008] ajoute les icônes auditives à la comparaison, pour des catégories d'objets : il faut 8,5 cycles d'apprentissage pour les *earcons*, 3 pour les icônes auditives et 1,1 pour les *spearcons* ou le vocal.

L'analyse comparée de ces différentes approches ainsi que le respect du cahier des charges de l'étude ont permis d'établir une stratégie de design sonore. Pour la navigation 3 principales fonctions ont été identifiées :

- la **différenciation** des menus multimédia : chaque menu est associé à une famille de sons qui doit permettre une bonne distinction entre les menus ;
- le **défilement vertical** : le son associé à un item doit indiquer son niveau de profondeur dans la structure (par exemple, indiquer que l'item « Répertoire » se trouve dans le sous-menu « Téléphone » sur menu « Communication ») ;
- le **défilement horizontal** : au sein d'un même niveau hiérarchique, les sons associés aux items doivent indiquer l'avancement dans la liste (par exemple, indiquer que « Répertoire » est le premier item de la liste du menu « Communication », tandis que « SMS » est le quatrième).

Pour l'ensemble des sons hiérarchiques, des solutions ont été proposées suivant deux directions.

D'une part, une approche conventionnelle, mélangeant les concepts d'*earcons* et d'icônes auditives (LIB1), chacun étant assigné à des fonctions spécifiques : des icônes sonores pour la différenciation, des *earcons* issus d'une synthèse sonore paramétrable par modélisation physique pour les défilements vertical et horizontal, avec en plus un procédé d'hybridation de ces 2 composantes qui favorise l'homogénéité du rendu. Nous appellerons par la suite cette librairie « hybride ».

D'autre part, une approche plus originale basée sur la notion de « musème » (LIB2), issue des travaux d'analyse musicologique de Philip Tagg [Tagg, 1982] et qui renvoie à l'idée d'une unité de sens exprimée dans un contexte musical. Pour le compositeur, il s'agit ici de représenter de manière métaphorique – grâce à un motif musical – chaque menu principal (fonction différenciation), la durée et la nature mélodique de ce motif étant ensuite modifiées afin d'assurer respectivement les fonctions de défilement vertical et horizontal et de couvrir ainsi la surface complète de l'arborescence.

IV.4.2.2.1 Sons de confirmation d'action (« *feedback sounds* »)

Ils servent à confirmer par le son l'action réalisée par l'utilisateur afin de lui permettre de savoir que son action est ou non correcte et prise en compte par le système. Trois

sons de confirmation d'action sont proposés. Ils sont identiques pour les 3 librairies.

- Son « butée » : ce son est joué dans les listes d'items pour indiquer à l'utilisateur que le focus (curseur) qui se déplace à l'écran sur les items de la liste est parvenu en butée.
- Son « défilement » : ce son est joué dans les listes d'items qui n'utilisent pas de sons hiérarchiques (dans les listes qui utilisent des sons hiérarchiques, aucun son de défilement n'est joué car le son hiérarchique a été créé de façon à avoir une attaque franche pour bien marquer le passage d'un item à un autre).
- Son « liste vide » : ce son d'erreur est joué dans les listes d'items lorsque l'utilisateur sélectionne (action « ok ») un item qui n'ouvre aucun écran. Il s'agit du même son que le son « butée ». En complément de LIB1, 2 et 3, la librairie LIB4 est constituée de ces seuls sons de confirmation d'action, car ce sont ceux habituellement présents sur les systèmes multimédia usuels tels que téléphones portables. Nous appellerons par la suite cette librairie « Guidage sonore Off » ou « Sans son ».

IV.4.2.2.2 Sons hiérarchiques

Ces sons sont associés aux différents items de menu de LIB1 et LIB2 afin de renseigner l'utilisateur sur l'univers fonctionnel visité (Communication, Navigation, Véhicule ...) et la profondeur verticale et horizontale de la structure du menu. L'objectif de ces sons est de faciliter l'apprentissage du contenu et de la structure du menu afin de permettre une navigation plus rapide dans le système.

Lorsqu'une liste de menu est ouverte, le son hiérarchique associé à l'item sur lequel est positionné le focus est joué. Lorsque l'utilisateur déplace le focus dans cette liste, le son hiérarchique de l'item atteint est joué. De plus, si l'utilisateur revient à la liste du niveau de menu précédent (action « Retour simple »), le focus se place automatiquement sur l'item précédemment sélectionné et le son associé à cet item est joué.

IV.4.2.2.3 Spearcons et synthèse vocale

L'intérêt de la synthèse vocale est de ne nécessiter aucun apprentissage. Dans la LIB3, la synthèse vocale remplace les sons hiérarchiques des librairies 1 et 2 (cf. Légende - figure 140). Dans les 3 librairies, nous utilisons la synthèse vocale dans les listes Répertoire et Artistes afin de faciliter le repérage dans les listes nominatives.

D'un point de vue pratique, la synthèse vocale est réalisée avec voix féminine française Alice de la synthèse TTS Acapela. Elle est précalculée dans le cadre de l'expérimentation, sachant que, sur véhicule de série, elle serait générée à la volée pour s'adapter à la nature dynamique du contenu de certains sous-menus (par exemple, la « Liste d'albums » du menu Musique). A partir de cette synthèse, nous avons produit les spearcons grâce au logiciel Audacity, avec une compression de tempo égale à 30% (<http://audacity.sourceforge.net/>). Un silence de 0,2 s. sépare le spearcon de la synthèse vocale.

IV.4.2.2.4 Extraits musicaux prototypiques

Le but est de faciliter la recherche de données musicales aux utilisateurs qui souvent ne connaissent pas les titres des morceaux et des albums disponibles dans leur librairie musicale. La durée des échantillons varie de 2 à 5 secondes. Pour les listes Playlists, Genres et Albums, les échantillons sont construits à partir d'un morceau représentatif de la playlist, du genre et de l'album survolé. Pour la liste Morceaux, les échantillons sont construits à partir d'un moment caractéristique du morceau. Ces extraits contiennent de la voix si le morceau est chanté, et, si possible, le moment où le titre est chanté.

IV.4.2.3 Evaluation

L'évaluation répond à 3 objectifs : vérifier, hors situation de conduite, la qualité des 2 librairies de sons hiérarchiques ; vérifier, en situation de conduite simulée, l'acceptabilité des sons ainsi que leur efficacité ; recueillir l'avis des participants sur les différents sons du guidage sonore.

Nos hypothèses à tester sont les suivantes : (1) la sonification permet de réaliser plus rapidement les tâches de navigation dans le système multimédia, tout en conduisant ; (2) elle limite des détournements de regard hors de la scène de conduite ; (2 bis) elle permet une meilleure focalisation de l'attention sur la conduite ; (3) les sons hiérarchiques sont plus acceptables sur le long terme que la synthèse vocale ; (4) les sons hiérarchiques de type musème sont plus acceptables sur le long terme que ceux de type hybride, car ils font moins directement référence à des sons environnementaux.

IV.4.2.3.1 Protocole d'évaluation

Participants. 24 personnes ont participé à l'évaluation du système (11 F., 13 H., entre 21 et 48 ans, tous salariés de Renault).

Installation. L'évaluation de la qualité des librairies hiérarchiques et le recueil de l'avis de participants sur le guidage sonore a été réalisé sur table. L'évaluation de l'acceptabilité et de l'efficacité des sons a été elle réalisée sur un simulateur afin de placer les participants en situation de conduite (roulage sur voie rapide).

Déroulement général. L'évaluation s'est déroulée en 7 étapes : {1} une présentation du test ; {2} une phase d'évocation des 2 librairies hiérarchiques (LIB1, 2) ; {3} une évaluation, à l'aide de la maquette (présentation sonore et visuelle), de la cohérence entre les sons hiérarchiques et les fonctions associées ; {4} un choix entre LIB1 et LIB2 ; {5} une familiarisation avec la maquette (+ pad gestuel) associée à la librairie hiérarchique choisie dans la phase précédente ; {6} la réalisation de 6 tâches courantes (e.g, appeler un correspondant) sur la maquette, en situation de conduite simulée, suivant 3 conditions expérimentales : avec la librairie choisie à l'étape 4 (LIB1 ou 2), avec la librairie vocale (LIB3), et avec le guidage sonore Off à l'exception des sons « défilement », « liste vide » et « butée » (LIB4); {7} un questionnaire permettant aux participants de verbaliser librement sur leurs préférences et leurs stratégies de jugement. La durée totale du test est d'environ 1h30 ; l'étape 6 dure en moyenne 30 minutes (10 minutes par condition).

IV.4.2.4 Résultats

IV.4.2.4.1 Qualité des librairies hybrides et musèmes

Phase d'évocation libre {étape 2}. Dans la librairie hybride, seul le son Téléphone est correctement associé à la fonction par la majorité des participants (74%). Le son Images est compris par 43% des participants, tandis que le son Communication évoque des fonctions telles que Téléphone et Internet liées au même univers fonctionnel Communication. Les autres sons sont eux rarement associés à la bonne fonction, et le son Navigation n'est identifié par aucun des participants. La majorité des sons évoquent des fonctions diverses. En revanche, près de la moitié des participants associe le son Musique à un son d'erreur (type Windows).

Dans la librairie musème, aucun son n'est correctement associé à la fonction qu'il représente. Le son Musique, le mieux identifié, l'est par 30% des participants. La plupart des sons sont associés spontanément à des sons d'alerte ou d'erreur.

Cohérence son-fonction {étape 3}. Cette étape correspond davantage à la manière dont

les utilisateurs découvriraient le système en situation réelle que l'étape d'évocation libre. Les sons sont ici globalement jugés cohérents avec la fonction qu'ils représentent, en particulier les sons hybrides Communication, Véhicule et Images et les sons musèmes Musique et Vidéos. En revanche, dans la librairie hybride, la moitié des participants considèrent que le son Musique n'est pas du tout adapté à la fonction. Dans la librairie musème, ce sont les sons Images et Communication qui sont considérés comme non adaptés à la fonction par la moitié des participants. Le tableau ci-dessous reprend ces résultats :

Librairie hybride (LIB1)	Librairie musème (LIB2)
Communication (63%) Véhicule (42%) Images (38%)	Musique (54%) Vidéos (54%)
Navigation (25% / 33%) Vidéos (21% / 21%)	Navigation (21% / 33%) Véhicule (25% / 33%)
Musique (63%)	Images (63%) Communication (50%)

Tableau 18. Evaluation de la cohérence son/fonction

Continuité du thème sonore dans les menus Communication et Musique {étape 3}. La continuité du thème sonore est jugée satisfaisante par la grande majorité des participants dans les menus Musique (83%), Communication (96% entre Communication et ses items, 92% entre Téléphone et ses items) de la librairie musème. En revanche, dans la librairie hybride, cette continuité du thème sonore n'est jugée satisfaisante par la majorité des participants que dans le menu Musique (96%). Si le 1^{er} niveau du menu Communication satisfait quant à lui 96% des participants, 42% des participants trouvent que le sous-menu Téléphone n'est pas homogène.

Dans les commentaires généraux formulés par les participants, il est intéressant de souligner que certains déplorent la perte du caractère évocateur des sons du dernier niveau de menu, impliquant une mémorisation difficile. Néanmoins, 29% des participants soulignent et apprécient que la profondeur du menu soit donnée par la simplification des sons à mesure que l'on s'enfonce dans la structure.

Préférence entre les librairies hybride et musème {étape 4}. La majorité des participants (79%) disent préférer la librairie hybride à la librairie musème (13%). La librairie hybride est préférée car elle est jugée plus explicite, suggestive, intuitive, ce qui facilite la mémorisation, l'apprentissage et la familiarisation avec le système. Notons que 8% des participants aimeraient toutefois faire un mix entre ce qu'ils ont préféré dans les deux librairies, retenant les sons Musique et Vidéos de la librairie musème et les sons restants de la librairie hybride.

IV.4.2.4.2 Efficacité du guidage sonore {étape 6}

Les 3 variables recueillies lors de cette étape sont : la durée chronométrée de réalisation de la tâche, le niveau de confort de conduite ressenti (échelle en 6 points allant de « très inconfortable » à « très confortable »), la fréquence de détournement de regard de la route ressentie (évaluée directement par le participant sur une échelle en 4 points allant de « rarement » à « très souvent »).

Des analyses de la variance, avec un intervalle de confiance fixé à 95%, nous permettent d'étudier l'influence qu'ont sur ces 3 variables numériques dépendantes les

variables explicatives catégorisées suivantes : « condition de sonification » (3 conditions : librairie vocale, librairie hiérarchique et sons OFF – sauf sons de confirmation d'action) et « tâches » (T1 : Appeler le coiffeur, T2 : Appeler Pizza Rapido, T3 : Lancer le morceau Divine idylle, T4 : Lancer le morceau Superstition, T5 : Lancer l'artiste Gare du Nord, T6 : Lancer l'artiste Ray Charles). Il s'agit d'un plan expérimental à mesures répétées, dont les variables inter sujets identifiées sont le genre et la classe d'âge. Des tests de Fisher ont permis de comparer les différentes modalités des variables (par exemple : sons hiérarchiques vs. vocal vs. sons OFF pour la variable « condition de sonification »).

L'analyse réalisée porte sur la dernière des 3 conditions testées par les participants afin de s'affranchir de l'influence de l'ordre de présentation et de se concentrer sur la situation où le sujet a essayé les divers guidages proposés par le système (dans un ordre contrebalancé entre les sujets) et a désormais bien en main le système. L'étude a été effectuée sur 22 des 24 participants (un participant n'a pu conduire sur simulateur, un autre avait une mauvaise vision).

Effet du guidage sonore sur la durée des tâches. Pour les 3 conditions de guidage, le graphe ci-dessous indique la moyenne et l'écart-type de la durée de chaque tâche, ainsi que pour l'ensemble des tâches (« T1-T6 »).

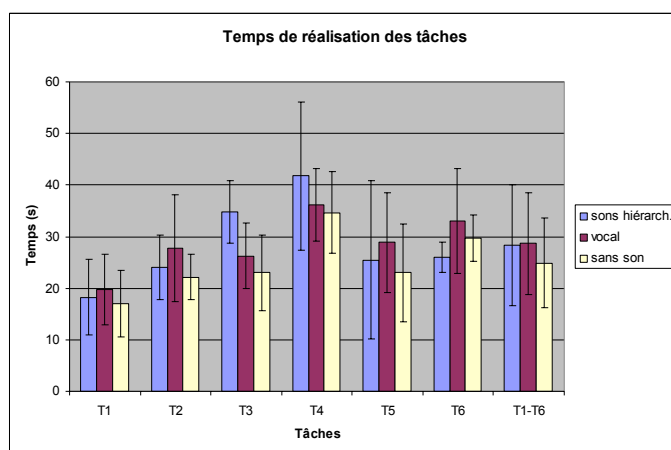


Figure 141. Temps de réalisation des tâches (moyenne +/- un écart type).
Par tâche (« T1 » à « T6 »), et en moyenne (« T1-T6 »).

Un effet tout juste significatif du guidage sonore sur la durée de réalisation des tâches est constaté ($p=0.047$). Le test de Fisher indique que la seule différence porte sur le vocal plus long que sons OFF ($p=0.02$). Les sons hiérarchiques sont similaires au vocal ($p=0.9$) et à sons OFF ($p=0.08$).

Si notre étude porte plus particulièrement sur l'influence de la condition de sonification, il est néanmoins intéressant d'examiner l'influence des autres variables. Celle de la tâche est significative ($p<0.0001$). La tâche 1 («appeler coiffeur») se révèle significativement plus rapide que les autres tâches ($p=0.003$ pour T1 vs. T2), et T4 (« lancer le morceau Superstition ») significativement plus longue ($p=0.009$ pour T5 vs. T8). La rapidité de réalisation de la tâche 1 peut s'expliquer par le fait que le chemin d'accès à la donnée recherchée («coiffeur») est le plus court des chemins d'accès parcourus.

Le genre n'a pas d'influence ($p=0.4$), en revanche la classe d'âge en a ($p<0.0001$), ce qui est une des explications de l'importance des écarts type.

Effet du guidage sonore sur le confort perçu. La condition de sonification a une

influence significative ($p < 0.0001$) sur le niveau de confort de conduite perçu.

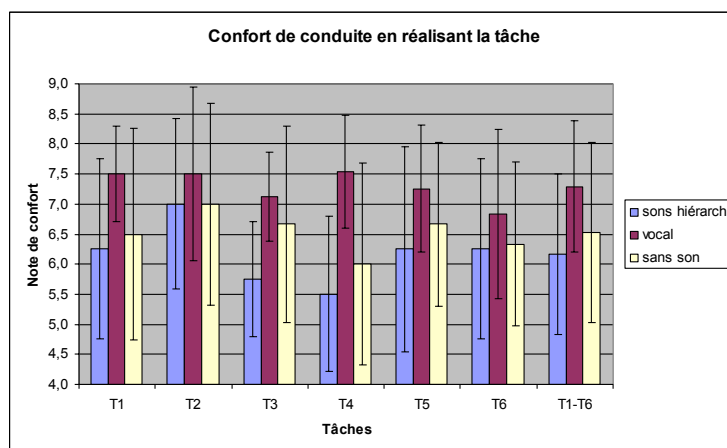


Figure 142. Confort perçu (moyenne +/- un écart type)

Les participants jugent en effet leur conduite plus confortable avec la librairie vocale que lorsque le guidage sonore est désactivé ($p < 0.0001$). En revanche, aucun effet significatif de la librairie hiérarchique n'est observé sur la perception de confort de conduite : le confort de conduite ressenti est similaire entre les conditions librairie hiérarchique et Sons OFF ($p = 0.2$; et $p = 0.002$ pour sons hiérarchiques vs. vocal).

Il n'y a ici pas d'influence de la variable « tâche » ($p = 0.5$). Le genre reste non influent ($p = 0.6$), et inversement pour l'âge ($p < 0.0001$).

Effet du guidage sonore sur le détournement du regard perçu. La condition de sonification influence la perception de détournement de regard ($p < 0.0001$). Les participants ont l'impression de détourner plus souvent leur regard de la scène de conduite lorsque le guidage sonore est coupé que lorsque le guidage vocal est activé ($p < 0.0001$).

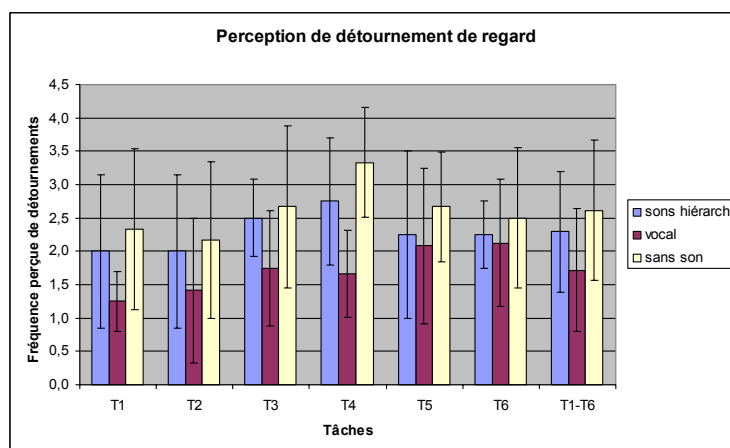


Figure 143. Détournement de regard perçu (moy. +/- un écart type)

En revanche, aucune différence significative n'est observée entre Sons OFF et la condition hiérarchique ($p = 0.2$, tandis que $p = 0.006$ pour les sons hiérarchiques vs. vocal). La tâche a un effet significatif ($p = 0.04$) : les tâches de téléphonie donnent aux participants l'impression de moins requérir de détournements de regard tandis que les tâches T5, T7 et T8 ($p = 0.05$ pour T2 vs. T8).

L'influence des facteurs inter sujets est similaire à celle observée sur les deux

précédentes variables dépendantes ($p=0.01$ pour l'âge, $p=0.7$ pour le genre).

Commentaires associés au confort de conduite et aux détournements de regard. Lors de l'étape 6, les participants ont expliqué les notes de confort et de détournement de regard attribuées après chaque tâche en invoquant plusieurs éléments du guidage sonore.

90% des participants soulignent leur intérêt pour les lettres alphabétiques associées au défilement lettre à lettre dans les listes. Ces lettres sont jugées utiles et limitent les détournements de regard, au profit d'une conduite plus sûre.

71% des participants disent préférer utiliser l'interface visuelle plutôt que les extraits musicaux. Ces derniers sont jugés inutiles car impliquent de connaître la musique disponible, nécessitent du temps pour les écouter alors qu'un coup d'oeil rapide vers l'écran permet de trouver la musique recherchée.

52% des participants disent ne pas apprécier les spearcons : incompréhensibles et donnant l'impression d'un bégaiement, ils sont source de confusion et de stress et ne constituent pas une aide pour l'utilisateur.

IV.4.2.4.3 Avis post conduite des participants {étape 7}

Satisfaction relative aux sons. Les librairies vocale et hiérarchique évaluées sont jugées de bonne qualité.

L'appréciation des sons hiérarchiques porte essentiellement sur les sons les mieux identifiés lors de l'étape 3 (Téléphone et Communication de la librairie hybride, Communication et Musique de la librairie musée). Inversement, les sons peu appréciés sont les moins évocateurs, hormis le son Véhicule de la librairie hybride, bien identifié mais mal jugé car trop familier (bruit de moteur).

Concernant les sons de confirmation d'action, les sons de défilement et d'erreur / butée sont jugés utiles et évocateurs et sont largement appréciés (respectivement 92 et 75%).

Les extraits musicaux satisfont 67% des participants car ils permettent notamment de pallier l'oubli du titre des morceaux et remplissent une fonction de confirmation.

Enfin les spearcons ne sont pas du tout appréciés (13% seulement de participants satisfaits) car, jugés incompréhensibles, ils perturbent la concentration sur la conduite.

Utilisation du guidage sonore sur le long terme. Une forte majorité (92%) indique que la sonification rend la conduite plus sûre, mais qu'elle ne permet pas de supprimer le visuel (83%).

Par ailleurs, 71% des participants pensent qu'ils continueraient d'utiliser le guidage sonore dans leur véhicule sur le long terme. De plus, 71% des participants préféreraient continuer d'utiliser la même librairie sonore plutôt que d'en changer pour conserver le bénéfice de l'apprentissage.

Quelques participants utiliseraient la librairie vocale lors de la phase de découverte du système (librairie plus explicite, ne requérant aucun apprentissage) puis utiliseraient sur le long terme la librairie hiérarchique, moins lassante, plus courte, plus agréable et plus personnalisée. Après l'appropriation du système, la librairie hiérarchique serait ainsi utilisée par 35% des participants, la librairie vocale par 30,5% des participants, et 30,5% des participants opteraient pour une librairie mixte vocale et hiérarchique, le vocal étant utilisé pour remplacer les sons non explicites du dernier niveau de menu. Seulement 1 personne préfère la condition non sonifiée.

IV.4.2.5 Discussion

Notre hypothèse de réduction de la durée de réalisation des tâches de navigation dans

le système multimédia grâce à la sonification n'est pas validée. Les sons hiérarchiques utilisés dans le haut du menu, joints à des sons vocaux dans les listes, n'apportent pas de gain par rapport au cas de la simple sonification par les sons de butée, d'erreur et de défilement. Le vocal présent dès le haut du menu entraîne quant à lui un allongement de la durée de réalisation des tâches. Sur les sons hiérarchiques, notre résultat est similaire à celui de Leplâtre [Leplâtre et al., 2000] qui a appliqué le principe des earcons hiérarchiques pour la sonification d'un menu de téléphone portable. Cette étude compare deux groupes (l'un dispose du menu visuel seul, l'autre du visuel et des sons) qui doivent effectuer des tâches de navigation sur le téléphone. Elle montre que la sonification diminue significativement le nombre d'appuis de touche et d'erreurs mais qu'en revanche elle ne change pas la durée entre deux appuis sur les touches (et donc le temps moyen pour effectuer la tâche). Notre résultat sur le vocal contredit celui de Palladino [Palladino et al., 2008], qui a comparé le temps de recherche dans une liste de 50 contacts téléphoniques, avec visuel, dans trois conditions : sans son, avec vocal, avec spearcon + vocal ; elle montre que la condition n'influe pas sur le temps de réalisation de la tâche.

Notre deuxième hypothèse de moindre perception de détournement de regard en présence de sonification, et son corollaire de meilleure focalisation sur la conduite, est validée, tant par les résultats recueillis à l'étape 6 (influence significative de la sonification vocale) qu'après la conduite, où 92% des participants estiment leur conduite plus sûre avec le son. Il serait intéressant, sur une maquette disposant d'une interface moins expérimentale et plus aboutie que celle utilisée ici, de collecter des données quantitatives de détournement de regard, à l'aide d'un oculomètre, afin de savoir si l'attention visuelle du conducteur est réellement moins sollicitée lorsque le guidage sonore est activé.

Si les sons hiérarchiques n'induisent pas en conduite un ressenti plus sécuritaire, les personnes indiquent a posteriori les considérer comme plus acceptables que le vocal : certains les adopteraient d'emblée, d'autres après une phase d'apprentissage de la structure du menu avec le vocal, si bien qu'à terme la majorité utiliserait les sons hiérarchiques (35% emploieraient la librairie hiérarchique et 30,5% une librairie mixte hiérarchique et vocale). Si la librairie vocale est préférée parce qu'elle facilite la découverte du système, aucune interprétation des prompts vocaux n'étant nécessaire, la librairie hiérarchique est in fine retenue car jugée plus agréable et moins lassante. Une librairie mixte, alliant sons hiérarchiques sur les premiers niveaux de menu et sons vocaux sur le dernier niveau de menu serait également appréciée car elle permettrait de pallier les difficultés de compréhension et de mémorisation des sons hiérarchiques du dernier niveau de menu. Notre hypothèse selon laquelle les sons hiérarchiques sont plus acceptables que le vocal est donc validée, pour les premiers niveaux du menu.

La majorité des participants (71%) affiche son intention de conserver à long terme le guidage sonore. Ce résultat, obtenu après une courte utilisation des sons (30 minutes en conduite environ) est à tempérer par le résultat recueilli par Helle [Helle et al., 2001] sur un téléphone portable sur lequel des earcons bien conçus ont été installés : au bout de trois semaines, 9 participants sur 17 avaient enlevé les sons, les jugeant gênants dans les lieux publics, trop longs et trop aigus. Certes le contexte est ici différent : le milieu automobile est moins ouvert sur l'extérieur, et la technologie sonore disponible sur un système multimédia embarqué est visiblement très supérieure à celle du téléphone portable utilisé par Helle. Une évaluation longitudinale (prêt d'un véhicule pendant une période définie) permettrait d'observer l'effet du guidage sonore sur

l'appropriation du système et d'étudier l'évolution de son utilisation sur le long terme. Par ailleurs, soulignons le souhait des participants de disposer de sons courts : nous avons créé les sons hiérarchiques avec ce principe de design, nous sommes donc incités à poursuivre dans cette voie.

Les participants n'apprécient pas les sons musèmes, sauf les mieux identifiés, même s'ils reconnaissent leur meilleure aptitude que les sons hybrides à assurer la continuité du thème sonore. Notre hypothèse selon laquelle ces sons seraient plus acceptables que les sons hybrides n'est donc pas validée. Plusieurs sons de la librairie hybride doivent néanmoins être retravaillés pour rendre la fonction qu'ils représentent plus évidente. C'est le cas notamment du son Musique qui évoque davantage un son d'erreur. Nous obtenons grâce à la comparaison entre ces deux librairies sonores le résultat suivant : l'acceptabilité des sons est directement liée à leur taux de reconnaissance, élevé lorsque le son entretient une relation forte avec l'item à représenter. Ce point justifie notre principe d'utiliser des composantes iconiques dans les sons hiérarchiques, et nous conduit à abandonner les sons hiérarchiques au profit du vocal dès que la relation son / item se relâche. Notre conclusion sur l'intérêt de la composante iconique rejoint les travaux de Dingler [Dingler et al., 2008] qui comparent la facilité d'apprentissage du vocal, des icônes auditives et des earcons.

Nous avons intégré les spearcons à notre système afin de proposer une augmentation de la vitesse de production sonore en cohérence avec la vitesse de défilement dans la liste. Cependant les spearcons sont rejetés car ils ne sont pas compris. Notre résultat contredit ceux de Palladino [Palladino et al., 2007] et de Dingler [Dingler et al., 2008] selon lesquels quasiment une seule présentation du spearcon suffit pour qu'il soit appris. Si les spearcons de Palladino sont d'un type différent des nôtres (compression logarithmique chez Palladino vs. linéaire), ceux de Dingler sont similaires, la qualité des spearcons que nous avons utilisés ne semble donc pas en cause. Quant aux travaux sur l'acceptabilité des spearcons, la seule étude portant sur ce sujet, celle de Walker [Walker et al., 2009] n'apporte pas d'élément clair.

Enfin, la grande majorité (83%) des sujets n'envisagent pas de se passer du visuel. Ce résultat est en cohérence avec les résultats de Palladino [Palladino et al., 2007 ; 2008], qui indiquent que le temps de recherche d'un item dans une liste est sensiblement plus court avec visuel que sans, et ce même après dix répétitions. L'analyse des verbatim recueillis en conduite nous permet de voir que le guidage sonore est principalement utilisé dans la phase de recherche initiale. Dans cette phase, les sons associés au menu (en particulier les sons hiérarchiques hybrides et la synthèse vocale, explicites) et les lettres alphabétiques permettent à l'utilisateur de s'approcher au plus près de l'information recherchée (par exemple, atteinte de la lettre D lors de la recherche du morceau « Divine Idylle »). Puis, lors de la phase d'affinement de la recherche (recherche de l'item « Divine Idylle » dans la lettre D) l'utilisateur privilégie souvent l'interface visuelle pour repérer et atteindre l'information recherchée : l'information est repérée d'un simple coup d'oeil. Dans cette phase, la synthèse vocale et les extraits musicaux se révèlent peu utiles. L'association d'un prompt vocal à l'ensemble des items survolés peut d'ailleurs être jugée lassante et donner l'impression d'un bégaiement du système lors du défilement rapide dans les listes (car la synthèse vocale est coupée lorsque le focus passe rapidement d'un item à l'autre). Néanmoins, la synthèse vocale et les extraits musicaux sont utilisés dans la phase finale, pour confirmer à l'utilisateur que l'item sur lequel il s'est arrêté est bien celui recherché.

IV.4.2.6 Conclusions et perspectives

Les résultats obtenus confirment l'intérêt du guidage sonore sur un système multimédia automobile. Bien que la sonification du système ne permette pas de naviguer et d'accéder plus rapidement aux données, les participants préfèrent la conserver car, au final, leur conduite leur apparaît plus sûre. Les sons proposés doivent cependant être évocateurs, et ce à tous les niveaux de l'arborescence. Le guidage sonore vocal leur donne l'impression d'avoir une conduite plus confortable et de devoir moins souvent détourner le regard de la route.

L'utilisation des lettres alphabétiques lors du défilement rapide dans les listes recueille une forte adhésion. Des travaux récents menés sur l'utilisation de lettres alphabétiques en synthèse vocale (Spindex) lors du défilement dans les listes, confirment l'efficacité de ce type de retour sonore. Mais la comparaison de différentes variantes de Spindex, effectuée par Jeon [Jeon et al., 2009], révèle que celle utilisée dans notre maquette (Spindex de type « Minimal », à savoir « A » prononcé sur le premier item commençant par A puis « B » sur le premier item commençant par B, sans son sur les items intermédiaires de A) est la moins appréciée. Il serait alors intéressant de nous orienter vers l'utilisation des variantes de Spindex les plus appréciées (comme celle de type « atténué » où le son A est prononcé à un niveau restreint sur les items intermédiaires de A) afin d'allier efficacité et acceptabilité.

Enfin, alors que l'intervention d'un designer sonore s'est révélée et demeure indispensable pour la création des sons hiérarchiques associés aux items du menu, la création des extraits musicaux et des prompts en synthèse vocale doit, elle, être automatisée pour que l'intégration de ces sons soit envisageable dans un véhicule. En effet, autant les items du menu n'évolueront pas (ils sont figés par le constructeur), autant le contenu des listes des menus Musique et Communication va être régulièrement modifié par les utilisateurs du véhicule.

IV.4.3 Extension de l'étude

Les premiers résultats obtenus lors de la phase initiale du projet de sonification d'une structure hiérarchique (voir § IV.4.2 précédent) ont ouvert la voie à des travaux complémentaires notamment dans le domaine de la mesure de l'attention visuelle et de l'altération de la qualité de conduite au cours de la navigation dans le système. Pour cela, un protocole expérimental basé sur une technique d'oculométrie a été mis au point et testé sur une stratégie de sonification issue des solutions proposées dans l'étude initiale par le compositeur/designer sonore Andrea Cera.

Cette extension d'étude, inscrite dans le cadre d'une poursuite de la collaboration avec le constructeur automobile (Renault), a été l'occasion d'un rapprochement scientifique avec un laboratoire de recherche disposant des moyens matériels et méthodologiques relatifs à la technique oculométrique (Pascal Gaillard, Julien Tardieu – Université de Toulouse 2 / PETRA / équipe Octogone). Les résultats de l'étude sont en cours de soumission dans le journal *Applied Ergonomics* [Tardieu et al., 2014]

IV.4.3.1 Rappel du cadre général

Le cadre général de l'étude est donc l'utilisation de la modalité auditive pour la transmission d'informations dans un paradigme de type interface homme-machine (IHM). En effet, dans ce contexte, la communication s'effectue le plus souvent sous une forme visuelle au moyen de représentations textuelles, graphiques, iconiques ou symboliques. Il est cependant admis que les représentations sonores (*auditory display*)

peuvent parfois jouer un rôle non négligeable et tirer profit de spécificités contextuelles telles qu'une présentation visuelle dégradée ou saturée, un état d'attention multiple, etc. [Deatherage, 1972 ; Gaver, 1986 ; Herman, 2008].

La présente étude est restreinte à un paradigme de navigation dans une structure hiérarchique et pose la question suivante : le son permet-il d'atteindre une cible – ou trouver un chemin – dans un menu arborescent contenant différents types d'objets, de manière efficace, rapide et sans détourner le regard d'un objectif visuel déterminé ? Elle fait suite à une première phase de formalisation, d'expérimentation et de conclusion partielle du problème comprenant :

- la définition de trois fonctionnalités de base régissant ce paradigme de navigation : la différenciation (des catégories entre elles), le repérage vertical (lien entre les classes d'une même famille) et le repérage horizontal (lien entre les objets d'une même classe) ;
- la définition – et l'implémentation – de trois stratégies de sonification, sur la base des concepts fondamentaux d'*earcon* et d'*auditory icon* [Blattner et al., 1989] : hybride (approche originale combinant les deux concepts de base), musée (approche de "musicalisation" des *auditory icons*, issue des travaux de Tagg [Tagg, 1982]), vocale (approche traditionnelle basée sur le procédé de *text-to-speech*) ;
- la mise en oeuvre d'une procédure expérimentale d'évaluation visant principalement à juger la qualité des sons (hors contexte), ainsi que leur efficacité et acceptabilité en contexte de conduite ;
- les conclusions partielles issues des résultats de l'expérience : une suprématie de la sonification par rapport à l'utilisation de la voix (au moins pour les premiers niveaux de la structure), une amélioration possible des concepts sonores et une évolution nécessaire du protocole expérimental pour recueillir des données objectives sur des variables importantes relatives au détournement d'attention.

IV.4.3.2 Nouveau protocole d'évaluation

Sur la base des données précédentes, une poursuite de l'étude a donc été réalisée au moyen d'une nouvelle approche expérimentale permettant de mieux quantifier l'impact de la sonification sur le comportement et la performance d'un individu. La technique oculométrique a été choisie pour atteindre cet objectif car elle semble pertinente et efficace dans le contexte automobile.

En premier lieu, et en prenant en compte les résultats précédents sur les potentialités d'amélioration de la qualité de sonification, une étape intermédiaire de nouvelle conception sonore a été effectuée par A. Cera, compositeur/designer sonore associé au projet. Cette phase a abouti à une seule et unique librairie de sons implémentée sur la base du modèle "hybride" (LIB1) intégrant différentes modifications sur les sons (p. ex., nouvelle sonification du menu "Navigation") ou sur la structure elle-même (suppression/ajout de branches de l'arborescence).

IV.4.3.2.1 Cadre et objectif

Dans le cadre de la technique de mesure par oculométrie, l'objectif de ce travail complémentaire est donc précisément de mieux évaluer l'impact de la sonification de la structure sur le mouvement du regard (*gaze movement*). Cette évaluation est réalisée en condition de laboratoire, au moyen d'une expérience de **double tâche** :

- une **tâche primaire** (T1) de détection visuelle de cible afin d'imposer une attention soutenue simulante, dans une certaine mesure, la tâche de conduite ;
- une **tâche secondaire** (T2) de navigation dans un menu matérialisée par une

interface graphique construite sur la base de la structure hiérarchique du système multimédia correspondant à l'application.

IV.4.3.2.2 Dispositif expérimental

Les sujets sont assis devant l'écran 17 pouces d'un dispositif de détection oculaire (*eye-tracker*) Tobii T120⁴⁹ sur lequel sont présentées les deux tâches (T1, T2) dans deux fenêtres séparées, et sont munis d'un casque qui diffusent soit les instructions, soit les balises auditives de la sonification (voir figure 144).

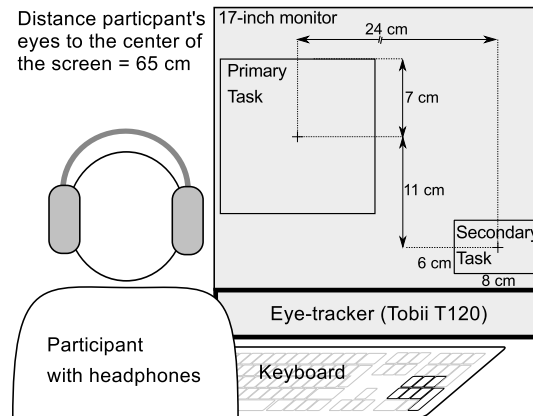


Figure 144. Dispositif expérimental et dimensionnement

Les sujets utilisent les touches dédiées d'un clavier pour naviguer dans le menu hiérarchique. Le dispositif de détection (*tracking*) est intégré à l'écran au bas duquel sont situés des caméras et des émetteurs infra-rouges. Pour une détection optimale, la tête des sujets doit rester face au centre de l'écran, à une distance approximative de 65 cm. ; le système admettant, malgré tout, une certaine tolérance de mouvement.

Les dimensions des éléments constituant les tâches et des fenêtres elles-mêmes sont déterminées en fonction de la configuration standard d'un habitacle automobile (voir figure 144). L'angle du regard entre le centre des deux fenêtres est le même que l'angle généralement rencontré entre le point de fixation de conduite du conducteur et la position du système multimédia, situé généralement au milieu du tableau de bord.

IV.4.3.2.3 Stimuli

La cible de la tâche primaire (T1) est un *smiley* jaune (voir figure 145), la tâche consistant à appuyer aussi vite que possible sur la barre d'espace du clavier lorsque celle-ci apparaît à l'écran. Certaines des cibles sont remplacées de manière aléatoire par un distracteur : un *smiley* inversé (voir figure 145).



Figure 145. Cible de la tâche primaire (T1) : *smiley* (gauche) et distracteur (droite)

L'intervalle de temps entre 2 stimuli visuels (cible ou distracteur) varie entre 1000 msec. et 2500 msec. par pas de 250 msec. ; le stimulus restant à l'écran pendant 800 msec. De plus, chaque stimulus apparaît aléatoirement à l'intérieur de la fenêtre associée à la tâche primaire. Le diamètre réelle de la cible est de 1 cm. Les couleurs de fond de l'écran sont réglées en gris avec une luminance de 160 cd/m².

⁴⁹ <http://www.tobii.com/fr/>

L'environnement Processing⁵⁰ est utilisé pour programmer l'interface visuelle, contrôler la présentation des stimuli, communiquer avec le *eye-tracker* pour recueillir les données de mouvements oculaires et assurer la synchronisation avec la tâche secondaire. Les données oculaires sont échantillonnées à une fréquence de 50 mesures par seconde.

Pour la tâche secondaire (T2), la structure et les éléments visuels du menu (couleurs, icônes, animations, etc.) sont semblables à ceux effectivement disponibles dans un véritable système multimédia automobile. La structure est cependant légèrement simplifiée et modifiée par rapport à celle correspondant à la première phase d'étude (figure 140, § IV.4.2.2) : seuls 2 menus sont développés pour l'expérience ("navigation" et "musique"), constituant ainsi 2 exemples réalistes d'interactions que le système offre au conducteur en situation de conduite (voir figure 146).



Figure 146. Interface de navigation (nouvelle version) utilisée pour la tâche secondaire (T2)
Haut-gauche : fenêtre d'accueil ("Home") ; haut-droit : menu "musique" ;
Bas-gauche : menu "navigation", bas-droit : menu "nouvelle destination".

Les sujets naviguent dans le menu en utilisant des touches dédiées sur le clavier permettant les mouvements haut/bas, droite/gauche et les actions de sélection, retour au menu précédent et réécoute des instructions. Le 1^{er} niveau du menu ("Home") présente 9 icônes qui peuvent être sélectionnées par un curseur (figure 146, haut-gauche). De plus, tous les menus – à part "Nouvelle destination" – sont constitués de listes verticales classées par ordre alphabétique. Le menu "Nouvelle destination" possède, quant à lui, un clavier virtuel sur lequel le sujet peut se déplacer. Dans ce menu, le sujet doit écrire un mot (p. ex., le nom d'une ville) en sélectionnant les lettres une à une (figure 146, bas-droite).

Cette tâche est programmée dans l'environnement Max/MSP (v5.1.7) qui permet à la fois la présentation du menu hiérarchique (avec / sans sonification) et le contrôle des paramètres de l'expérience : communication avec l'environnement Processing pour le déclenchement des cibles visuelles, mesure du temps de réaction, présentation des instructions de navigation, gestion des blocs d'expérience (entraînement et test).

IV.4.3.2.4 Plan d'expérience

Les variables indépendantes sont réparties dans un plan mixte comprenant :

- 1 facteur inter-sujets (*between subject*) : modalité sonore utilisée dans T2 (sonification ou silence),

⁵⁰ <https://www.processing.org>

- 2 facteurs intra-sujets (*within subject*) : longueur du chemin à effectuer dans T2 (court, moyen ou long) et absence/présence de la tâche secondaire T2 lors de la tâche primaire T1.

Les variables dépendantes sont constituées de différentes données collectées au cours de l'expérience : temps de réaction à la détection de la cible (TR), nombre d'erreurs ("omission" ou "fausse alarme"), temps de navigation ainsi que position du regard en fonction de 3 zones spécifiques (*areas of interest*) :

- AOI1 : à l'intérieur de la fenêtre associée à la tâche T1 ;
- AOI2 : à l'intérieur de la fenêtre associée à la tâche T2 ;
- AOI0 : en dehors des fenêtres de T1 et T2.

Les trois principales hypothèses de l'expérience sont les suivantes :

- H1 : la sonification augmente le temps passé à scruter la fenêtre de la tâche T1 ;
- H2 : la tâche T2 diminue la performance de la tâche T1 ;
- H3 : cet effet est plus important en condition silencieuse qu'en condition sonifiée.

IV.4.3.2.5 Procédure

Le principe de l'expérience est une tâche double dans laquelle il est demandé au sujet de privilégier la tâche primaire (T1). Une minute après le début de la tâche T1, une instruction vocale propose au participant de réaliser une tâche secondaire (T2) qui consiste à trouver un élément donné dans la structure du système.

L'expérience est constituée de 12 essais différents groupés en 3 blocs de 4 essais chacun : 3 éléments à chercher et un nom de ville à saisir (voir figure 147 pour illustration). Les 3 blocs sont présentés de manière aléatoire entre les sujets. Au début de chaque bloc, les instructions s'affichent sur l'écran et donnent le détail des 4 tâches de navigation/saisie à effectuer dans le bloc. Chaque essai se déroule ensuite de la manière suivante : i/ tâche simple (T1) ; ii/ tâche double (après 1 min.) où il est demandé d'atteindre un élément. Ces 2 étapes étant répétées 4 fois pour chaque essai d'un bloc. Une pause est par ailleurs prévue entre chaque bloc.

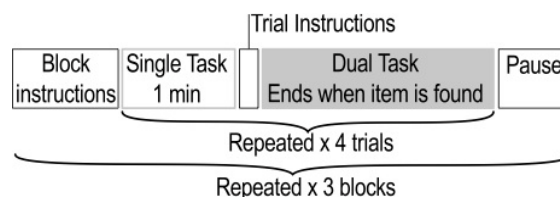


Figure 147. Schéma de principe du protocole expérimental

46 sujets (âge moyen 23 ans) ont pris part à l'expérience. La moitié d'entre eux ont réalisé la tâche T2 en condition silencieuse, l'autre moitié en condition sonifiée. En outre, une phase d'entraînement permet au sujet de se familiariser avec le contenu du menu et les commandes du système de navigation.

IV.4.3.2.6 Premiers résultats

Concernant la performance de la tâche primaire (T1), les résultats montrent une augmentation du temps de réaction (TR) lorsque le sujet est en condition de double tâche (figure 148). Une analyse de variance montre que cette différence de TR entre condition de simple tâche (T1) et de double tâche (T1+T2) est significative ($F(1,45)=30.02, p<0.001$).

Par ailleurs, les données montrent un comportement similaire quelque soit la condition de sonification – silence vs. sonifiée – (figure 148) et une analyse de variance confirme que l’effet de la condition de sonification n’est pas significatif ($F(1,45)=0.44$, $p=0.504$).

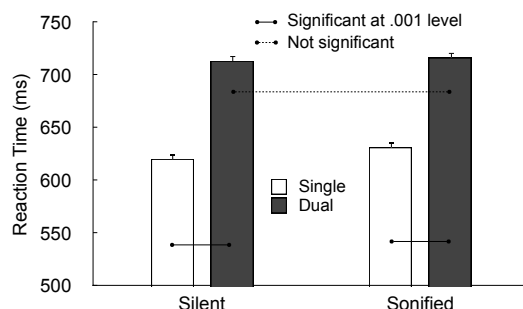


Figure 148. Temps de réaction de la tâche T1.
Fonction de la condition de sonification et de la nature de la tâche (simple/doube)

Pour la performance de la tâche secondaire (T2), les résultats donnant le temps de parcours dans l’arborescence (temps entre 1^{er} appui de touche après l’instruction et la sélection de l’objet-cible par le sujet) montrent logiquement une corrélation avec la longueur du chemin (figure 149); les différences entre le temps de navigation pour un chemin court, moyen ou long étant très significatives ($F(2,90)=57.09$, $p<0.001$). En revanche, même si les données montrent que les sujets mettent en moyenne plus de temps à atteindre leur objectif de navigation en condition sonifiée, plutôt qu’en condition silencieuse (figure 149), l’analyse de variance révèle que cette différence n’est pas significative ($F(1,45)=3.98$, $p=0.052$).

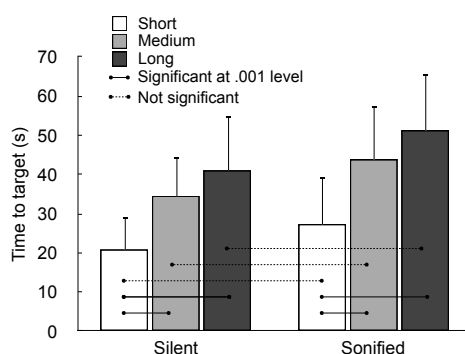


Figure 149. Temps de navigation au cours de la tâche secondaire (T2)
Fonction de la condition de sonification et de la longueur du chemin

Enfin, les résultats concernant les mouvements du regard, c’est-à-dire le pourcentage de temps passé à scruter chaque zone spécifique de l’interface (AOI1 : dans la fenêtre de T1 ; AOI2 : dans la fenêtre de T2 ; AOI0 : hors des fenêtres de T1 et T2) montrent une augmentation du temps passé dans AOI1 et une diminution du temps passé dans AOI2 en condition sonifiée par rapport à la condition silencieuse (figure 150). En outre, une analyse de variance révèle que ces différences sont fortement significatives : pour AOI1 ($F(1,39)=137.3$, $p<0.001$) et pour AOI2 ($F(1,39)=118.1$, $p<0.001$).

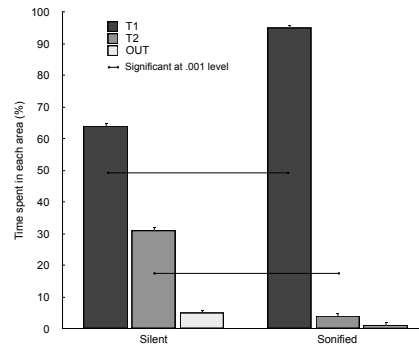


Figure 150. Pourcentage de temps passé dans chaque zone d'attention
Fonction de la condition de sonification et de la zone d'intérêt (AOI1 : T1 ; AOI2 : T2 ; AOI0 : ni T1/T2)

IV.4.3.2.7 Discussion

En premier lieu, les résultats de l'expérience montrent un impact fort de la sonification sur le temps passé dans la zone dédiée à chaque tâche : elle augmente pour T1 et diminue pour T2. Ce qui signifie que l'attention portée à la tâche primaire est renforcée – et, a fortiori, l'attention portée à la tâche secondaire est diminuée – par la sonification ; autrement dit, en condition sonifiée, le sujet se concentre davantage sur la tâche primaire, simulant la conduite, que sur la tâche secondaire de navigation. Ce résultat valide notre première hypothèse de travail (H1) et milite en faveur d'un complément auditif au système de type navigation en contexte automobile.

Cela étant, les résultats relatifs au temps de réaction montrent que la performance de la tâche T1 est malgré tout altérée par la configuration de double tâche (augmentation du temps de réaction) et ce quelque soit la condition de sonification. Ce résultat valide notre seconde hypothèse (H2) mais invalide la troisième (H3). De plus, au vu des conclusions précédentes, ce résultat peut paraître étonnant en condition sonifiée puisque, a priori, la navigation s'effectue principalement à l'oreille (peu de mouvements du regard en direction de la tâche T2).

Il peut néanmoins s'expliquer par une augmentation de la charge cognitive (*cognitive load*) en situation de double tâche ou bien, plus précisément, par un phénomène intermodal de type cécité – ou surdité – inattentionnelle, c'est-à-dire une difficulté à détecter un événement visuel lorsque l'on est engagé dans une tâche auditive, ou inversement (voir [Most et al., 2005] et [Dalton et al., 2012] pour des études sur ce type de phénomènes dans la modalité respectivement visuelle et auditive).

Par ailleurs, les données relatives au temps de parcours pour atteindre l'objectif de navigation démontrent que la condition de sonification n'a pas d'influence sur la tâche T2. Autrement dit, la sonification ne facilite pas la navigation dans la structure hiérarchique. Les raisons de cette absence de "valeur ajoutée" de la modalité auditive proviennent certainement de la consigne expérimentale donnant la priorité à la tâche primaire et ne faisant, en aucun cas, mention d'une quelconque performance ou rapidité de navigation. Ce résultat met en évidence l'importance des instructions données au sujet et le biais expérimental qu'elles peuvent induire dans l'interprétation des résultats.

IV.4.3.3 Conclusion

Grace à l'oculométrie, ce complément d'étude permet d'apporter des premiers éléments sur le fait que la sonification d'une interface de navigation – comme par exemple, celle d'un système multimédia automobile en interaction avec le conducteur –

peut s'avérer d'une aide non négligeable pour réduire la distraction visuelle en conduite. Sur cette base, nous proposons des règles simples de design sonore : les *auditory icons* doivent plutôt être utilisées lorsque les éléments sont faciles à représenter, et les *earcons* lorsque des indices hiérarchiques sont requis ; enfin les messages vocaux sont plutôt à dédier aux objets de type listes ou sous-menus.

L'étude pourrait maintenant se prolonger en portant une attention particulière à la tâche primaire, simulant la situation de conduite. En effet, le modèle choisi pour l'expérience (détection d'une cible visuelle) peut simuler, en première approximation, une attention de même nature que celle exigée pour la conduite mais ne satisfait en aucun cas à tous les types de stimulation auxquels un conducteur est soumis en condition de roulage.

Du point de vue expérimental, l'étape suivante pourrait donc impliquer un véritable simulateur de conduite ; l'utilisation de ce genre de dispositif soulève néanmoins de nouvelles questions méthodologiques (écriture des scénarios, durée des tests, etc. ...) qu'il sera important de poser et de résoudre.

IV.5 Conclusion

Cette quatrième partie relative à la thématique du **design sonore** a donc permis dans un premier temps de tenter d'élaborer une définition de la discipline ou, tout au moins d'en dessiner les contours. Sur des bases étymologiques, historiques et bibliographiques, le design sonore s'avère donc être une conception sonore appliquée à différents secteurs d'activité (produits manufacturés, architecture/urbanisme, cinéma/spectacles vivants, etc.) s'attachant à donner une forme et un sens à des sons intentionnellement créés pour répondre à une fonction ou une esthétique donnée.

De plus, la partie introductive montre que cette thématique se positionne clairement comme un domaine de recherche à part entière en développant une approche méthodologique qui articule une relation entre des connaissances scientifiques et une pratique artistique intuitive et singulière, et en fédérant également une communauté de chercheurs et créateurs de plus en plus importante et de mieux en mieux structurée.

Le développement d'une partie des travaux personnels a pour vocation d'alimenter ce domaine en résultats expérimentaux et de contribuer ainsi à la réflexion autour des questions de recherche et de pratique en design sonore. Ils présentent, notamment, le bien fondé d'une telle approche dans le contexte des interfaces homme-machine (clavier ATM ou système multimédia embarqué) en s'appuyant sur le postulat que la modalité auditive peut, dans certains cas, compléter voire suppléer la modalité visuelle pour transmettre des informations ou des sensations globales. Certains de ces travaux instruisent également la délicate question de la capacité à innover dans le domaine du sonore, c'est-à-dire proposer des solutions inédites qui respectent les règles et contraintes inhérentes à un contexte donné, tout en s'en affranchissant et en défrichant des territoires ou des processus sonores inexplorés (clignotant).

Parmi d'autres aspects, cette dernière ouverture représente une perspective originale et motivante dont l'étude pourrait permettre d'éclairer et d'examiner de manière approfondie les relations possibles entre créativité et design sonore. Dans ce cadre de travail renouvelé, les questions fondamentales posées pourraient alors se formuler de la manière suivante : comment concilier conception innovante et conception sonore ? *L'inoui* est-il possible en design sonore ?

C'est notamment sur cette ligne d'action – et dans la mesure du possible – qu'une partie de mes prochaines recherches pourra être axée, tout en conservant et poursuivant les thèmes abordés jusqu'à présent, et qui constituent ma principale expérience et mon savoir-faire dans le domaine.

Conclusion

Conclusion

Ce document fait état de plusieurs travaux dans différents axes de recherche articulés autour de grands thèmes – et processus – de l’acoustique : i/ l’émission du son avec la synthèse sonore par modélisation physique, ou comment concevoir des sons avec des moyens informatiques et des connaissances sur la physique des instruments ; ii/ la propagation du son avec l’étude du rayonnement acoustique et sa modélisation par des moyens électroacoustiques, ou comment rendre au haut-parleur sa dimension spatiale en redonnant accès aux paramètres de directivité et en réduisant l’écart entre l’instrument réel et sa version amplifiée ; iii/ la perception du son avec l’étude du timbre des sons musicaux et environnementaux, ou comment comprendre et modéliser les mécanismes perceptifs et cognitifs qui sous-tendent notre appréhension du monde sonore quotidien ou qui nous permettent de juger de la qualité d’un instrument. En outre, un quatrième volet présente un domaine d’application directement ouvert à ces thématiques : le design sonore, ou comment concevoir des sons fonctionnels et esthétiques sur la base de connaissances et de règles établies scientifiquement, et le plus souvent associées à des contraintes technologiques, économiques ou sociales fortes.

Les travaux présentés – parfois *in extenso*, s’ils ont fait l’objet d’une publication scientifique – sont, en premier lieu, systématiquement encadrés par une problématique générale et un état de l’art du domaine afin de les restituer dans un contexte théorique et, le cas échéant, méthodologique ou technologique. Ils ont, par ailleurs, pour objectif principal de démontrer une contribution personnelle aux grandes thématiques évoquées ci-dessus, en détaillant les résultats obtenus et les perspectives offertes. Parmi ces résultats, on peut noter de manière sélective : la contribution à l’étude de la physique de l’anche libre et l’implémentation d’un modèle physique d’accordéon dans l’environnement de synthèse sonore *Modalys*, l’étude du rayonnement d’instruments à vent ainsi que la réalisation de plusieurs prototypes fonctionnels d’un système multi haut-parleurs à directivité contrôlable, par ailleurs utilisés dans différents contextes artistiques et musicaux (installations sonores, pièces d’étude, oeuvres mixtes), l’intégration des études sur le timbre musical dans l’implémentation d’un moteur de recherche par similarité pour une base de données sonores, la contribution au développement des connaissances sur le timbre des sons environnementaux par une méta-analyse de plusieurs études spécifiques mettant en évidence une relative universalité de certaines dimensions perceptives, ou bien encore l’apport de données scientifiques concernant la valeur ajoutée d’un design sonore d’interface homme-machine en termes d’attention et de concentration visuelle.

La rédaction de ce document a donc été l’occasion de faire un bilan sur une activité de recherche pluridisciplinaire, déployée sur plusieurs années et construite autour d’un objectif de développement d’une discipline appliquée à la conception et au *design* de sons. Cet exercice d’intégration, et de réflexion, a également permis d’entrevoir de possibles suites d’étude – naturellement contenues dans les conclusions et perspectives des différents travaux présentés – et, de manière plus prospectif, d’imaginer de potentiels axes de travail et de recherche pour les années à venir.

Bibliographie

Bibliographie

Abdi H. (1987). Introduction au traitement statistique des données expérimentales. Presses Universitaires, Grenoble.

Adrien J. M. (1988). Etude des structures complexes vibrantes – application à la synthèse par modèles physiques. Thèse de Doctorat, Université Paris VI.

Adrien J. M. (1991). The Missing Link: Modal Synthesis. *In* Representations of Musical Signals, G. De Poli, A. Piccialli, and C. Roads Eds. MIT Press, Cambridge, Massachusetts, chap. 8, pp. 269-297 [ISBN 0-262-04113-8]

Arabie P., Shepard R. N. (1973). Representation of similarities as additive combinations of discrete, overlapping properties. *In* Mathematical Psychology Meetings, Montreal.

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

Aures W. (1985). Der sensorische wohlklang als funktion psychoakustischer empfindungsgrößen. *Acta Acustica united with Acustica*, 58(5), 282-290.

Avizienis R., Freed A., Kassakian P., Wessel D. (2006). A Compact 120 independent element spherical loudspeaker array with programable radiation patterns. *In* Audio Engineering Society Convention 120, AES.

Baalman M. A. J. (2003). Application of Wave Field Synthesis in the composition of electronic music. *ICMC , International Computer Music Conference*, Singapore.

Backus J. (1963). Small-Vibration Theory of the Clarinet. *The Journal of the Acoustical Society of America*, 35(3), 305-313.

Barrass S. (1997). Auditory information design. Doctoral disseration of Philosophy, Australian National University.

Barsalou L. W. (1999). Perceptions of perceptual symbols. *Behavioral and brain sciences*, 22(04), 637-660.

Barthélemy J.P., Guénoche A. (1988) : "Les arbres et les représentation des proximités". Masson éd., Paris.

Bassereau J. F., Duchamp R. (1996). Du Différentiel Sémantique au profil sensoriel. *Revue Sciences et Techniques de la Conception*, 5, 45-74.

Benade A. (1990). Fundamentals of Musical Acoustics. New York: Dover Publications. (*originally published 1976*)

Behler, G. K. (2006). Sound source with adjustable directivity. *The Journal of the Acoustical Society of America*, 120(5), 3224-3224.

Behler G. K. (2007). Sound source for the measurement of room impulse responses for auralization. ICA, 19th International Congress on Acoustics, Madrid.

Behler G. K., Pollow M. (2008). Variable Source-Directivity Using Dodecahedron-Loudspeakers. *The Journal of the Acoustical Society of America*, 123(5), 3614-3614.

Bensoam J., Misdariis N., Vergez C., Caussé R. (2001). Finite element method for sound and structural vibration : musical application with Modalys sound synthesis program based in modal representation. SCI, Systemics, Cybernetics & Informatics, Orlando, FL.

Berkhout A. J., de Vries D., Vogel P. (1993). Acoustic control by wave field synthesis. *The Journal of the Acoustical Society of America*, 93(5), 2764-2778.

Bezaf, M. C. (2007). Perception des bruits d'impact – Application au bruit de fermeture de porte automobile. Thèse de doctorat, Université de Provence – Aix-Marseille I (2007).

Bismarck (von) G. (1974). Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acta Acustica united with Acustica*, 30(3), 146-159.

Bijsterveld K., Krebs S. (2013). Listening to the Sounding Objects of the Past: The Case of the Car. *In* Franinovic K. & Serafin S. (Eds), *Sonic Interaction Design*, MIT Press, 3-25.

Bissinger G. (2004). Contemporary generalized normal mode violin acoustics. *Acta Acustica united with Acustica*, 90(4), 590-599.

Blattner M. M., Sumikawa D. A., Greenberg R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1), 11-44.

Blauert J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.

Blauert J., Jekosch U. (1997). Sound-quality evaluation – A multi-layered problem. *Acta acustica united with acustica*, 83(5), 747-753.

Blum T., Keislar D., Wheaton J., Wold E. (1997). Audio databases with content-based retrieval. *In* *Intelligent multimedia information retrieval* (pp. 113-135), MIT Press.

Bogaards, N., Röbel, A., & Rodet, X. (2004). Sound analysis and processing with audiosculpt 2. ICMC, International Computer Music Conference, Miami, USA.

Bonnet, C. (1986). *Manuel pratique de psychophysique*. A. Colin.

Bonnet M. D. (2002). La simulation numérique : le rôle des modèles formalisés, à la frontière de la recherche scientifique et de la création musicale. Thèse de Doctorat, Ecole des Hautes Etudes en Sciences Sociales, EHESS.

- Bouasse H. (1929). Instruments à vent. Librairie Delagrave, Paris.
- Bouasse H. (1931). Tourbillons, Vol. 1. Librairie Delagrave, Paris.
- Boulez P. (1983). L'acoustique et la musique contemporaine. ICA, 11th International Congress on Acoustics, vol. 8, 213-216.
- Bregman A. S. (1994). Auditory scene analysis : the perceptual organization of sound. MIT Press.
- Bresciani J. F. (1996). Etude du rayonnement des instrument de musique en vue du contrôle de la reproduction par des haut-parleurs. Thèse de Doctorat, Université du Maine, Le Mans
- Brewster S. A., Wright P. C., Edwards A. D. (1995). Guidelines for the creation of earcons. *In Adjunct Proceedings of HCI, Human-Computer Interaction*, Huddersfield, UK.
- Bruti L., Cuccu D., Gaetani R. (2005). Method for reproducing the sound of an accordion electronically. U.S. Patent No. 6,946,594. Washington, DC.
- Busson S. (2006). Individualisation d'indices acoustiques pour la synthèse binaurale. Thèse de Doctorat, Université de la Méditerranée-Aix-Marseille II.
- Cabe P. A., Pittenger J. B. (2000). Human sensitivity to acoustic information from vessel filling. *Journal of experimental psychology: human perception and performance*, JEP, 26(1), 313-324.
- Caclin A., McAdams S., Smith B. K., Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America*, 118(1), 471-482.
- Cadoz, C., Luciani A., Florens J. L. (1984). Responsive Input Devices and Sound Synthesis by Simulation of Instrumental Mechanisms: The CORDIS System. *CMJ, Computer Music Journal*, 8(3), 60-73.
- Campbell M. (2013). Objective evaluation of musical instrument quality: a grand challenge in musical acoustics. *International Congress on Acoustics, ICA, Montreal*.
- Caramiaux B., Susini P., Bianco T., Bevilacqua F., Houix O., Schnell N., Misdariis N. (2011). Gestural embodiment of environmental sounds: an experimental study. *In Proceedings of the International Conference on New Interfaces for Musical Expression, NIME 2011*.
- Carello C., Anderson K. L., Kunkler-Peck A. J. (1998). Perception of object length by sound. *Psychological science*, 9(3), 211-214.
- Carroll J., Chang J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika*, 35, 283-319.

Caussé R., Misdariis N., Thoréton S. (1997). Caractérisation objective de la qualité d'un instrument de musique. Actes du 4ème Congrès Français d'Acoustique, 1, 621-624.

Caussé R., Misdariis N., Ricot D. (1999a). Studies of accordion reed vibrations – Applications in sound synthesis. The Journal of the Acoustical Society of America, 106(4), 2286-2287, Proceeding of the Acoustical Society of America meeting, Columbus.

Caussé R., Begnis C., Misdariis N. (1999b). Assessment of musical instrument quality criteria : the notion of openness for a trumpet. The Journal of the Acoustical Society of America, 105(2), 1215-1216, Proceedings of European Acoustics Association convention (Forum Acusticum)/ Acoustical Society of America meeting, Berlin.

Chaigne A., Kergomard J. (2008). Acoustique des instruments de musique, Eds Belin.

Chion M. (1994). Audio-Vision: sound on screen. Columbia University Press, New York.

Cocron P., Bühler F., Franke T., Neumann I., Krems J. F. (2011). The silence of electric vehicles – blessing or curse. *In* Paper accepted to appear in Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC.

Conversy S. (2000). Conception d'icônes auditives paramétrées pour les interfaces homme-machine. Thèse de Doctorat, Université de Paris-Sud XI, Orsay.

Cook P., Essl G., Tzanetakis G., Trueman D. (1998a). Multi-speaker Display Systems for Virtual Reality and Spatial Audio Projection. *In* Proceedings of International Conference on Auditory Display, ICAD, Glasgow, UK.

Cook P., Trueman D. (1998b). A database of measured musical instrument body radiation impulse responses, and computer applications for exploring and utilizing the measured filter functions. *In* Proceedings of the International Symposium on Musical Acoustics, Leavenworth, Washington (pp. 303-308).

Corteel E. (2004). Adaptations de la Wave Field Synthesis aux conditions réelles. Thèse de doctorat, Université Pierre et Marie Curie, Paris VI.

Cottingham, J. P. (2011). Acoustics of free-reed instruments. Physics Today, 64(3), 44-48.

Cottingham J. P. (2007). Reed Vibration in Western Free-Reed Instruments. *In* Proceedings of the International Congress on Acoustics, ICA, Madrid.

Dalton P., Fraenkel N. (2012). Gorillas we have missed: Sustained inattentional deafness for dynamic events. Cognition, 124(3), 367–372.

Daniel J. (2000). Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimedia. Thèse de doctorat, Université Pierre et Marie Curie, Paris VI.

De Vries D. (1996). Sound reinforcement by wavefield synthesis: Adaptation of the synthesis operator to the loudspeaker directivity characteristics. *Journal of the Audio Engineering Society*, 44(12), 1120-1131.

Deatherage B. H. (1972). Auditory and Other Sensory Forms of Information Presentation. *In* Van Cott H. P. & Kinkade R. G. (Eds.), *Human Engineering Guide to Equipment Design*, 123-160 (Revised Edition).

Derogis P., Causse R. (1994). Caractérisation du rayonnement de la table d'harmonie d'un piano droit. *Le Journal de Physique IV*, 4(C5), C5-609.

Derogis P. (1997). Analyse des vibrations et du rayonnement de la table d'harmonie d'un piano droit et conception d'un système de reproduction du champ acoustique. Thèse de Doctorat, Université du Maine, Le Mans.

Deutsch, D. (Eds) (1982). *The psychology of music*. Academic Press.

Dingler T., Lindsay J., Walker B. N. (2008). Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. *In* *Proceedings of the 14th International Conference on Auditory Display*, Paris, France.

Dubois D. (Eds) (1997). *Catégorisation et Cognition : De la Perception au Discours*. Éditions Kimé, Paris.

Dubois F., Baudet G. Chamard J. C. (2012). EVADER: Electric Vehicle Alert for Detection and Emergency Response. *Proceedings Acoustics'2012 / Congrès Français d'Acoustique*, CFA, Nantes.

Eckel G., Iovino F., Caussé R. (1995). Sound synthesis by physical modelling with Modalys. *In* *Proceedings of International Symposium on Musical Acoustics*, ISMA, Dourdan.

Edworthy J., Loxley S., Dennis I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 33(2), 205-231.

Ehrenfels C. (von) (1890). Über Gestaltqualitäten. *Vierteljahrschrift für Wissenschaftliche Philosophie*, 14, 249-292.

Ellis D. P. W. (2003). Sinewave and Sinusoid + Noise Analysis/Synthesis in Matlab. <http://labrosa.ee.columbia.edu/matlab/sinemodel/>

Eriksson M., Bresin R. (2010) Improving running mechanisms by use of interactive sonification. *In* *Proceedings of the 3rd interactive sonification workshop*, ISON, Stockholm, Sweden.

Eysenck M. W. (2000). *Cognitive psychology: A student's handbook*. Taylor & Francis.

- Farina A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. *In* Audio Engineering Society Convention 108, AES.
- Fastl H. (1997). The psychoacoustics of sound-quality evaluation. *Acta Acustica united with Acustica*, 83(5), 754-764.
- Faure A. (2000). Des sons aux mots, comment parle-t-on du timbre musical? Thèse de Doctorat, Ecole des Hautes Etudes en Sciences Sociales, EHESS.
- Favre C. (1999). Analyse en normes L1 et L0 des distances et des préférences : planification en analyse sensorielle, application au confort d'accueil de sièges automobiles. Thèse de Doctorat, Université de Rennes II.
- Fernstrom M., Brazil E., Bannon L. (2005). HCI design and interactive sonification for fingers and ears. *IEEE Multimedia*, 12(2), 36-44.
- Fletcher N. H., Rossing T. (1998). *The Physics of Musical Instruments – 2nd Edition*, New- York, Springer-Verlag.
- Freed D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *The Journal of the Acoustical Society of America*, 87(1), 311-322.
- Frere A., Susini P., Misdariis N., Weber R., Peteul-Brouillet C., Guyader G. (2014). Vibrations' influence on Dieselness perception », *Applied Acoustics*, 77, 59-70.
- Fritz C., Curtin J., Poitevineau J., Lorrel-Samuels P., Tao F. C., (2012), Player preferences among old and new violins, *PNAS*, 109(3), 760–763.
- Frohlich B., Barrass S., Zehner B., Plate J., Gobel M. (1999). Exploring geo-scientific data in virtual environments. *In* Visualization'99. Proceedings, 169-173, IEEE.
- Gaillard P. (2009). Laissez-nous trier ! TCL-LabX et les tâches de catégorisation libre de sons, Le sentir et le dire. Concepts et méthodes en psychologie et linguistique cognitive.
- Garcia A. (1983). La directionnalité des instruments de musique. *Bulletin du Groupe d'Acoustique Musicale (G.A.M.), LAM-UPMC*, , No. 110.
- Gauthier P. A. (2007). Synthèse de champs sonores adaptative. PhD thesis, Université de Sherbrooke, Canada.
- Gaver W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-computer interaction*, 2(2), 167-177.
- Gaver W. W. (1988). Everyday listening and auditory icons. PhD thesis, University of California, San Diego.
- Gaver W. W. (1989). The SonicFinder: An interface that uses auditory icons. *Human-Computer Interaction*, 4(1), 67-94.

Gaver W. W. (1993). What is the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5(1), 1–29.

Gaver W. W. (1994). Using and creating auditory icons. In Kramer G. (Ed.) *Auditory display: sonification, audification and auditory interfaces*. Westview Press, New York.

Gazengel B., Petiot J. F., Soltes M. (2012). Objective and subjective characterization of saxophone reeds. *Proceed. Acoustics'2012 / Congrès Français d'Acoustique*, CFA, Nantes.

Gerzon M. A. (1985). Ambisonics in multichannel broadcasting and video. *Journal of the Audio Engineering Society*, 33(11), 859-871.

Gibson J. J. (1977). The Theory of Affordances. In Shaw R.E. & Bransford J. (Eds.), *Perceiving, Acting, and Knowing*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Gibson J. J. (1978). The ecological approach to the visual perception of pictures. *Leonardo*, 227-235.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton, Mifflin and Company, Boston, MA.

Giordano B. L., McAdams S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. *The Journal of the Acoustical Society of America*, 119(2), 1171-1181.

Giordano, B., Susini, P., Bresin, R.. (2013). Perceptual evaluation of sound-producing objects. In Franinovic K. & Serafin S. (Eds), *Sonic Interaction Design*, MIT Press.

Grassi M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & psychophysics*, 67(2), 274-284.

Grassi M., Casco C. (2010). Audiovisual bounce-inducing effect: When sound congruence affects grouping in vision. *Attention, Perception, & Psychophysics*, 72(2), 378-386.

Grey J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, 61(5), 1270-1277.

Guastavino C. (2003). Etude sémantique et acoustique de la perception des basses fréquences dans l'environnement sonore urbain. Thèse de Doctorat, Université Pierre et Marie Curie, Paris VI.

Guénoche A., Barthelémy J. P. (1993). Similitude, arbres et typicalité. In *Sémantique et cognition*, CNRS, Paris, 205-224.

Guski R., Felscher-Suhr U., Schuemer R. (1999). The concept of noise annoyance: how international experts see it. *Journal of sound and vibration*, 223(4), 513-527.

Guski R., Troje N. F. (2003). Audiovisual phenomenal causality. *Perception & Psychophysics*, 65(5), 789-800.

Guyot F. (1996). Etude de la perception sonore en termes de reconnaissance et d'appréciation qualitative : une approche par la catégorisation. Thèse de Doctorat, Université du Maine, Le Mans.

Hajda J. M., Kendall R. A., Carterette E. C., Harshberger M. L. (1997). Methodological issues in timbre research. *In* Deliège I., & Ioboda J. S. (Eds.), *Perception and cognition of music*, 253-306, Psychology Press, Hove, U.K.

Hammershøi D., Møller H. (2002). Methods for binaural recording and reproduction. *Acta Acustica united with Acustica*, 88(3), 303-311.

Helle S., Leplâtre G., Marila J., Laine P. (2001). Menu sonification in a mobile phone-a prototype study. *In* *Proceed. International Conference on Auditory Display, ICAD*, 255-260, Espoo, Finland.

Helmholtz (von) H. L. F. (1860). De l'action des cordes du violon. *Annales de la Société de Philosophie de Glasgow*.

Hermann T., Hunt A. (2005) An introduction to interactive sonification. *IEEE Multimedia* 12(2), 20-24.

Hermann T., Bovermann T., Riedenklau E., Ritter H. (2007) Tangible computing for interactive sonification of multivariate data. *In* *Proceedings of the 2nd International workshop on interactive sonification, ISON*, York, UK.

Hermann T. (2008). Taxonomy and definitions for sonification and auditory display. *In* *Proceedings of the 14th International Conference on Auditory Display, ICAD'08*, Paris.

Hiller L., Ruiz P. (1971). Synthesizing musical sounds by solving the wave equation for vibrating objects. *Journal of the Audio Engineering Society*, 19(6), 462-470 (Part 1) ; 19(7), 542-551 (Part 2).

Hirschberg A., Gilbert J., Wijnands A. P. J., Valkering A. M. C. (1994). Musical aero-acoustics of the clarinet. *Le Journal de Physique IV*, 4(C5), C5-559.

Houdé O., Kayser D., Koenig O., Proust J., Rastier F. (1998). *Vocabulaire de sciences cognitives. Psychologie et sciences de la pensée*. Presses Universitaires de France, Paris.

Houx O. (2003). Catégorisation auditive des sources sonores. Thèse de Doctorat, Université du Maine, Le Mans.

Houx O., Lemaitre G., Misdariis N., Susini P., Visell Y., Franinovic K., Rocchesso D. (2007). Everyday sound classification: Sound perception, interaction and synthesis Part I - State of the art : deliverable 4.1 of the CLOSED project.
http://www.closed.ircam.fr/uploads/media/closed_deliverable_4.1_public.pdf

Houx O., Lemaitre G., Misdariis N., Susini P., Urdapilleta I. (2012). A Lexical Analysis of Environmental Sound Categories. *Journal of Experimental Psychology: Applied*, JEP, 18(1), 52-80.

Howard J. H., Ballas (1980). A. Syntactic and semantic factors in the classification of nonspeech transient patterns. *Perception and Psychophysics*, 28(5), 431-439.

Hutchins C. M., Benade V. (1997). Research papers in violin acoustics, 1975-1993: with an introductory essay, 350 years of violin research. Acoustical society of America, New York, Woodbury.

Iverson P., Krumhansl C. L. (1993). Isolating the dynamic attributes of musical timbre. *The Journal of the Acoustical Society of America*, 94(5), 2595-2603.

Jacob K. D., Birkle T. K. (1990). Prediction of the full-space directivity characteristics of loudspeaker arrays. *Journal of the Audio Engineering Society*, 38(4), 250-259.

Jacobsen F., Figueroa S. B., Rasmussen K. (2004). A note on the concept of acoustic center. *The Journal of the Acoustical Society of America*, 115(4), 1468-1473.

Jansson E. (2002). Acoustics for violin and guitar makers - Chapter VII: the function of the violin », KTH, Dept of Speech, Music and Hearing, 4th edition.

Jekosch U. (1999). Meaning in the context of sound quality assessment. *Acta Acustica united with Acustica*, 85(5), 681-684.

Jeon M., Walker B. N. (2009). "Spindex": Accelerated Initial Speech Sounds Improve Navigation Performance in Auditory Menus. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 53(17), 1081-1085, SAGE Publications.

Johnson S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3), 241-254.

Jossic M., Meurisse T., Misdariis N., Mamou-Mani A. (2014). Perception du contrôle actif sur instruments à vent. *Actes du Congrès Français d'Acoustique, CFA, Poitiers*.

Jot J. M., Warusfel O. (1995). Spat: a spatial processor for musicians and sound engineers. *In CIARM: International Conference on Acoustics and Musical Research*, Ferrara.

Jot J. M., Cerveau L., Warusfel O. (1997). Analysis and synthesis of room reverberation based on statistical time-frequency model. *Proc. 103rd Audio Eng. Soc. Convention*, New-York, USA.

Jot J. M., Cerveau L., Warusfel O. (1997). Analysis and synthesis of room reverberation based on a statistical time-frequency model. *In Audio Engineering Society Convention 103*, AES.

Jot J. M., Wardle S., Larcher V. (1998). Approaches to binaural synthesis. *In Audio Engineering Society Convention 105*, AES.

Kahle E. (1995). Validation d'un modèle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d'opéras . Thèse de doctorat, Université du Maine, Le Mans.

Karplus K., Strong A. (1983). Digital synthesis of plucked-string and drum timbres. *Computer Music Journal*, 7(2), 43-55.

Kassakian P., Wessel D. (2003). Design of low-order filters for radiation synthesis. *In* Audio Engineering Society Convention 115, AES.

Kassakian P., Wessel D. (2004). Characterization of spherical loudspeaker arrays. *In* Audio Engineering Society Convention 117, AES.

Khaykin D., Rafaely B. (2010). Spherical loudspeaker array modeling and control using different basis function representations. *In* Proceed. of the 2nd International Symposium on Ambisonics and Spherical Acoustics, Paris, France.

Kendall R. A., Carterette E. C. (1993). Verbal attributes of simultaneous wind instrument timbres. I. von Bismarck's adjectives. *Music Perception: An Interdisciplinary Journal*, (10)4, 445-467.

Klatzky R. L., Pai D. K., Krotkov E. P. (2000). Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environment*, 9(4), 399-410.

Koopman P. D., Cottingham J. P. (1996). Acoustical properties of free reeds. *Quarterly Reed Organ Soc*, 15(3-4), 17-23.

Krimphoff J. (1993). Mémoire de DEA. Université du Maine, Le Mans. (*non publié*)

Krimphoff J., McAdams S., Winsberg S. (1994). Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique. *Le Journal de Physique IV*, 4(C5), C5-625.

Krumhansl C. L. (1989). Why is musical timbre so hard to understand? *In* Nielzen S. & Olsson O. (Eds.), *Structure and Perception of Electroacoustic Sound and Music*, 43-53, Elsevier, Amsterdam.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis *Psychometrika*, 29(1), 1-27.

Kruskal J. B., Wish M. (1978). Multidimensional scaling. University Paper series on Quantitative Applications in the Social Sciences, Sage Publications, Beverly Hills, CA.

Kunkler-Peck A. J., Turvey M. T. (2000). Hearing shape. *Journal of Experimental psychology: human perception and performance*, 26(1), 279-294.

Lafay G., Rossignol M., Misdariis N., Lagrange M., Petiot J. F. (2014). A new experimental approach for urban soundscape characterization based on sound manipulation : A pilot study. *In* Proceedings of 26th International Symposium on Musical Acoustics, Le Mans.

Lakatos S., McAdams S., Caussé R. (1997). The representation of auditory source characteristics: simple geometric sources. *Perception & psychophysics*, 59(8), 1180-1190.

Langlois, S., Le Nindre, B., Peteul, C., Misdariis, N. (2005). Man-machine interfaces in cars : technological impact of sound design. *In* Proceed. Forum Acusticum, Budapest.

Langlois S., Loiseau S., Tardieu J., Cera A., Misdariis N. (2010). Evaluation de la sonification d'un système multimédia automobile. *In* Conference Internationale Francophone sur l'Interaction Homme-Machine, 173-180, ACM.

Lartillot O., Toiviainen P. (2007). A Matlab toolbox for musical feature extraction from audio. *In* International Conference on Digital Audio Effects, 237-244.

Lavandier M. (2005). Différences entre enceintes acoustiques : une évaluation physique et perceptive. Thèse de doctorat, Université de la Méditerranée, Aix- Marseille II.

Le Ny J. F. (1989). Le développement actuel de la psychologie scientifique. *L'année psychologique*, 89(2), 213-238.

Legendre P., Legendre L. (1998), Numerical Ecology. Development in environmental modelling. Elsevier, 2nd english edition.

Leipp E. (1971). Acoustique et musique: données physiques et technologiques, problèmes de l'audition des sons musicaux, principes de fonctionnement et signification acoustique des principaux archétypes d'instruments de musique, les musiques expérimentales, l'acoustique des salles. Masson et Cie, Editeurs.

Leipp E. (1977). La machine à écouter: essai de psycho-acoustique. Masson, Paris

Leishman T. W., Rollins S., Smith H. M. (2006). An experimental evaluation of regular polyhedron loudspeakers as omnidirectional sources of sound. *The Journal of the Acoustical Society of America*, 120(3), 1411-1422.

Lemaitre G., Susini P., Winsberg S., McAdams S., Letinturier B. (2007). The sound quality of car horns: a psychoacoustic study of timbre. *Acta Acustica united with Acustica*, 93(3), 457-468.

Lemaitre G., Susini P., Winsberg S., McAdams S., Letinturier B. (2009a). The sound quality of car horns: Designing new representative sounds. *Acta Acustica united with Acustica* 95(2), 356-372.

Lemaitre G., Houix O., Visell Y., Franinovic K., Misdariis N., Susini P. (2009b). Toward the design and evaluation of continuous sound in tangible interfaces: The Spinotron. *International Journal of Human-Computer Studies*, 67(11), 976-993.

Lemaitre, G., Houix, O., Misdariis, N., Susini, P. (2010). Listener expertise and sound identification influence the categorization of environmental sounds. *Journal of Experimental Psychology: Applied*, 16(1), 16-32.

Lemaitre G., Houix O., Susini P., Visell Y., Franinovic K. (2012). Feelings elicited by auditory feedback from a computationally augmented artifact: The Flops. *Affective Computing, IEEE Transactions on*, 3(3), 335-348.

Lemaitre, G., Rocchesso, D. (2013). On the effectiveness of vocal imitation and verbal descriptions of sounds. *The Journal of the Acoustical Society of America* vol. 135(2), 862-873.

Leplâtre G., Brewster S. (2000). Designing Non-Speech Sounds to Support Navigation in Mobile Phone Menus. *In Proceedings of International Conference on Auditory Display*, 190-199, Atlanta, USA.

Lesueur C. (1988). *Rayonnement acoustique des structures*, Eyrolles.

Lichte W. H. (1941). Attributes of complex tones. *Journal of Experimental Psychology* 28(6), 455-480.

Lock D., Schiemer G. (2006). Orbophone: a new interface for radiating sound and image », *Proceedings of the 2006 International Conference on New Interfaces for Musical Expression*, 89-92, Paris, France

Loyez P. (1992), *Techniques des haut-parleurs et enceintes acoustiques*, Editions Fréquences.

Lutfi R. A., Oh E. L. (1997). Auditory discrimination of material changes in a struck-clamped bar. *The Journal of the Acoustical Society of America*, 102(6), 3647-3656.

Lyon, R. H. (2000). *Designing for product sound quality*. CRC Press.

Maffiolo V. (1999). *De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain*. Thèse de doctorat, Université du Maine, Le Mans.

Malham D. G (1995). *Basic Ambisonics*. University of York,.
http://www.york.ac.uk/inst/mustech/3d_audio/ambis2.htm

Marozeau J., de Cheveigné A., McAdams S., Winsberg S. (2003). The dependency of timbre on fundamental frequency. *The Journal of the Acoustical Society of America*, 114(5), 2946-2957.

Marshall K. D. (1985a). The influence of the musician on the vibrational behavior of a violin. *The Journal of the Acoustical Society of America*, 77(S1), S44-S45.

Marshall K. D. (1985b). Modal analysis of a violin. *The Journal of the Acoustical Society of America*, 77(2), 695-709.

Mas A., Leipp E. (1977). Une enceinte acoustique instrumentale. *Bulletin du GAM* No. 89.

Matlin M. W. (2001). *La cognition: une introduction à la psychologie cognitive*. De Boeck Supérieur.

McAdams S. (1993). Recognition of auditory sound sources and events. *In* McAdams S. and Bigand E. (Eds), *Thinking in sound: the cognitive psychology of human audition*, Oxford University Press.

McAdams S., Bigand E. (1994). *Penser les sons: psychologie cognitive de l'audition*. Presses Universitaires de France, Paris.

McAdams, S. (1994a). La reconnaissance de sources et d'événements sonores. *Penser les sons*, 157-213 (chap.6), Presses Universitaires de France, Paris.

McAdams S. (1994b). Audition : physiologie, perception et cognition. *In* Richelle M., J. Requin J. & Robert M. (Eds), *Traité de psychologie expérimentale*, 283-344, Presses Universitaires de France, Paris.

McAdams S., Winsberg S., Donnadiou S., De Soete G., Krimphoff J. (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychological Research*, 58(3), 177-192.

McAdams S., Susini P., Misdariis N., Winsberg S. (1998). Multidimensional characterisation of perceptual and preference judgements of vehicle and environmental noises. *In* Proceedings of Euro-Noise'98 conference, Munich, Germany.

McAdams S., Misdariis N. (1999). Perceptual-based retrieval in large musical sound databases. *In* Proceedings of the Human Centered Processes Conference, 445-450, Brest.

McAdams S., Chaigne A., Roussarie V. (2004). The psychomechanics of simulated sound sources: material properties of impacted bars. *The Journal of the Acoustical Society of America*, 115(3), 1306– 1320.

McAdams S. (2012). Musical timbre perception. *In* Deutsch D. (Ed.), *The psychology of music*, Academic Press.

McIntyre J., Zago M., Berthoz A., Lacquaniti F. (2001). Does the brain model Newton's laws?. *Nature neuroscience*, 4(7), 693-694.

McIntyre M. E., Schumacher R. T., Woodhouse J. (1983). On the oscillations of musical instruments. *The Journal of the Acoustical Society of America*, 74(5), 1325-1345.

Mercier D. (2002), *Le livre des techniques du son - Tome 1*. Paris, Dunod, 3^{ème} édition.

Meurisse T., Mamou-Mani A., Caussé R., Sharp D. (2013), Active control applied to simplified wind musical instrument. *In* Proceedings of International Congress on Acoustics, Montreal, Canada.

Meurisse T., Mamou-Mani A., Caussé R., Sharp D. (2013). Active control applied to simplified wind musical instrument. *In* Proceedings of Meetings on Acoustics 19(1), Acoustical Society of America.

Miller J. R., Carterette E. C. (1975). Perceptual space for musical structures. *The Journal of the Acoustical Society of America*, 58(3), 711-720.

Millot L., Cuesta C., Valette C. (2001). Experimental results when playing chromatically on a diatonic harmonica. *Acta Acustica united with Acustica*, 87(2), 262-270.

Minard A. (2013). Perception et confort acoustiques des systèmes de traitement d'air. Thèse de doctorat, Université de La Rochelle (France) et Université de Liège (Belgique).

Misdariis N., Smith B. K., Pressnitzer D., Susini P., McAdams S. (1998a). Validation of a multidimensional distance model for perceptual dissimilarities among musical timbres. *In Proceedings of Joint meeting of the 16th International Congress on Acoustics, 135th meeting of Acoustical Society of America, The Journal of the Acoustical Society of America*, vol. 103(1), 3005.

Misdariis N., McAdams S. (1998b). Validation d'un modèle multidimensionnel de distance pour la dissemblance perceptive des timbres musicaux. Actes des 4èmes Journées Françaises sur la Psychologie de la Musique, Bréau, France.

Misdariis N., Ricot D., Caussé R. (2000). Modélisation physique de la vibration d'une anche d'accordéon. Actes du Congrès Français d'Acoustique, 281-284, Lausanne.

Misdariis N., Warusfel O., Caussé R. (2001a). Radiation control on a multi-loudspeakers device. *In Proceedings of the International Symposium on Musical Acoustics*, Perugia.

Misdariis N., Nicolas F., Warusfel O., Caussé R. (2001b). Radiation control on a multi-loudspeaker device: la Timée. *In Proceedings of International Computer Music Conference*, 306-309, La Havane.

Misdariis N., Warusfel O., Caussé R., Nicolas F. (2002). Le contrôle de la directivité par un système multi haut-parleurs : applications. Actes des Journées d'Informatique Musicale, JIM, Marseille, 121-129.

Misdariis N., Defrance G., Warusfel O. (2006). Contrôle du Rayonnement Appliqué à la Synthèse Sonore: Description de "Synthèse Additive Spatiale". Actes du 8ème Congrès Français d'Acoustique, Tours.

Misdariis N., Lang A., Katz B. F., Susini P. (2008). Perceptual effects of radiation control with a multi-loudspeaker device. *The Journal of the Acoustical Society of America*, 123(5), 3665-3665, Acoustics08, Paris

Misdariis N., Boyer E., Caussé R., Noisternig M. (2010a). Le contrôle du rayonnement appliqué à la reproduction d'un instrument de musique. Actes du 10^{ème} Congrès Français d'Acoustique, CFA, Lyon.

Misdariis N., Minard A., Susini P., Lemaitre G., McAdams S., Parizet E. (2010b). Environmental Sound Perception: Metadescription and Modeling Based on Independent Primary Studies. *EURASIP Journal on Audio, Speech, and Music Processing*.

Misdariis N., Tardieu J., Langlois S., Loiseau S. (2011). Menu Sonification in an Automotive Media Center : Design and Evaluation. *In Kolski C. (Ed.), Human-Computer Interactions in Transport*, ISTE Ltd - John Wiley & Sons Inc., 233-281.

Misdariis N., Cera A. (2013). Sound signature of Quiet Vehicles: state of the art and experience feedbacks. *In* Inter-Noise and Noise-Con Congress and Conference Proceedings 247(5), 3333-3342, Institute of Noise Control Engineering.

Most S. B., Scholl B. J., Clifford E. R., Simons D. J. (2005). What You See Is What You Set: Sustained Inattentional Blindness and the Capture of Awareness. *Psychological Review*, 112(1), 217-242.

Monichon P., Leipp E., Abott A., Lorin E. (1972). L'accordéon. *Bulletin du Groupe d'Acoustique Musicale, GAM*, No. 59.

Msallam R., Dequidt S., Causse R., Tassart S. (2000). Physical model of the trombone including nonlinear effects. Application to the sound synthesis of loud tones. *Acta Acustica united with Acustica*, 86(4), 725-736.

Mzali M. (1997). Influence de la directivité des sources sur la perception d'une scène sonore. *Mémoire Master ATIAM, Université Pierre et Marie Curie, Paris VI*.

Neisser U. (1967). *Cognitive psychology*. Appleton-Century-Crofts Eds.

Neisser U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. WH Freeman/Times Books/Henry Holt & Co.

Nicol R. (199). Etude de la restitution du son spatialisé dans une zone étendue : application à la téléprésence. *Thèse de doctorat, Université du Maine, Le Mans*.

Nicolas F. (2000). Comment la musique prend-elle acte d'un espace architectural. Colloque « Espace [re][dé]composé », Ircam.
<http://www.entretemps.asso.fr/Nicolas/Timee/Espace.html>

Norman D. A. (1999). Affordance, conventions, and design. *Interactions*, 6(3), 38-43.

Olson H. F. (1952). *Musical Engineering*, McGraw-Hill, New York

Osgood C. E. (1952). The nature and measurement of meaning. *Psychological bulletin*, 49(3), 197-237.

Otondo F., Rindel J. H. (2005). A new method for the radiation representation of musical instruments in auralizations. *Acta Acustica united with Acustica* 91(5), 902-906.

Ozcan E., van Egmond R. (2007). Memory for product sounds: the effect of sound and label type. *Acta Psychologica*, 126(3), 196-215.

Ozcan E., van Egmond R. (2009). Product Sound Design: An Inter-Disciplinary Approach?. *In* Undisciplined! Design Research Society Conference, Sheffield Hallam University, UK.

Palladino D. K., Walker B. N. (2007). Learning rates for auditory menus enhanced with spearcons versus earcons. *In* Proceedings of the 13th International Conference on Auditory Display, ICAD, 274-279.

Palladino D. K, Walker B. N. (2008). Efficiency of spearcon-enhanced navigation of one dimensional electronic menus. *In Proceedings of the 14th International Conference on Auditory Display, ICAD.*

Parizet E., Guyader E., Nosulenko V. (2008). Analysis of car door closing sound quality. *Applied Acoustics*, 69(1), 12-22.

Pasqual A. M., Arruda J. R., Herzog P. (2008a). Effects of enclosure design on the directivity synthesis by spherical loudspeaker arrays. *The Journal of the Acoustical Society of America*, 123(5).

Pasqual A. M., Arruda J. R., Herzog P. (2008b). Optimal array pattern synthesis with desired magnitude response. *The Journal of the Acoustical Society of America*, 123(5), 3643.

Pasqual A. M., Arruda J. R., Herzog P. (2010a). A comparative study of Platonic solid loudspeakers as directivity controlled sound sources. *In Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics, Paris.*

Pasqual A. M., de França A., Roberto J., Herzog P. (2010b). Application of acoustic radiation modes in the directivity control by a spherical loudspeaker array. *Acta acustica united with Acustica*, 96(1), 32-42.

Patterson R. D., Edworthy J., Shailer M. J., Lower M., Wheeler P. (1986). Alarm sounds for medical equipment in intensive care areas and operating theatres. Research Report No. AC598, Institute of Sound and Vibration Research, University of Southampton, UK.

Patterson R. D., Robinson K., Holdsworth J., McKeown D., Zhang C., Allerhand M. (1995). Complex sounds and auditory images. *Auditory Physiology and Perception*, 83, 429-446.

Pätynen J., Lokki T. (2010). Directivities of symphony orchestra instruments. *Acta Acustica united with Acustica*, 96(1), 138-167.

Peeters G. (2004). A large set of audio features for sound description (similarity and description) in the cuidado project. Research Report, Ircam, Paris.
http://www.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf

Peeters G., McAdams S., Krimphoff J., Susini P., Misdariis N., Bennett S. (2001). Procédé de caractérisation du timbre d'un signal sonore selon au moins un descripteur. INPI, France, CNRS Ircam, France Telecom.

Peeters G., Giordano B. L., Susini P., Misdariis N., McAdams S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5), 2902-2916.

Petiot J. F., Caussé R. (2007). Perceptual Differences between cellos: a subjective/objective study. *In Proceedings of International Symposium on Musical Acoustics, Barcelona.*

- Pierce J. R. (1992). The Science of Musical Sound. Scientific American/Freeman (general).
- Placido L., Mamou-Mani A., Sharp D. (2011), Investigating perceptual differences between two trumpets of the same model type, *Applied Acoustics*, 72(12), 907-914.
- Pollow M., Behler G. K. (2009). Variable Directivity for Platonic Sound Sources Based on Spherical Harmonics Optimization, *Acta Acustica united with Acustica*, 95(6), 1082-1092.
- Pollow M., Behler K., Schultz F. (2010). Musical Instrument Recording for Building a Directivity Database. *In* 36th German Annual Conference on Acoustics, DAGA'10, Berlin.
- Pomberger H. (2008). Angular and radial directivity control for spherical loudspeaker arrays. Diploma thesis, University of Music and Dramatic Arts, Graz, Austria.
<http://iem.kug.ac.at/fileadmin/media/iem/altdaten/projekte/acoustics/awt/winkel/pomberger.pdf>
- Pompei F. J. (1999). The Use of Airborne Ultrasonics for Generating Audible Sound Beams. *Journal of the Audio Engineering Society*, 47(9), 726-731.
- Prezat F., (2006). Analyse du rayonnement de la voix parlée et chantée, rapport de stage LIMSI-CNRS.
- Pulkki V. (1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6), 456-466.
- Qi H., Hartono P., Suzuki K., Hashimoto S. (2002). Sound database retrieved by sound. *Acoustical Science and Technology* 23(6), 292-300.
- Rath M., Rocchesso D. (2005). Continuous sonic feedback from a rolling ball. *IEEE Multimedia*, 12(2), 60-69.
- Rath M., Schleicher D. R. (2008). On the relevance of auditory feedback for quality of control in a balancing task. *Acta Acustica United Acustica*, 94(1), 12-20.
- Rayleigh J. W. S. B. (1945). *Theory of Sound*, New-York, Dover.
- Repp B. H. (1987). The sound of two hands clapping: An exploratory study. *The Journal of the Acoustical Society of America*, 81(4), 1100-1109.
- Ricot D. (1999). Modélisation physique de la vibration d'une anche d'accordéon, Mémoire de fin d'études, Ecole Centrale de Lyon.
- Ricot D., Caussé R., Misdariis N. (2005). Aerodynamic excitation and sound production of blown-closed free reeds without acoustic coupling: The example of the accordion reed. *The Journal of the Acoustical Society of America*, 117(4), 2279-2290.
- Roads C., Strawn J. (1985). *Foundations of computer music*. MIT Press.
- Roads C. (1996). *The computer music tutorial*. MIT press.

Robare P. (2009). Sound in Product Design. Master of Design in Interaction Design, Carnegie Mellon University School of Design.

Rocchesso D., Polotti P. (2008) Designing continuous multisensory interaction. *In* contribution to the Sonic Interaction Design workshop, CHI conference, Florence, Italy.

Rocchesso D., Polotti P., delle Monache S. (2009). Designing continuous sonic interaction. *International Journal of Design*, 3(3).

Rodet X. (2001). The additive analysis-synthesis package. Ircam tutorial, <http://recherche.ircam.fr/equipes/analyse-synthese/DOCUMENTATIONS/additive/index-e.html>

Rodriguez W. (2003). Le design sonore, naissance d'une catégorie musicale. Mémoire de DEA, Ecole des Hautes Etudes en Sciences Sociales, EHESS, Paris.

Rosch E. (1978). Principles of categorization. In E. Rosch and B. Lloyd, editors, *Cognition and Categorization*, pages 27–48. Lawrence Erlbaum Associates, Hillsdale, NJ, 1978.

Rosenblatt F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408.

Roussarie V. (1999). Analyse perceptive de structures vibrantes simulées par modèle physique. Thèse de doctorat, Université du Maine, Le Mans.

Rousseau A. (1996). Modélisation du rayonnement des instruments à vent à trous latéraux. Rapport de stage, Ircam (*non publié*).

Russ M. (2008). Sound synthesis and sampling. Taylor & Francis, Focal Press, 3rd edition.

Saint-Loubry, B. (2000). Sound quality mark for economics cars. In *Internoise and Noise-Con Congress and Conference Proceedings*, 3, 4323-4327.

Sakamoto, N., Gotoh, T., Atoji, N., Aoi, T. (1975). Frequency Response of an Electrostatic Horn-Tweeter with Electret. *In* AES – Audio Engineering Society Convention 52nd.

Schaeffer P. (1966). *Traité des objets musicaux*. Paris: Editions du Seuil.

Schafer R. M. (1977). *The tuning of the world*. Random House Inc.

Schafer, R. M. (1979). *Le paysage sonore*. J.-C. Lattès.

Schafer, R. M. (2010). *Le paysage sonore: le monde comme musique*. Éditions Wildproject.

Schaffert N., Mattes K., Effenberg A. O. (2010) Listening to the boat motion: acoustic information for elite rowers. *In* *Proceedings of the 3rd Interactive sonification workshop, ISON*, Stockholm, Sweden.

Scherer K. R., Dan E. S., Flykt A. (2006) What determines a feeling's position in affective space? A case for appraisal. *Cognition & Emotion*, 20(1), 92–113.

Schiffman S., Reynolds M. L. Young F. W. (1981). Introduction to multidimensional scaling - Theory, Methods and Applications. Academic Press, New-York.

Schmeder A. (2009). An exploration of design parameter for human interactive systems with compact spherical loudspeaker arrays. *In Proceedings of Ambisonics Symposium*, Graz.

Schwarz, G. (1978). Estimating the dimensions of a model. *The annals of statistics*, 6(2), 461-464.

Shepard R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27(2), 125-140.

Shepard R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, 27(3), 219-246.

Shepard R. N., Arabie P. (1979). Additive clustering: Representation of similarity as combinations of discrete overlapping properties. *Psychological Review*, 86(2), 87-123.

Siekierski E., Derquenne C., Martin N. (2001). Sensory evaluation of air-conditioning noise: Sensory profiles and hedonic tests. *In Proceedings of International Congress on Acoustics*, Rome.

Slaney, M. (1993). An efficient implementation of the Patterson-Holdsworth auditory filter bank. Apple Computer, Perception Group, Tech. Rep, 35, 8.

Slaney M. (1998). Auditory Toolbox, version 2. Interval Research Corporation Technical Report No. 1998-010. <http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/>

Sloutsky V. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, 7(6), 246-251.

Smith, B. K. (1995). PsiExp: an environment for psychoacoustic experimentation using the IRCAM musical workstation. *In Society for music perception and cognition conference'95*, University of Berkeley, CA.

Smith J. O. (1992). Physical modelling using digital waveguides. *Computer Music Journal*, 16(4), 74-91, special issue: Physical Modeling of Musical Instruments, Part I.

Solomon L. N. (1958). Semantic approach to the perception of complex sounds. *The Journal of the Acoustical Society of America*, 30(5), 421-425.

St-Hilaire A. O., Wilson T. A., Beavers G. S. (1971). Aerodynamic Excitation of the Harmonium Reed. *Journal of Fluid Mechanics*, 49(04), 803-816.

Stan G. B., Embrechts J. J., Archambeau D. (2002). Comparison of different impulse response measurement techniques. *Journal of the Audio Engineering Society*, 50(4), 249-262.

Stanton N. A., Edworthy J. (Eds) (1999). Human factors in auditory warnings. Ashgate Publishing, Farnham.

Susini, P. McAdams S., Winsberg S. (1997a). Caractérisation perceptive des bruits de véhicules. *In Actes du 4^{ème} Congrès Français d'Acoustique, CFA, Marseille.*

Susini P., McAdams S., Winsberg S. (1997b). Perceptual characterisation of vehicules noises. EEA Symposium: Psychoacoustic in Industry and Universities. Eindhoven, The Netherlands.

Susini P., McAdams S., Winsberg S. (1999) A multidimensional technique for sound quality assessment. *Acta Acustica united with Acustica*, 85(5), 650-656.

Susini P., McAdams S., Winsberg S., Perry I., Vieillard S., Rodet X. (2004). Characterizing the sound quality of air-conditioning noise. *Applied Acoustics*, 65(8), 763-790.

Susini P., Misdariis N., Houix O., Lemaitre G. (2009). Does a "natural" sonic feedback affect perceived usability and emotion in the context of use of an ATM?. *In Proceedings of Sound and Music Computing conference, Porto.*

Susini P. (2011). Le design sonore, un cadre expérimental et applicatif pour explorer la perception sonore”. Habilitation à Diriger des Recherches. Université Méditerranée, Aix-Marseille II.

Susini P., Lemaitre G., McAdams S. (2011). Psychological Measurement for Sound Description and Evaluation. *Measurement With Persons*, Berglund B., Rossi G. B., Townsend J. T. & Pendrill L. R. (Eds), Taylor & Francis Group, New York, 227-254.

Susini P., Misdariis N., Lemaitre G., Houix O. (2012). Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback. *Journal on Multimodal User Interface*, 5(3), 175-186.

Susini P., Houix O., Misdariis N. (2014). Sound design: an applied, experimental framework to study the perception of everyday sounds. *The New Soundtrack*, Edinburgh University Press.

Tagg P. (1982). Teoria degli affetti. *In Analysing popular music: theory, method and pratice*. *Popular Music*, 2, 37-67.

Tardieu J., Misdariis N., Langlois S., Gaillard P., Lemercier C. (2014). Sonification of in-vehicle interface reduces gaze movements under dual-task condition. *Applied Ergonomics (en cours de soumission)*.

Tarnow V. (1974). Sound radiation from loudspeaker systems with the symmetry of the platonic solids. *Brüel&Kjær, Technical Review*.

Tonon T. (2009). Accordion reeds, cavity resonance, and pitch bend. *The Journal of the Acoustical Society of America*, 126(4), 2217-2217.

Topper D., Burtner M., Serafin S. (2002). Spatio-operational spectral (s.o.s.) synthesis. *In* Proceedings of the 5th International Conference on Digital Audio Effects, DAFX'02, Hamburg, Germany.

Tractinsky N., Katz A., Ikar D. (2000) What is beautiful is usable. *Interacting with Computers*, 13(2), 127–145.

Trueman D., Bahn C., Cook P. R. (2000). Alternative Voices for Electronic Sounds: Spherical Speakers and Sensor-Speaker Arrays (SenSAs). *In* Proceedings of 19th International Computer Music Conference, Berlin.

Tuuri K., Mustonen M. S., Pirhonen A. (2007). Same sound–different meanings: A novel scheme for modes of listening. *In* Proceedings of the 2nd Audio Mostly Conference on Interaction with Sound.

Upop Montreal (2011). Les grands paradigmes des sciences cognitives. http://www.upopmontreal.com/wp-content/uploads/2011/11/II_COURS_4-22-nov-2011-v2.pdf

Vanderveer N. J. (1979). Ecological acoustics: human perception of environmental sounds. Doctoral dissertation, Cornell University, New-York (*unpublished*).

Varela F. J., Rosch E., Thompson E. (1992). The embodied mind: Cognitive science and human experience. MIT press.

Vergez C., Bensoam J., Misdariis N., Caussé R. (2000). Modalys : sound Synthesis by Nonlinear Interaction Between Modal Objects. *In* Proceedings of the 140th Acoustical Society of America meeting, Atlanta.

Vial S. (2011). Court traité du design. Presses Universitaires de France.

Vinet H., Assayag G., Burred J. J., Carpentier G., Misdariis N., Peeters G., Roebel A., Schnell N., Schwarz D., Tardieu D. (2011). Sample Orchestrator : gestion par le contenu d'échantillons sonores. *Revue Traitement du Signal*, 3, 417-468, Eds Lavoisier.

Viollon S. (2000). Influence des informations visuelles sur la caractérisation de la qualité acoustique de l'environnement urbain. Thèse de doctorat, Université de Cergy-Pontoise.

Vos H., Warusfel O., Misdariis N., De Vries D. (2003). Analysis and reproduction of the frequency spectrum and directivity of a violin. *Journal of the Acoustical Society of the Netherlands*, 167.

Walker B. N., Nance A., Lindsay J. (2006). Spearcons: Speech-based earcons improve navigation performance in auditory menus. *In* Proceedings of the International Conference on Auditory Display, London, UK, 63-68.

Walker B. N., Kogan A. (2009). Spearcon performance and preference for auditory menus on a mobile phone. *In* Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments, 445-454, Springer Berlin Heidelberg.

Warren-Jr W. H., Verbrugge R. R. (1984). Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology: Human perception and performance*, 10(5), 704–712.

Warusfel O., Derogis P., Caussé R. (1997). Reproduction of directivity patterns using multi-loudspeaker sources. *In Proceedings 103rd Audio Engineering Society Convention*, New-York, USA.

Warusfel O., Misdariis N., Caussé R. (2000). Physical and perceptual modeling of sound-source radiation. *In Proceedings of the 140th Acoustical Society of America meeting*, Atlanta.

Warusfel O., Misdariis N. (2001). Directivity synthesis with 3D array of loudspeakers: application for stage performance. *In Proceedings of the 4th International Conference on Digital Audio Effects, DAFX'01*, Limerick, Ireland.

Warusfel O., Misdariis N., Caulkins T., Corteel E. (2004a). Radiation control applied to sound synthesis: an attempt for "Spatial Additive Synthesis". *In Proceedings of the 150th Acoustical Society of America meeting*, New-York.

Warusfel O., Misdariis N. (2004b). Sound source radiation syntheses: From performance to domestic rendering. *In Audio Engineering Society Convention 116*, AES.

Warusfel O., Corteel E., Misdariis N., Caulkins T. (2004c). Reproduction of sound source directivity for future audio applications. *In Proceedings of International Congress on Acoustics*, Kyoto.

Watkinson J. (1998). *The art of sound reproduction*. Taylor & Francis.

Weinreich G., Arnold E. B. (1980). Method for measuring acoustic radiation fields. *The Journal of the Acoustical Society of America*, 68(2), 404-411.

Weinreich G. (1985). Sound hole sum rule and the dipole moment of the violin. *The Journal of the Acoustical Society of America*, 77(2), 710-718.

Weinreich G. (1997a). Directional tone color. *The Journal of the Acoustical Society of America* 101(4), 2338-2346.

Weinreich G. (1997b). The DTC loudspeaker. *The Journal of the Acoustical Society of America* 101(5), 3071-3071.

Wessel D. (1979). Timbre Space as a Musical Control Structure. *Computer Music Journal*, 3(2), 45-52, published by MIT Press.

Wessel D. (1991). Instruments That Learn, Refined Controllers, and Source Model Loudspeakers. *Computer Music Journal*, 15(4), 82-86.

Williams M. (1992). The Basic Principles of ISSR (Individual Sound Source Reinforcement). *In Audio Engineering Society Convention 93*, AES.

Winsberg S., Carroll J. D. (1989). A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. *Psychometrika* , 54(2), 217-229.

Winsberg S., De Soete G. (1993). A latent class approach to fitting the weighted Euclidean model, CLASCAL. *Psychometrika* , 58(2), 315-330.

Woodcock S. (2003). MATLAB econometrics toolbox.
<http://www.sfu.ca/~swoodcoc/software/software.html>

Woodard S. E. (2005). Methods to increase sound fidelity and quality produced from piezoelectric devices. *Journal of Sound and Vibration*, 280(1), 127-240.

Zotter F., Noisternig M. (2007a). Near-and farfield beamforming using spherical loudspeaker arrays. *In* Proceed. of 3rd Congress of the Alps Adria Acoustics Association.

Zotter F., Sontacchi A., Holdrich R. (2007b). Modeling a spherical loudspeaker system as multipole source. *Fortschritte der Akustik*, 33(1).

Zotter F., Höldrich R. (2007c). Modeling radiation synthesis with spherical loudspeaker arrays. *In* Proceedings of 19th International Congress on Acoustics, Madrid.

Zotter F., Schmeder A., Noisternig M. (2008a). Crosstalk cancellation for spherical loudspeaker arrays. *In* 34th German Annual Conference on Acoustics, DAGA'08.

Zotter F., Pomberger H., Schmeder A. (2008b). Efficient directivity pattern control for spherical loudspeaker arrays. *The Journal of the Acoustical Society of America*, 123(5), 3643-3643.

Zwicker E. (1977). Procedure for calculating loudness of temporally variable sounds. *The Journal of the Acoustical Society of America*, 62(3), 675-682.

Zwicker E, Fastl H. (1990). *Psychoacoustics: Facts and models*. Springer-Verlag, Berlin.

Annexes

Annexes

Liste des Annexes :

- Annexe 1. Rayonnement acoustique du piston plan circulaire
- Annexe 2. Résolution de l'équation d'onde en coordonnées sphériques
- Annexe 3. Mesures de la source GSR
- Annexe 4. Mesures de la directivité du saxophone
- Annexe 5. Mesures de la directivité du cor de basset
- Annexe 6. Complementary data and initial results related to the four primary studies
- Annexe 7. Illustration of the experimental graphical user interfaces used in experiments 1 and 2
- Annexe 8. Details of acoustic features calculation

Annexe 1. Rayonnement acoustique du piston plan circulaire

Soit une source étendue de surface S encastrée dans un plan infini rigide S_H , l'**intégrale de Rayleigh** donne l'expression de la pression p en un point d'observation M situé dans le demi-espace en avant de la surface S , en fonction de la vitesse normale V_n sur la surface S (voir figure 1-1) :

$$p(r) = \frac{j\omega\rho_0}{2\pi} \int_S \frac{V_n(r_0)}{R} \cdot e^{-jkR} dS \quad \text{où } R = |r - r_0| \quad (\text{Eq. 1-1})$$

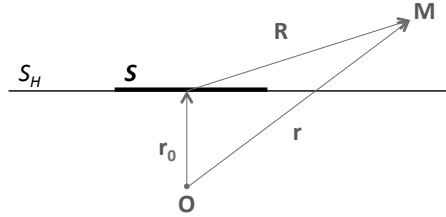


Figure 1-1. Source plane de surface S encastrée dans un plan infini S_H . A partir d'une origine O (arbitraire), r_0 : coordonnées du point courant sur la surface S ; r : coordonnées du point d'observation M .

Dans les conditions de champ lointain, c'est-à-dire lorsque $|r| \gg |r_0|$, on peut utiliser les premiers éléments du développement en série de R , à savoir :

$$R = |r - r_0| \approx r - r_0 \cdot \frac{r}{|r|} \quad (\text{Eq. 1-2})$$

afin d'écrire une approximation de l'intégrale de Rayleigh (approximation de Fraunhofer) sous la forme :

$$p(r) \approx j\omega\rho_0 \frac{e^{-jkr}}{2\pi r} \int_S V_n(r_0) \cdot e^{jk r_0 \cdot \frac{r}{|r|}} dS \quad (\text{Eq. 1-3})$$

Dans le cas du piston plan circulaire, les variables du problème peuvent s'exprimer dans un système de coordonnées sphériques (voir figure 1-2), soit :

$$r = (r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta) ; \quad r_0 = (r_0 \cos \varphi, r_0 \sin \varphi, 0)$$

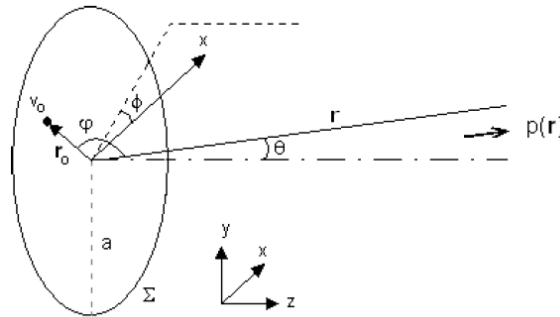


Figure 1-2. Piston plan – système de coordonnées sphériques (voir aussi Figure 23)

ce qui permet d'exprimer les éléments de l'équation Eq. 1-3 sous la forme :

$$r_0 \cdot \frac{r}{|r|} = r_0 \sin \theta (\cos \phi \cos \varphi + \sin \phi \sin \varphi) = r_0 \sin \theta \cos(\varphi - \phi) \quad (\text{Eq. 1-4})$$

et donc l'expression de l'approximation de la pression rayonnée en champ lointain par le piston plan, en fonction des variables d'angle et de distance, sous la forme :

$$p(r, \theta, \phi) \approx j\omega\rho_0 V_0 \frac{e^{-jkr}}{2\pi r} \int_0^a \int_0^{2\pi} e^{jkr_0 \sin\theta \cos\psi} r_0 d\psi dr_0 \quad \text{où } \psi = \varphi - \phi \quad (\text{Eq. 1-5})$$

Par ailleurs, en faisant intervenir la fonction de Bessel de 1^{ère} espèce d'ordre 0 :

$$J_0(z) = \frac{1}{2\pi} \int_0^{2\pi} e^{jz \cos\alpha} d\alpha \quad (\text{Eq. 1-6})$$

on peut écrire l'Eq.1-5, sous la forme :

$$p(r, \theta) \approx j\omega\rho_0 V_0 \frac{e^{-jkr}}{2\pi r} \int_0^a J_0(kr_0 \sin\theta) r_0 dr_0 \quad (\text{Eq. 1-7})$$

De plus, en utilisant la propriété reliant les fonctions de Bessel d'ordre 0 et 1 :

$$J'_1(z) = J_0(z) - \frac{J_1(z)}{z} \Leftrightarrow \int z \cdot J_0(z) dz = z \cdot J_1(z) \quad (\text{Eq. 1-8})$$

on peut exprimer la pression en champ lointain au moyen de la fonction de Bessel de 1^{ère} espèce d'ordre 1, selon le calcul suivant :

$$\begin{aligned} p(r, \theta) &\approx j\omega\rho_0 V_0 \frac{e^{-jkr}}{2\pi r} \cdot \frac{1}{(k \sin\theta)^2} \int_0^{ka \sin\theta} J_0(\mu) \mu d\mu \\ p(r, \theta) &\approx j\omega\rho_0 V_0 \frac{e^{-jkr}}{2\pi r} \cdot \frac{1}{(k \sin\theta)^2} [\mu J_1(\mu)]_0^{ka \sin\theta} \\ p(r, \theta) &\approx j\omega\rho_0 V_0 \frac{e^{-jkr}}{2\pi r} \cdot a \cdot \frac{J_1(ka \sin\theta)}{k \sin\theta}, \quad \text{avec } k = \omega/c \end{aligned} \quad (\text{Eq. 1-9})$$

et aboutir ainsi à l'Eq. 13 du § II.2.1 :

$$p(r, \theta) = j\rho_0 c V_0 \cdot \frac{ka^2}{2} \cdot \frac{e^{-jkr}}{r} \cdot \left[\frac{2J_1(ka \sin\theta)}{ka \sin\theta} \right] \quad (\text{Eq. 1-10})$$

qui fait apparaître une fonction de directivité ($D(\theta)$ – voir § II.2.1) faisant intervenir la fonction caractéristique $J_1(z)/z$ (voir figure 1-3).

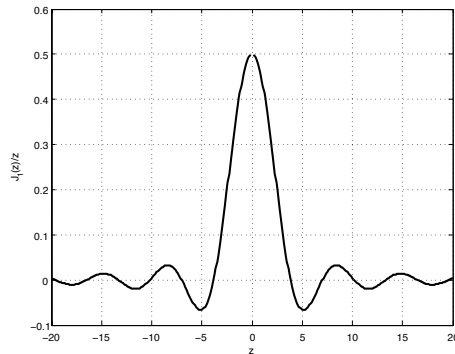


Figure 1-3. Représentation graphique de la fonction $J_1(z)/z$

Annexe 2. Résolution de l'équation d'onde en coordonnées sphériques

L'expression générale de l'équation d'onde est donnée par :

$$\frac{\partial^2 p}{\partial t^2} - c^2 \nabla^2 p = 0 \quad \text{où } \nabla^2 : \text{Laplacien} \quad (\text{Eq. 2-1})$$

Dans un système de coordonnées sphériques (voir figure 2-1), l'opérateur Laplacien prend une forme complexe, ce qui amène à faire l'hypothèse de séparation des variables spatiales et temporelles :

$$p(\vec{x}, t) = X(\vec{x})T(t) \quad (\text{Eq. 2-2})$$

où $T(t)$ s'exprime sur une base d'exponentielles complexes de pulsations $\omega = k.c$

On s'intéresse alors uniquement à la recherche des solutions spatiales satisfaisant à l'équation impliquant la variable spatiale x (équation différentielle d'Helmholtz) :

$$\nabla^2 X(\vec{x}) + k^2 X(\vec{x}) = 0 \quad (\text{Eq. 2-3})$$

Par ailleurs, dans un système de coordonnées sphériques (voir figure 2-1), cette variable spatiale s'exprime de la manière suivante :

$$x = r \sin \theta \cos \phi ; \quad y = r \sin \theta \sin \phi ; \quad z = r \cos \theta$$

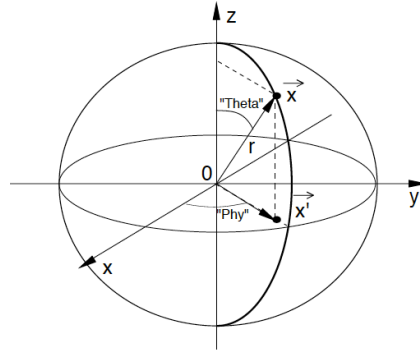


Figure 2-1. Système de coordonnées sphériques

ce qui permet d'expliciter l'opérateur Laplacien et d'aboutir à la formulation de l'Eq. 2-3 sous la forme suivante :

$$\left[\frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) \right] X(r, \theta, \phi) + r^2 k^2 X(r, \theta, \phi) = 0 \quad (\text{Eq. 2-4})$$

La résolution de l'Eq. 2-4 passe à nouveau par une séparation des variables spatiales – r , θ et ϕ , – en introduisant la fonction R qui décrit la partie radiale du champ sonore et les fonctions Θ et Φ qui représentent la distribution angulaire de ce même champ :

$$X(r, \theta, \phi) = R(r) \cdot \Theta(\theta) \cdot \Phi(\phi) \quad (\text{Eq. 2-5})$$

Le développement de l'Eq. 2-4 conduit alors à une formulation à variables séparées :

$$\frac{r^2 \sin^2 \theta}{R} \frac{d^2 R}{dr^2} + \frac{2r \sin^2 \theta}{R} \frac{dR}{dr} + r^2 k^2 + \frac{1}{\Phi} \frac{d^2 \Phi}{d\phi^2} + \frac{\sin \theta \cos \theta}{\Theta} \frac{d\Theta}{d\theta} + \frac{\sin^2 \theta}{\Theta} \frac{d^2 \Theta}{d\theta^2} = 0 \quad (\text{Eq. 2-6})$$

En limitant la description du champ sonore à la distribution angulaire, c'est-à-dire en ne considérant pas les phénomènes d'expansion radiale dans la recherche de solutions, on obtient pour l'Eq. 2-5 ci-dessus :

- une équation en ϕ d'expression suivante :

$$\frac{1}{\Phi} \frac{d^2 \Phi}{d\phi^2} + m^2 = 0 \quad (\text{Eq. 2-7})$$

dont les solutions s'expriment sous la forme :

$$\Phi(\phi) = \frac{1}{\sqrt{2\pi}} \cdot e^{im\phi} \quad (\text{Eq. 2-8})$$

- une équation en θ d'expression suivante :

$$\frac{1}{\sin^2 \theta} \left[-m^2 + \frac{\sin \theta \cos \theta}{\Theta} \frac{d\Theta}{d\theta} + \frac{\sin^2 \theta}{\Theta} \frac{d^2 \Theta}{d\theta^2} \right] + l = 0 \quad (\text{Eq. 2-9})$$

qui peut également s'écrire grâce à l'équation différentielle de Legendre, en choisissant la constante $l = n(n+1)$ et le changement de variable $x = \cos \theta$:

$$(1-x^2) \frac{d^2 \Theta}{dx^2} - 2x \frac{d\Theta}{dx} + \left(n(n+1) - \frac{m^2}{1-x^2} \right) \Theta = 0 \quad (\text{Eq. 2-10})$$

dont les solutions s'expriment sous la forme :

$$\Theta(x) = \sqrt{\frac{2n+1}{2}} \sqrt{\frac{(n-m)!}{(n+m)!}} P_n^m(x) \quad (\text{Eq. 2-11})$$

où les fonctions $P_n^m(x)$ sont les fonctions associées de Legendre, dérivées des polynômes de Legendre et s'expriment sous la forme :

$$P_n^m(x) = (-1)^m (1-x^2)^{m/2} \cdot \frac{d^m}{dx^m} P_n(x), \quad \text{pour } m \geq 0 \quad (\text{Eq. 2-12})$$

$$P_n^{-|m|}(x) = (-1)^{|m|} \frac{(n-|m|)!}{(n+|m|)!} P_n^{|m|}(x), \quad \text{pour } m < 0$$

et où $P_n(x)$ est le polynôme de Legendre d'ordre n d'expression :

$$P_n(x) = \frac{1}{2^n n!} \cdot \frac{d^n}{dx^n} (x^2 - 1)^n \quad (\text{Eq. 2-13})$$

La solution générale de l'équation d'onde – sous l'hypothèse d'une solution à variables séparées et en ne considérant que les variables angulaires ($R(r) = 1$) – aboutit ainsi à la formulation des harmoniques sphériques. On retrouve alors l'expression donnée au § II.2.2 (Eq. 15) :

$$\Theta(\theta) \Phi(\phi) = Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} \cdot P_n^m(\cos \theta) \cdot e^{jm\phi} \quad (\text{Eq. 2.14})$$

Annexe 3. Mesures de la source GSR

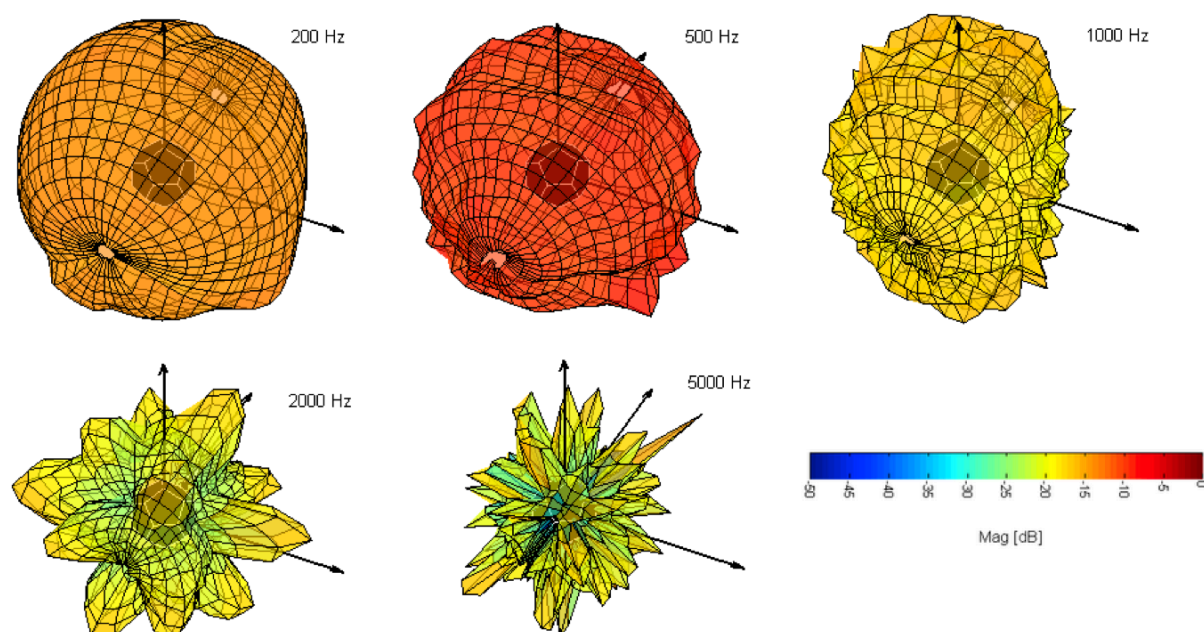


Figure 3-1. Mesure de la source GSR en configuration 'OMNI'

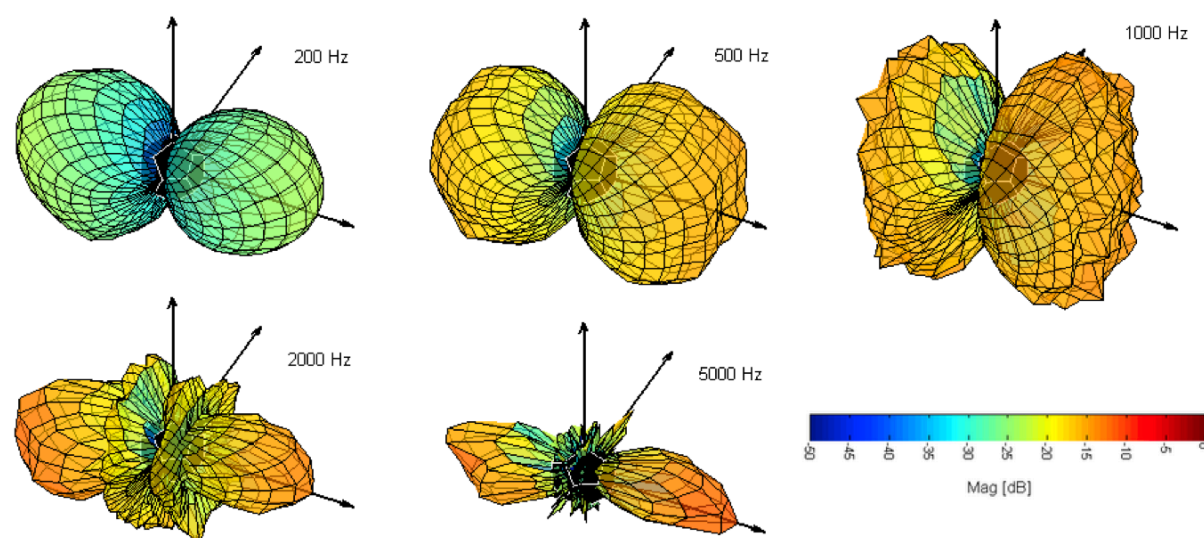


Figure 3-2. Mesure de la source GSR en configuration 'DIPO' (amplitude)

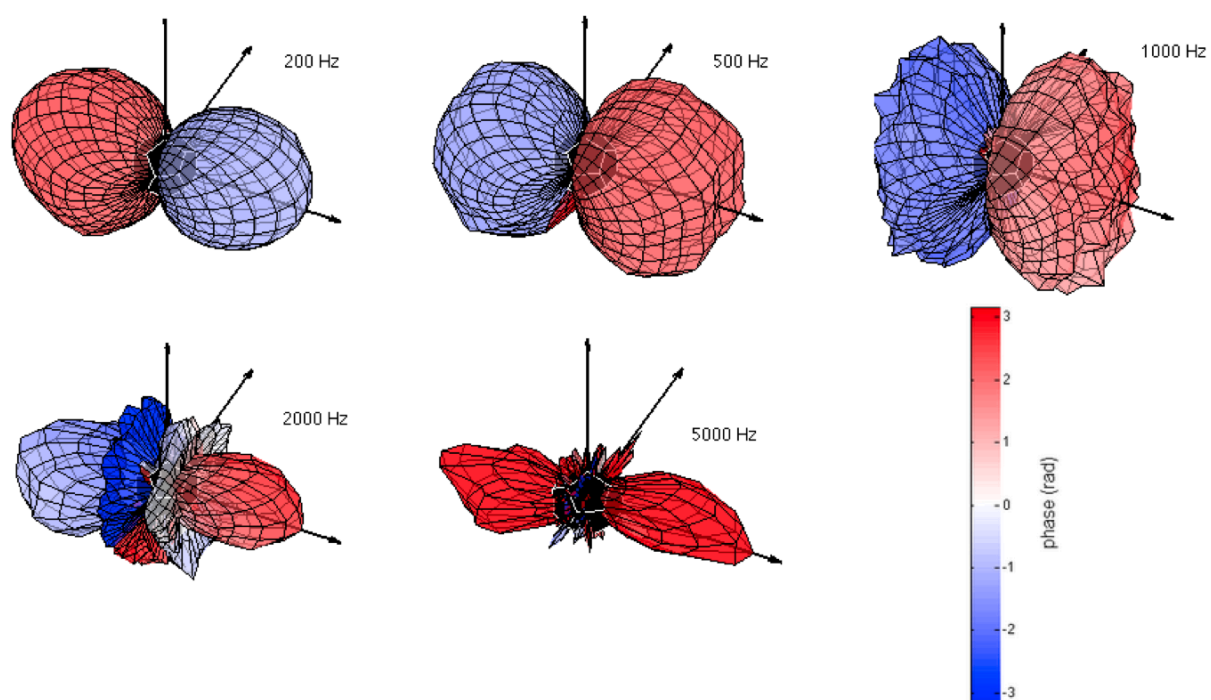


Figure 3-3. Mesure de la source GSR en configuration 'DIPOL' (phase)

Annexe 4. Mesures de la directivités du saxophone

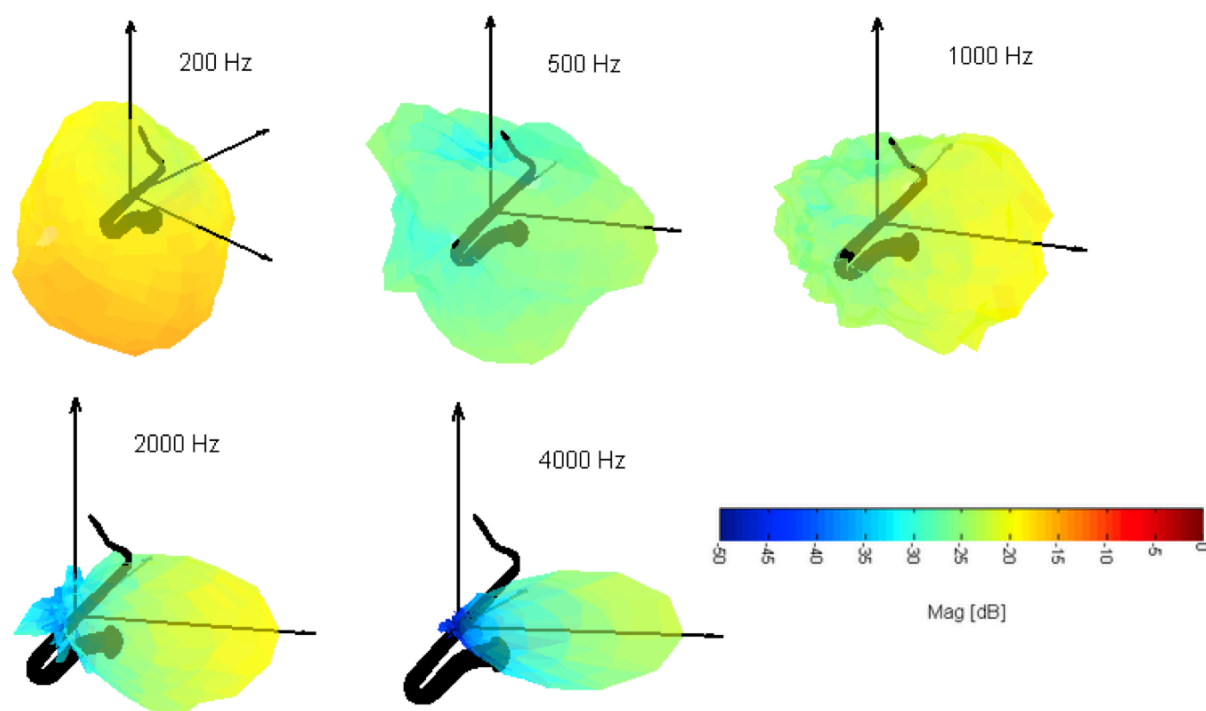


Figure 4-1. Mesure de la directivité du saxophone, configuration fermée (La bémol 104 Hz)

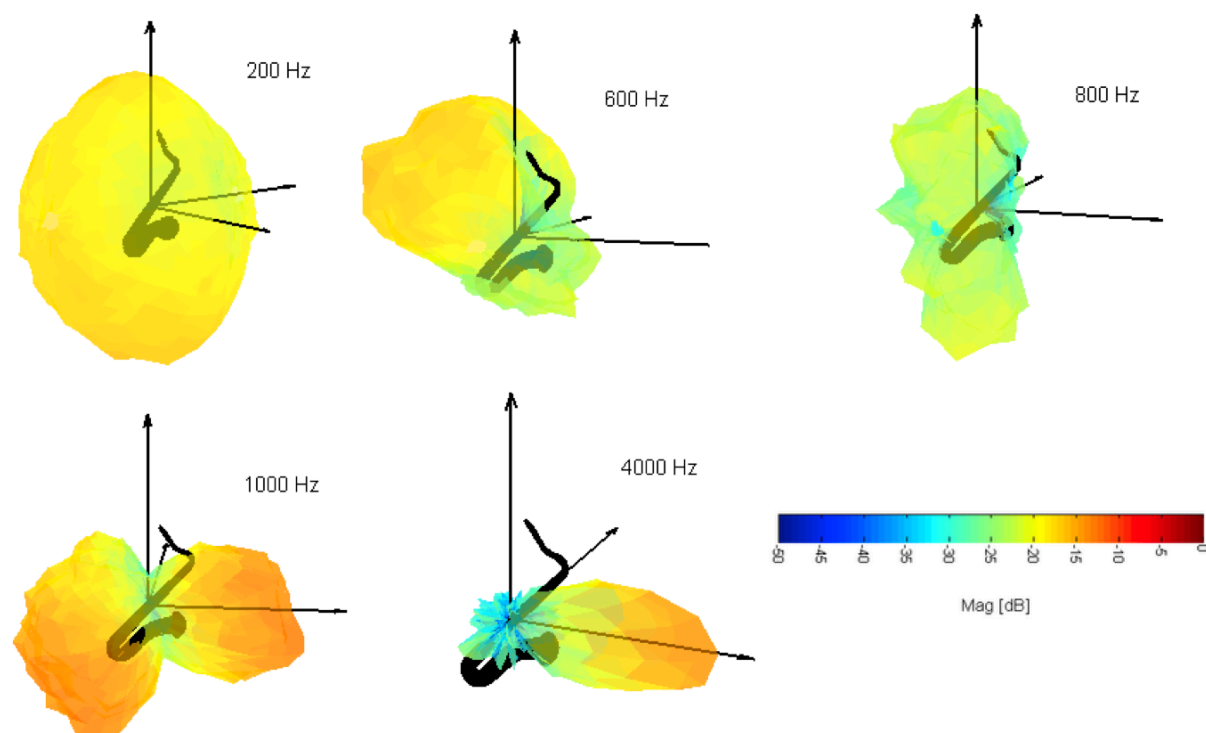


Figure 4-2. Mesure de la directivité du saxophone, configuration semi-fermée (Fa 175 Hz)

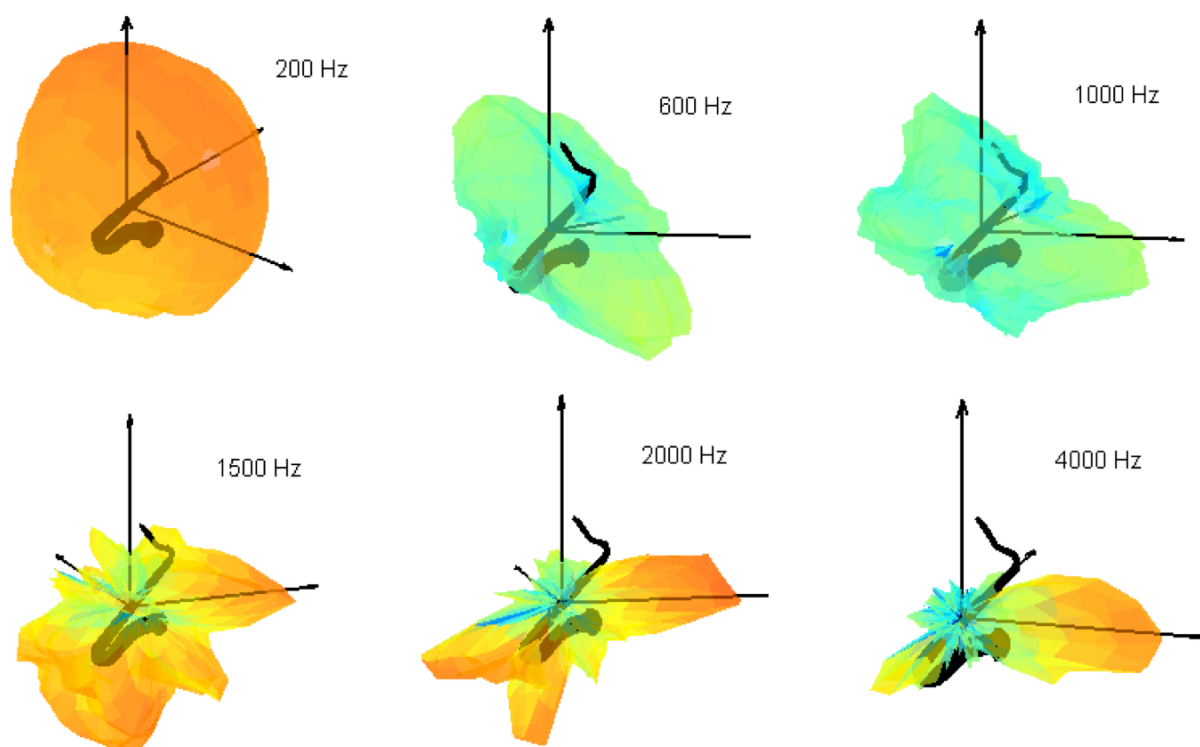


Figure 4-3. Mesure de la directivité du saxophone, configuration aigue (Mi bémol 622 Hz)

Annexe 5. Mesures de la directivité du cor de basset

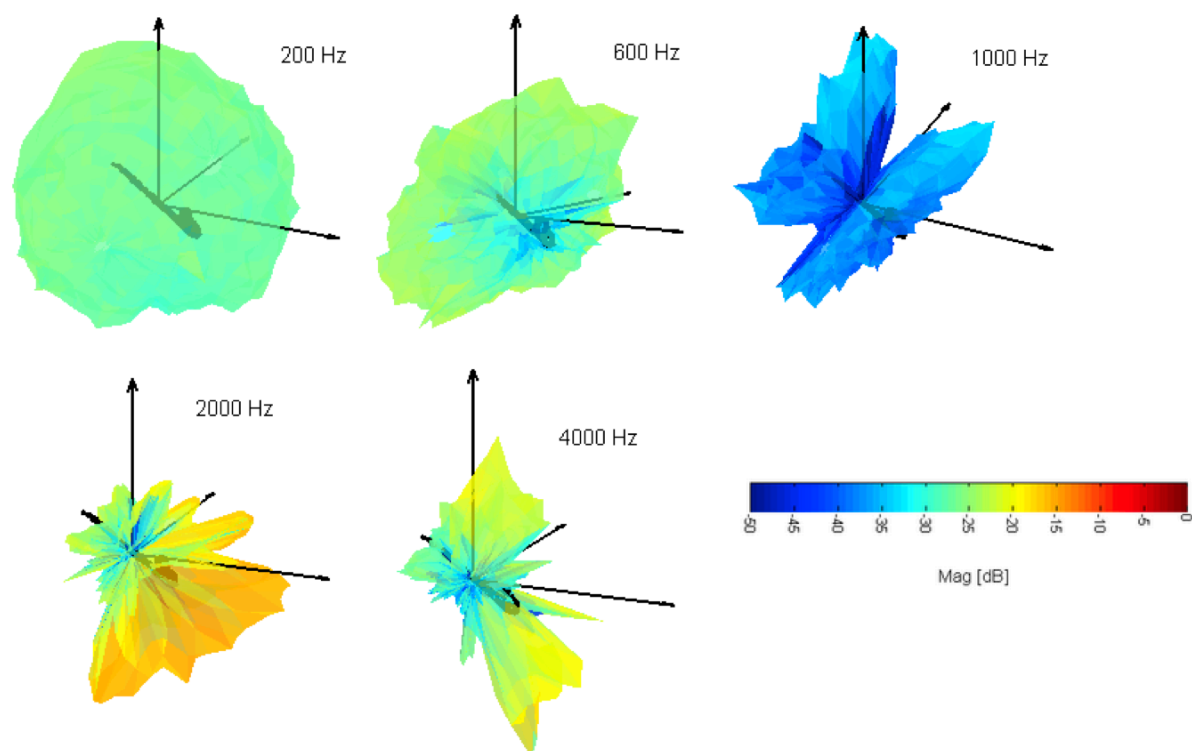


Figure 5-1. Mesure de la directivité du cor de basset, doigté ouvert (Do 261 Hz)

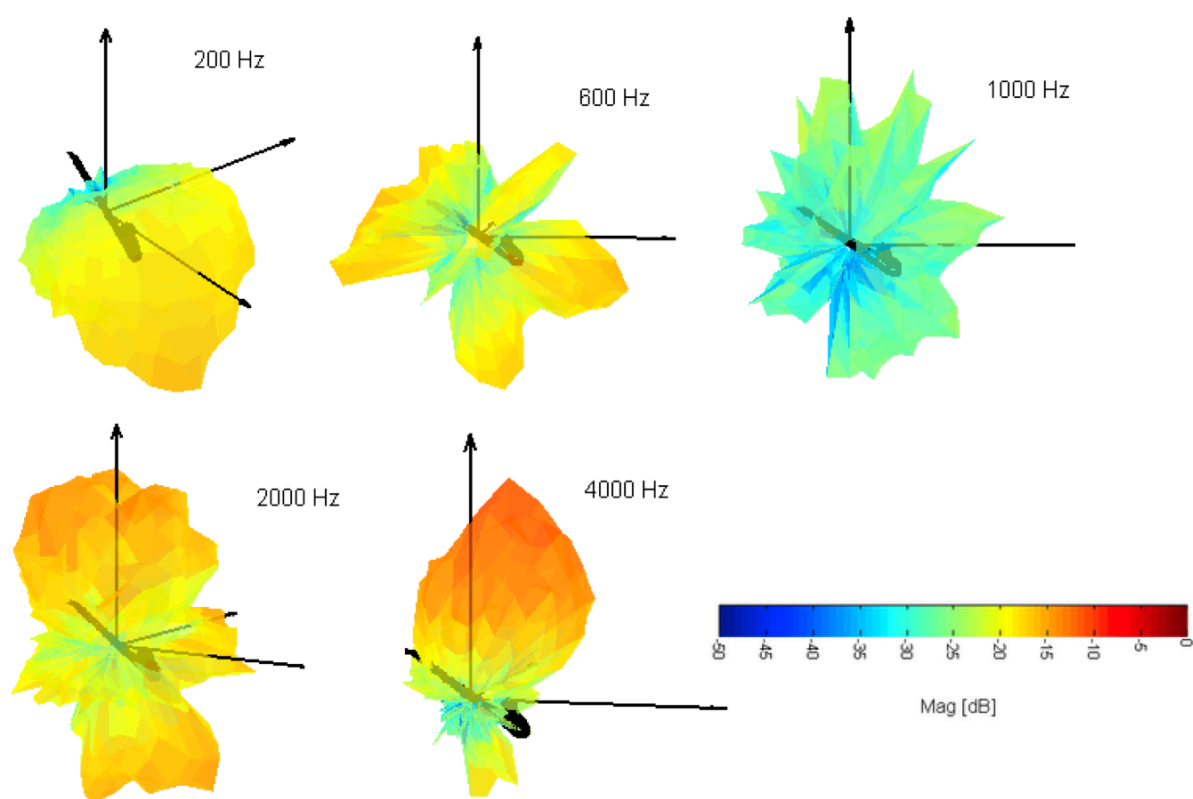


Figure 5-2. Mesure de la directivité du cor de basset, doigté fermé (Fa 87 Hz)

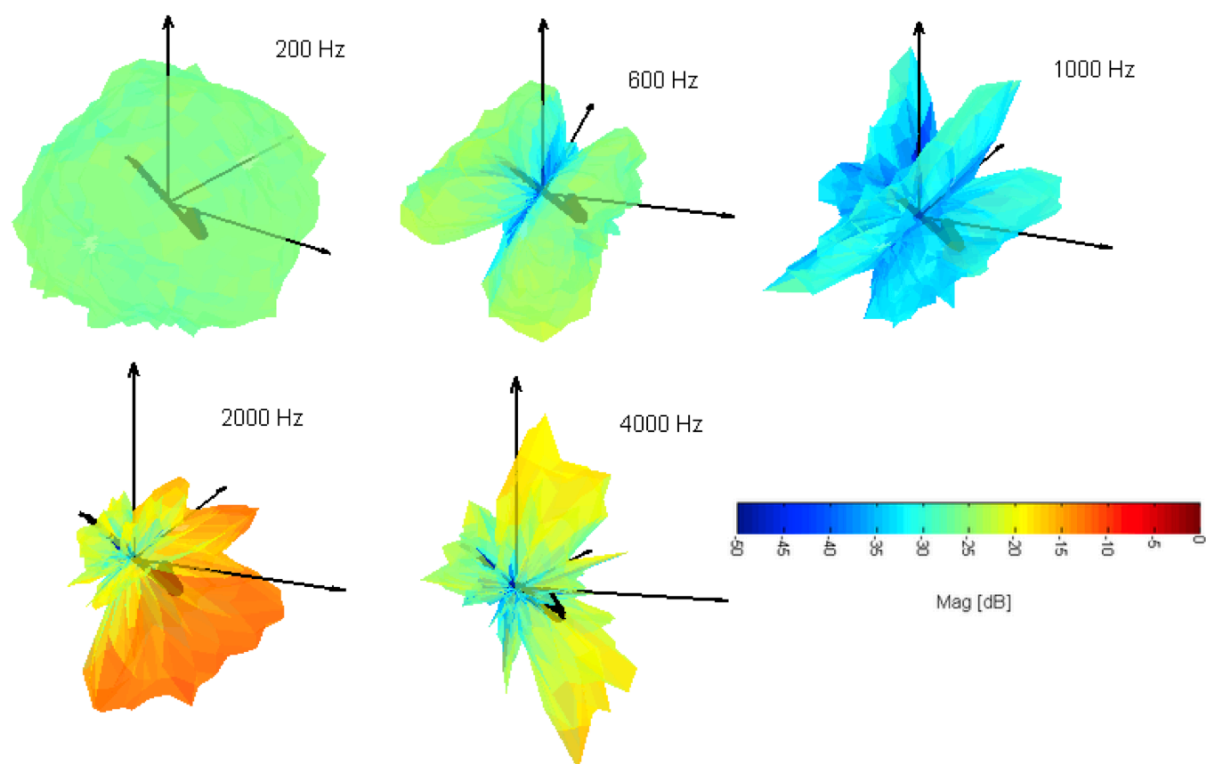


Figure 5-3. Mesure de la directivité du cor de basset, doigté semi-fermé (Fa 175 Hz)

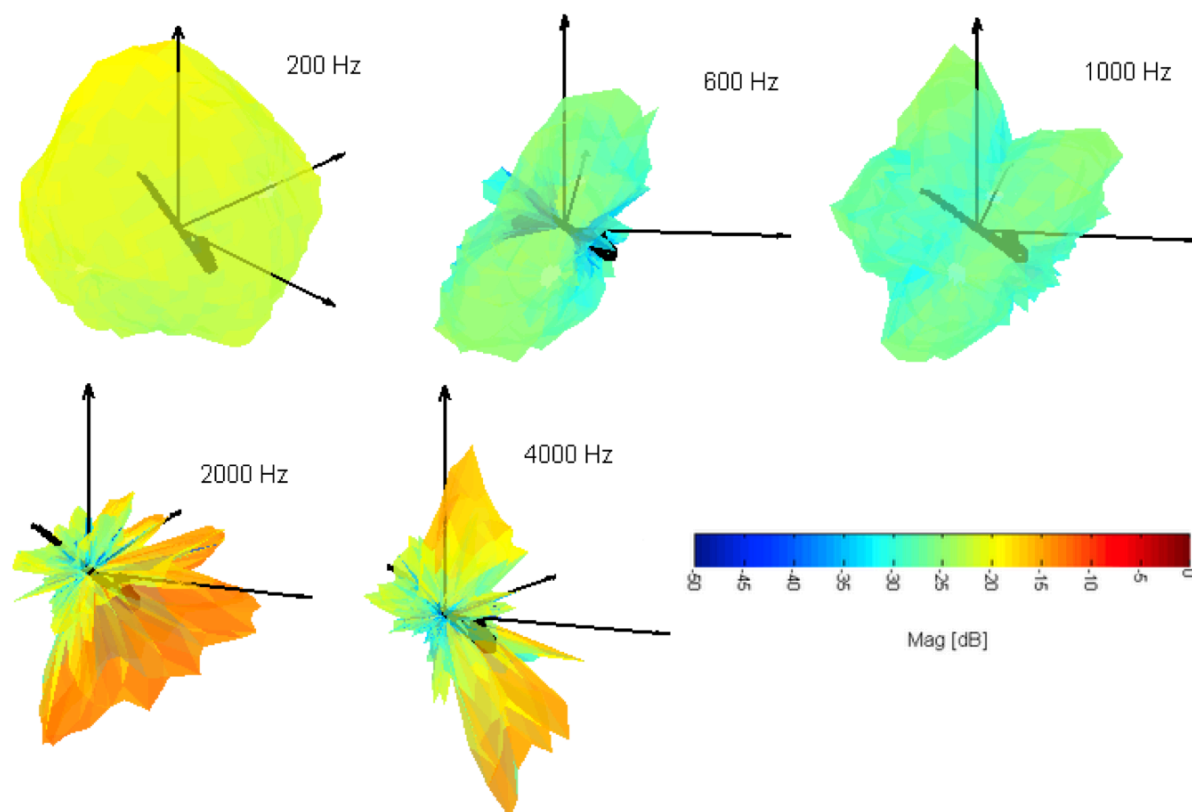


Figure 5-4. Mesure de la directivité du cor de basset, doigté aigu (Si 494 Hz)

Annexe 6 : Complementary data and initial results related to the 4 primary studies

Data Related to study A1

	Dimension 1	Dimension 2	Dimension 3
RAPmv-A	-0.81**	0.32	-0.33
CGg-ERB	0.35	-0.7**	-0.14
Dec (266-2300)	-0.32	0.00	-0.83**

Table 6-1. Correlation coefficients between the perceptual dimensions of study A1 and psychoacoustic descriptors ($df=14$, ** $p<0.01$).

- **RAPmv-A:** A-weighted harmonic-to-noise ratio. Both harmonic and noise parts were separated using additive analysis/synthesis (see [Rodet, 2001], for more detail on the separation technique). The feature is the ratio of their levels expressed in dB(A).
- **CGg-ERB:** ERB Spectral centroid. The frequency dimension is represented in ERB-rate (distance in terms of Equivalent-Rectangular Bandwidth (ERB) filters; see [Patterson et al., 1995] and [Slaney, 1993]).
- **Dec:** Harmonic spectral decrease. This feature is related to the shape of the spectral envelope computed from the harmonic components of the signal. In the present case, this feature is computed on the bandwidth of the spectrum, but represents the relative decrease in the envelope of the harmonic spectrum only between 266 Hz and 2300 Hz.

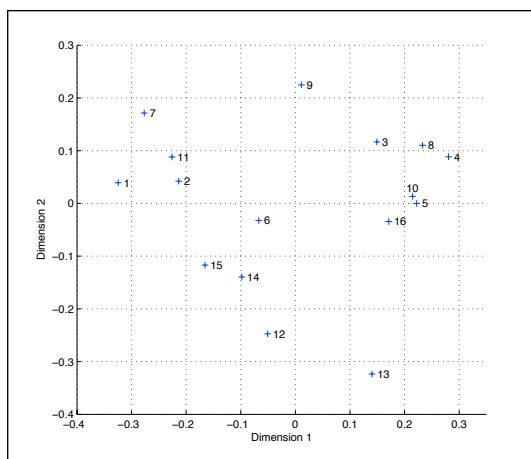


Figure 6-1. Study A1 perceptual space projected onto dimensions 1 and 2.

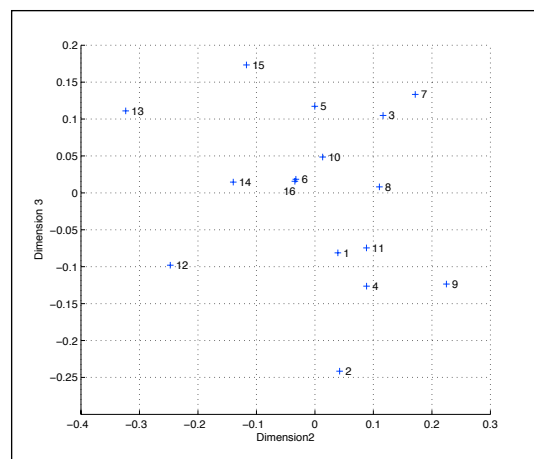


Figure 6-2. Study A1 perceptual space projected onto dimensions 2 and 3.

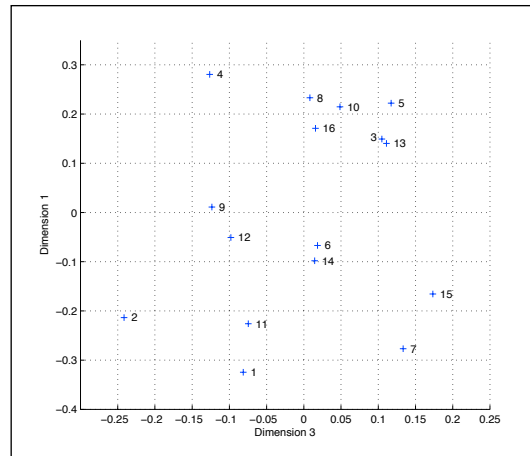


Figure 6-3. Study A1 perceptual space projected onto dimensions 3 and 1.

Data Related to study A2

	Dimension 1	Dimension 2
rad_2N/0.5N	0.93**	-0.29
CGg-C	-0.51	0.86**

Table 6-2. Correlation scores between the perceptual dimensions of study A2 and acoustic features ($df = 12$, ** $p < 0.01$).

- **rad_2N/0.5N**: 2N and 0.5N harmonics ratio, where N is deduced from the RPM value of engine rotation.
- **CGg-C**: Spectral centroid, with linear frequency using C-weighting.

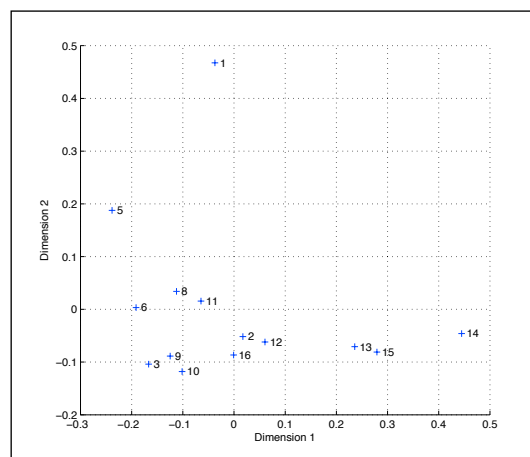


Figure 6-4. Study A2 perceptual space projected onto dimensions 1 and 2.

Data Related to study B

	Dimension 1	Dimension 2	Dimension 3
NHR-A	-0.97**	0.11	-0.26
SC_n-B	-0.32	0.73**	-0.15
N	0.26	0.04	0.84**

Table 6-3. Correlation coefficients between the perceptual dimensions of study B and acoustic features ($df=17$, ** $p < 0.01$).

- **NHR-A:** Feature corresponding to the relative balance of the harmonic (motor) and noise (air turbulence) components. The best correlation is obtained with the A-weighted version of this parameter.
- **SC_n-B:** B-weighted spectral centroid of the noise component. For this dimension, the emergence of a spectral pitch led us to consider the spectral centroid (SC). More precisely, we compute the SC of each of the two parts of the sound: the noise component (SC_n) and the harmonic component (SC_h). The best correlation with Dimension 2 is obtained for SC_n using B-weighting.
- **N:** Loudness. Indeed, even though the selected sounds are in the same range of loudness, they were not equalized in loudness.

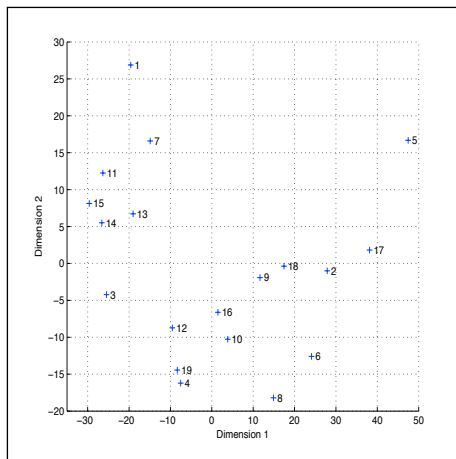


Figure 6-5. Study B perceptual space projected onto dimensions 1 and 2.

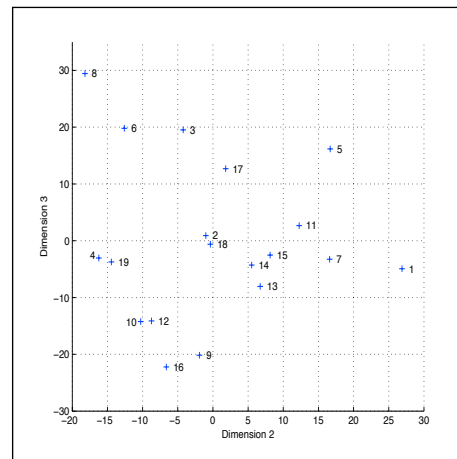


Figure 6-6. Study B perceptual space projected onto dimensions 2 and 3.

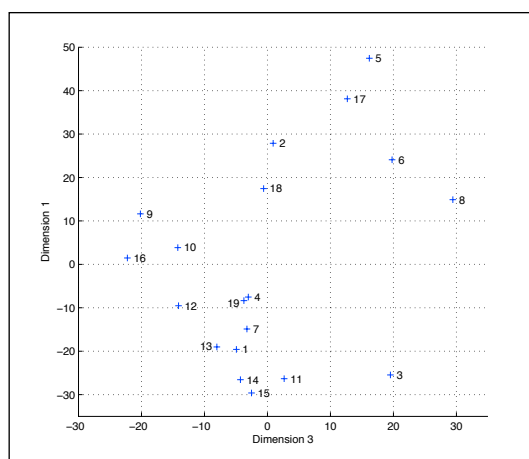


Figure 6-7. Study B perceptual space projected onto dimensions 3 and 1.

Data Related to study C

	Dimension 1	Dimension 2	Dimension 3
Roughness	-0.9**	-0.1	0.3
Spectral centroid	0.0	0.9**	0.1
Spectral deviation	0.3	-0.4	-0.8**

Table 6-4. Correlation coefficients between the perceptual dimensions of study C and the best-correlated psychoacoustic descriptors ($df = 20$, ** $p < 0.01$).

- **Roughness:** Feature modeled by the amplitude modulation rate of the temporal envelope (expressed in asper) and related to the sensation of auditory roughness.
- **Spectral centroid:** Feature describing the spectral distribution of the energy of the sound, computed from a frequency decomposition on the ERB scale [Marozeau et al., 2003]. It has been identified as corresponding to the sensation of "brightness".
- **Spectral deviation:** Feature related to the fine structure of the spectral envelope. It is computed based on the smoothness of the outputs of the filter-bank [Marozeau et al., 2003].

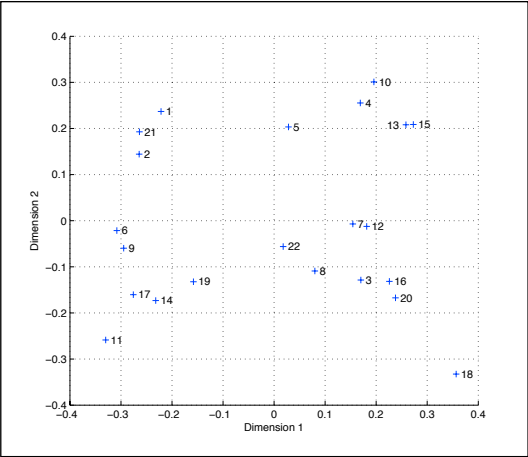


Figure 6-8. Study C perceptual space projected onto dimensions 1 and 2.

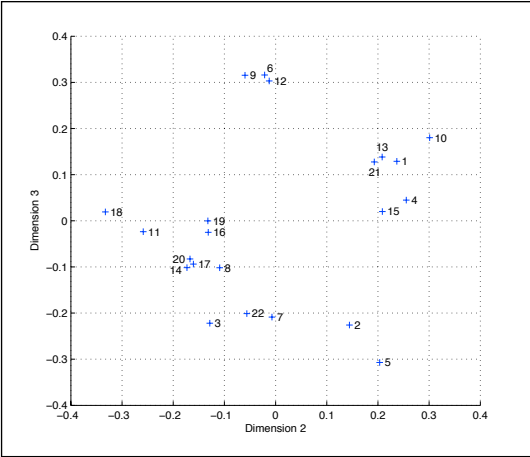


Figure 6-9. Study C perceptual space projected onto dimensions 2 and 3.

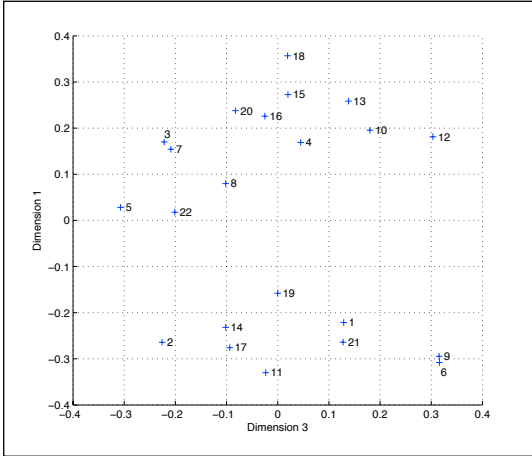


Figure 6-10. Study C perceptual space projected onto dimensions 3 and 1.

Data Related to study D

	Dimension 1	Dimension 2	Dimension 3
Sharpness	-0.90		
Spectral centroid	-0.93		
Cleanness indicator		0.87	
.....		

Table 6-5. Correlation coefficients between the perceptual dimensions of study D and acoustic features (*df* = 10).

- **Spectral centroid:** Feature describing the spectral distribution of the energy of the sound.

- **Sharpness** Feature defined by Aures [Aures, 1985], similar to spectral centroid with perceptual modeling.
- **Cleanness indicator:** Indicator that is derived from the temporal loudness calculation according to Zwicker's model [Zwicker et al., 1990]. The algorithm takes into account temporal integration and temporal masking. The proposed indicator is based on the temporal evolution of the curve.

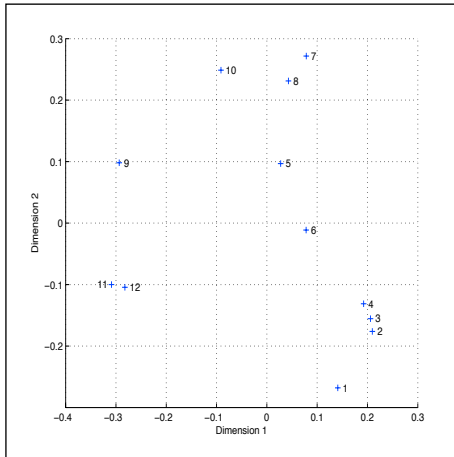


Figure 6-11. Study D perceptual space projected onto dimensions 1 and 2.

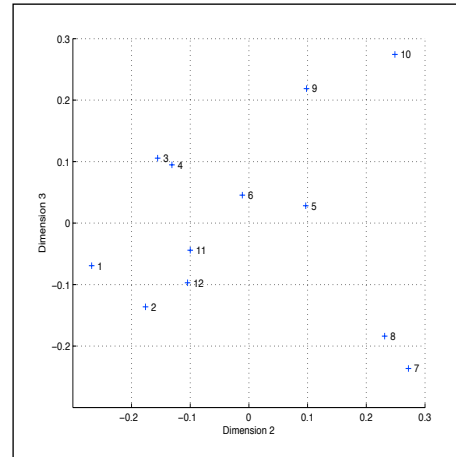


Figure 6-12. Study D perceptual space projected onto dimensions 2 and 3.

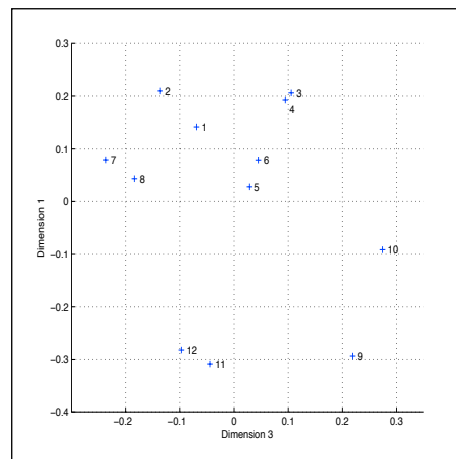


Figure 6-13. Study D perceptual space projected onto dimensions 3 and 1.

Annexe 8 : Details of acoustic features calculation

RMS value

The estimation of the RMS (Root-Mean-Square) value of the signal is frame-based and is calculated every 60 ms with a Blackman window. The feature is the mean value over time.

Loudness

Loudness is the intensive attribute of human hearing. It thus describes the subjective aspect of the intensity of a signal by considering masking effects that occur over the whole spectrum and the filtering steps of the hearing path. The loudness model used is the ISO 532-B model from [Zwicker et al., 1990].

Harmonic emergence feature

This feature is a Harmonic-to-Noise ratio, designed to convey the relative amounts of harmonic (or pseudo-harmonic) energy and noise energy in the signal. It is based on the *Pm2* partial extraction method (see [Bogaards al., 2004]). Once both harmonic and noise parts of the signal are extracted, the feature simply consists of the ratio of their respective loudnesses N_h and N_n as formalized in Eq. 8-1:

$$HNR = N_h / N_n \quad (\text{Eq. 8-1})$$

Spectral centroid

The spectral centroid is a weighted mean frequency of the spectrum of the signal. The calculation of this feature can be more or less complex. Its definition is quite similar to Zwicker et al.'s *sharpness* feature [Zwicker et al., 1990]. It uses a gammatone filterbank (from Auditory Toolbox, Slaney [Slaney, 1998]) that is based on the ERB-rate scale z (see [Marozeau et al., 2003] for more details). The resulting feature is the Perceptual Spectral Centroid as defined in Eq. 8-2:

$$PSC = \sum_z f_z \cdot N_z / \sum_z N_z \quad (\text{Eq. 8-2})$$

where N_z is the specific loudness in each channel (obtained by each gammatone filter) and f_z is the corresponding center frequency.

Spectral spread

The spectral spread describes how the spectrum is spread around its mean value, i.e. the spectral centroid defined above. The associated perceptual feature uses the same perceptual modeling as the *PSC* feature, thus giving the Perceptual Spectral Spread *PSS*, as defined in Eq. 8-3

$$PSS = \sum_z (f_z - PSC)^2 \cdot N_z / \sum_z N_z \quad (\text{Eq. 8-3})$$

Complex Brightness

This feature estimates the *brightness* sensation of a sound that combines a noisy and a harmonic part. It simply corresponds to the linear combination of the *PSC* values of both noisy and harmonic parts (respectively PSC_h and PSC_n) and the *PSS* value of the whole signal, as defined in Eq. 8-4:

$$\text{Complex brightness} = \alpha \cdot PSC_h + \beta \cdot PSC_n + \gamma \cdot PSS \quad (\text{Eq. 8-4})$$

where α , β and γ are linear coefficients.

Roughness

Roughness is a feature that quantifies the perceived modulation or graininess of a sound. When inharmonicity is strong, amplitude modulations can generate beating in some cases. When the beating becomes fast enough so that the modulations are no longer discriminated by the human ear, they seem to give a *rough* aspect to the sound. This roughness feature (also defined in [Grey, 1977]) mainly consists in estimating a modulation index at the output of every auditory filter, which is called the *partial roughness*. The overall roughness is the sum of all the partial roughnesses. From each auditory filter output, the modulation frequency f_{modi} and the modulation depth m_i are estimated with a temporal envelope calculation. The partial roughness is proportional to the product of the modulation frequency and the depth $f_{modi} \cdot m_i$. The roughness R is then calculated as the sum of the R_i , as mentioned in Eq. 8-5:

$$R_i = K \cdot f_{modi} \cdot m_i \quad ; \quad R = \sum_i R_i \quad (\text{Eq. 8-5})$$

where K is the proportionality coefficient

Cleanness indicator

This feature represents the short-term variations in the loudness of the signal. These variations, which usually occur between 20 and 100 Hz, are slow enough to be heard as a temporal phenomenon, but they are too fast to be heard as separate sound events (e.g., bounces, rattles, etc.). The feature corresponds to the amplitude of the spectrum of the instantaneous loudness $N(t)$, which is estimated every 3.3 msec., within this frequency band (see Eq. 8-6).

$$\text{Cleanness indicator} = \sum_{20-100 \text{ Hz}} |FFT_{256}(N(t))| \quad (\text{Eq. 8-6})$$

where FFT_{256} is the 256-point Fast Fourier Transform.

Nicolas MISDARIIS

Synthèse – Reproduction – Perception des Sons Instrumentaux et Environnementaux

Application au Design Sonore

Résumé

Ce mémoire présente une composition d'études et de travaux de recherche orientés autour de trois grandes thématiques : la synthèse, la reproduction et la perception des sons, en considérant à la fois les sons de nature musicale mais aussi environnementale. Il vise en outre un champ d'application, le design sonore, qui implique globalement la création intentionnelle de sons du quotidien. La structure du document est conçue selon un schéma relativement uniforme et comporte, pour chaque partie, une présentation générale de la thématique apportant des éléments théoriques et des données relatives à l'état de l'art, suivie de développements spécifiques permettant de converger vers les sujets d'étude propres à chaque thème – explicitement, formalisme modal dans la synthèse par modélisation physique, pour la partie "Synthèse" ; mesure et contrôle de la directivité des instruments de musique, pour la partie "Reproduction" ; timbre et identification des sources sonores, pour la partie "Perception" – puis d'une présentation détaillée des travaux personnels relatifs à chacun des sujets, le cas échéant, sous la forme d'un article publié. Ces divers éléments de connaissance et d'expérience proposent donc une contribution personnelle et originale, volontairement inscrite dans un cadre de recherche élargi, pluridisciplinaire et appliqué.

Mots-clés : synthèse sonore, modélisation physique, reproduction sonore, rayonnement acoustique, directivité, perception sonore, timbre, sons environnementaux, design sonore, sonification, interface homme-machine.

Résumé en anglais

This dissertation presents a composition of studies and research works articulated around three main topics : synthesis, reproduction and perception of sounds, considering both musical and environmental sounds. Moreover, it focuses on an application field, the sound design, that globally involves the conception of intentional everyday sounds. The document is based on a rather uniform structure and contains, for each part, a general presentation of the topic which brings theoretical elements together with an overview of the state-of-the-art, followed by more precise developments in order to focus on the specific matters related to each topic – in detail, modal formalism in sound synthesis by physical modeling, for the "Synthesis" section ; measurement and control of musical instruments directivity, for the "Reproduction" section ; timbre and sound sources identification, for the "Perception" section – and then followed by a detailed presentation of the personal works related to each matter, in some cases, in the form of published papers. Then, these several elements of knowledge and experience offer a personal and original contribution, deliberately put in a broad, multidisciplinary and applied framework.

Keywords : sound synthesis, physical modeling, sound reproduction, acoustic radiation, directivity, auditory perception, timbre, environmental sounds, sound design, sonification, computer-human interface.