Mémoire de fin d'études
Institut Supérieur de l'Electronique et du Numérique, Toulon
2007-2008

Catégorisation des sons de matériaux frappés : approches perceptive et statistique.

Patrick MARMAROLI

CNRS - Laboratoire de Mécanique et d'Acoustique
31 chemin Joseph Aiguier
13402 Marseille Cedex 20

Maîtres de stage : R. Kronland-Martinet et M. Aramaki

patrick.marmaroli@gmail.com

Table des matières

1 Introduction **7**

I Remerciements 7

II Résumé 8

III Abstract 9

IV Résumé des outils utilisés et du travail réalisé 10

2 Contexte général du stage **11**

I Le Laboratoire de Mécanique et d'Acoustique 11

II La problématique 13

3 Phase expérimentale **15**

I Constitution d'une base de données sonores 15

II Test perceptif 16

II.1 Protocole 16

II.2 Résultats obtenus pour les sons destinés à la calibration 17

II.3 Résultats obtenus pour les sons destinés à la validation 18

III Conclusion sur la phase expérimentale 20

4 Caractérisation acoustique des sons **21**

I Les descripteurs fréquentiels 22

I.1 Le Centre de Gravité Spectral : *CGS* 22

I.2 Le point de roll-off : *ROF* 23

I.3 L'étalement spectral : *SSP* 24

I.4 Le coefficient de dissymétrie : *SKN* 25

I.5 Le kurtosis : *KRT* 25

I.6 La décroissance spectrale : *SDC* 26

I.7 Le taux de croisement du spectre : *SCR* 26

II Les descripteurs temporels 26

II.1 Le logarithme du temps d'attaque : *LAT* . . . 27
II.2 Le centre de gravité temporel : *CGT* . . . 28
II.3 La valeur efficace : *RMS* . . . 29
II.4 La durée perceptive : *TED* . . . 29
II.5 L'amortissement : *AMO* . . . 30
II.6 Taux de passage par zéro : *ZCR* . . . 31
III Les descripteurs spectro-temporels . . . 32
III.1 Le flux spectral : *FSP* . . . 32
III.2 La rugosité : *RUG* . . . 33
III.3 Le centroïde de rugosité : *CGR* . . . 38
III.4 Le rapport CGR/CGS : *RSF* . . . 38
IV Conclusion sur la caractérisation acoustique . . . 39

5 Analyse statistique descriptive **41**
I Mise en forme des données . . . 41
II Présentation du logiciel *SPSS* . . . 42
III Analyse de corrélation . . . 43
IV L'Analyse en Composante Principale . . . 45
V Conclusion sur l'analyse descriptive . . . 47

6 Analyse statistique prédictive **49**
I Principe de la régression logistique binaire . . . 49
II Modèles prédictifs pour chaque catégorie . . . 50
II.1 Précautions d'usages . . . 50
II.2 Calibration . . . 51
II.3 Validation . . . 52
II.4 Autres calibrations . . . 53
III Modèle prédictif global . . . 56
IV Modèles prédictifs physiques . . . 57
IV.1 Modèle prédictif pour le type d'impact : dur / mou . . . 57
IV.2 Modèle prédictif du matériau réellement impacté . . . 57
V Conclusion sur l'analyse prédictive . . . 59

7 Application en temps réel **61**

8 Conclusion et perspectives **63**

Table des figures

3.1 Matériel utilisé pour les prises de sons 15

3.2 Interface développée pour le test perceptif 16

3.3 Sons destinés à la calibration. Taux de confusion (normalisé à 1) entre matériaux réellement impactés et matériaux perçus ; (à gauche) : représentation sous forme d'un diagramme en barres ; (à droite) : représentation sous forme de matrice, dans ce cas, les catégories « carton » et « céramique » ont été spécifiées car des enregistrements d'impacts sur des objets en carton et en céramique ont été soumis aux sujets. 17

3.4 Sons destinés à la validation. Taux de confusion (normalisé à 1) entre matériaux réellement impactés et matériaux perçus ; (à gauche) : représentation sous forme d'un diagramme en barres ; (à droite) : représentation sous forme de matrice. 19

4.1 CGS d'un son d'impact de verre 23

4.2 Amplitude (en haut) et énergie cumulée (en bas) du spectre d'un impact de verre. Le *ROF* avec un seuil de 85% est représenté sur les deux graphes par une ligne rouge verticale. . . 24

4.3 *SSP* d'un son d'impact de verre 25

4.4 *SCR* d'un son d'impact de bois (à gauche) et de verre (à droite). 26

4.5 Etapes de l'estimation du temps attaque sur un impact de métal. 28

4.6 Estimation du temps d'attaque d'un impact de bois, le *LAT* est le logarithme du temps qui s'écoule entre les deux points rouges. 28

4.7 *CGT* d'un son d'impact de verre 29

4.8 *TED* d'un son d'impact de verre (calculé sur le module du signal) : durée pendant laquelle le signal est au dessus de 40% de sa valeur maximale. 30

4.9 (en haut) : module du signal analytique calculé à partir de la transformée de Hilbert (décroissance exponentielle), (au milieu) : enveloppe du signal analytique estimée avec un filtre de butterworth, (en bas) : logarithme de l'enveloppe (comportement linéaire) et son approximation par une droite avec la fonction *polyfit* de Matlab. 31

4.10 (en haut) : représentation temps fréquence d'un signal composé d'un chirp quadratique (0-100), d'un chirp linéaire (100-200) et d'un signal stationnaire sinusoïdal (200-300), (en bas) : les coefficients de corrélation de Pearson correspondant dont la somme normalisée par $\frac{T}{\Delta t}$ (equation 4.16) nous donnera le *FSP*. 33

4.11 Les qualités perceptives de deux sons purs en fonction de leur écart fréquentiel. Les frontiéres entre les régions ne sont pas abruptes, notamment la frontière supérieure entre rugoisté et deux sons distincts dépend du registre, d'aprés Helmholtz (1877). 34

4.12 Les filtres deviennent plus étroits en basse fréquence. Figure extraite de [LLT] 37

4.13 Rugosité d'un impact de verre, (en haut) : en fonction du temps et de la fréquence, plus la couleur est foncée plus la rugosité est importante, (en bas) : rugosité de l'impact au cours du temps. 38

5.1 Matrice de corrélation des descripteurs de signaux 44

5.2 Les axes extraits par l'ACP 46

5.3 Projection des sons typiques dans l'espace engendré par les deux premiers axes de l'analyse en composante principale. 47

7.1 Interface de Taping : résultats suite à l'enregistrement d'un impact sur un verre avec un maillet dur, (à gauche) représentations de quelques descripteurs ; 1^{er} graphe : le *CGT*, l'*AMO*, 2^{e} graphe : l'*ATT*, 3^{e} graphe : le *CGS*, le *ROF*, le *SSP*, 4^{e} graphe : le *TED*, (à droite) : diagramme en barre du haut : type d'impact (ici le modèle estime qu'il y a 69% de chances que le maillet soit dur), diagrammes en barre du milieu : matériau réellement impacté (93% de chances que ce soit réellement du verre), diagrammes en barre du bas : matériau perçu (90% de chances que l'on entende du verre), sur la droite : valeurs des descripteurs correspondant au son enregistré. 62

Chapitre 1

Introduction

I Remerciements

Mes premiers remerciements s'adressent à mes co-encadrants Richard Kronland-Martinet[1] et Mitsuko Aramaki[2] pour leur confiance, leurs conseils et tout le savoir qu'ils m'ont transmis avec passion et enthousiasme. Je remercie également Loïc Brancheriau[3] pour sa précieuse aide en statistiques, et Thierry Voinier[1] pour ses conseils en informatique et en traitement de signal. Je remercie également toute les personnes que j'ai pu côtoyer au sein du laboratoire (Solvi Ystad, Philippe Guillemain, Adrien Merer, Mathieu Barthet, Fabrice Silva, Thibault Necciari, Christophe Vergez, Benjamin Ricaud, mon collègue de bureau : Vijay Ratiney,...) pour leur contribution à une excellente ambiance de travail. Enfin je salue le courage de tous les auditeurs qui ont subi les tests perceptifs, rien de tout ce qui suit n'aurait été possible sans eux.

[1] CNRS - Laboratoire de Mécanique et d'Acoustique (LMA)

[2] Institut des Neurosciences Cognitives de la Méditerranée (INCM)

[3] Centre de coopération internationale en recherche agronomique pour le développement (CIRAD)

II Résumé

Ce projet de recherche vise à développer un algorithme de catégorisation des sons d'environnement produits par des objets impactés basé sur des critères perceptifs.

Nous avons limité cette étude à la catégorisation perceptive de cinq classes de matériaux : le bois, le métal, le plastique, la pierre et le verre. Pour chacune de ces classes, nous avons tenté de définir les constituants fondamentaux des signaux d'impacts susceptibles de fournir une information perceptive pertinente sur la nature du matériau perçu. Cet algorithme de catégorisation est basé sur des modèles de prédiction statistiques.

L'étude s'est articulé suivant six phases :

1 Constitution d'une banque de sons d'impacts sur des objets en bois, en métal, en verre, en pierre et en plastique issus de notre environnement quotidien. Les géométries et les structurations de ces objets ont été choisi aussi variées que possible afin que la banque de sons soit exhaustive.

2 Implémentation d'un test d'écoute pour définir l'attribution de ces sons à l'une des cinq classes perceptives considérées (bois, métal, verre, pierre ou plastique).

3 Définition de descripteurs basés sur les caractéristiques acoustiques et perceptives des signaux. Nous nous sommes basés sur des descripteurs existants pour caractériser le timbre des sons.

4 Calibration et validation de modèles statistiques de classification des sons pour chaque catégorie de matériau.

5 Construction d'un algorithme global de catégorisation.

6 Implémentation d'une application basée sur ces modèles en temps réel.

III Abstract

The aim of this project is to implement an algorithm for the categorization of impact sounds based on perceptual criteria.

This classification is limited to five classes of materials : wood, metal, stone, plastic and glass. For each class we investigated the fondamental features characterising the acoustic signal that provide a relevant information. This algorithm of classification is based on predictive statistical models.

The study was divided in six steps :

1 Construction of a bank impact sounds from different materials (wood, metal, glass, plastic and stone) recorded on everyday life objects. We considered objects of different sizes and shapes in order to have an exhaustive sound data bank.

2 Design of a listening test to determine sound categories from a perceptual point of view as a function of the five classes of materials.

3 Definition of the sound descriptors based on acoustic and perceptual characteristics of signals. We considered desriptors known to be relevant for the perception of timbre and we estimated this descriptors.

4 Calibration and validation of the statistic models of classification of sounds for each class of materials.

5 Construction of a global algorithm of categorization.

6 Implementation of a real time application based on these models.

IV Résumé des outils utilisés et du travail réalisé

Tests perceptifs

- Campagne de mesures pour constituer une banque de données sonores (525 sons d'impacts sur tout type d'objets et de matériaux).
- Utilisation de *Matlab* pour implémenter et mettre en oeuvre un test perceptif.

Analyse

- Utilisation de *Matlab* pour analyser les réponses des auditeurs.
- Utilisation de *Matlab* pour calculer les valeurs des descripteurs pour chaque son.

Calibration et validation des modèles prédictifs

- Utilisation de *SPSS* pour effectuer l'analyse en composante principale sur les descripteurs.
- Utilisation de *SPSS* et de *Matlab* pour mettre au point des modèles prédictifs perceptifs et physiques.
- Utilisation de *Matlab* pour valider les modèles.

Application

- Utilisation de *Matlab* pour élaborer une application en temps réel.
- Démos devant les diverses personnes du laboratoire.

Autres

- *Latex* pour rédiger ce rapport.
- *PowerPoint* pour la soutenance.
- OS utilisés : *Windows XP* et *Mac OS 10.4.11*

Chapitre 2

Contexte général du stage

I Le Laboratoire de Mécanique et d'Acoustique

Présentation générale

Le LMA à été fondé en 1941 à partir des moyens du Centre d'Etudes de la Marine Nationale. Initialement « Centre de Recherches Industrielles et Maritimes », il est devenu par la suite « Centre de Recherches Physiques » en 1963, puis a adopté en 1973 son intitulé actuel « Laboratoire de Mécanique et d'Acoustique » pour s'adapter ainsi à l'évolution de ses axes de recherches.

Le LMA est une Unité Propre de Recherche (UPR) du département Sciences et Technologies de l'Information et de l'Ingénierie (ST2I) du CNRS. Il est associé aux universités de Provence (Aix-Marseille 1) et de la Méditerranée (Aix-Marseille 2) et a des liens étroits avec l'Ecole Centrale de Marseille (anciennement EGIM). L'effectif du laboratoire est d'environ 120 personnes (chercheurs, enseignants-chercheurs, ITA et doctorants).

Les installations spécialisées du LMA, notamment en acoustique, en font l'un des laboratoires universitaires français les mieux équipés dans son domaine.

Les activités de recherche

Les activités du LMA sont réparties en 4 axes :

- Matériaux
- Ondes
- Structures
- Sons

C'est au coeur de ce dernier axe que j'ai réalisé mon stage, plus précisément dans l'équipe « Modélisation, Synthèse et Contrôle des Signaux Sonores et Musicaux » qui sera présenté ci-dessous.

Collaborations et partenariats

Le LMA développe des collaborations étroites avec les universités et les écoles d'ingénieurs de la région Aix-Marseille. Il accueille des enseignants-chercheurs qui y effectuent leur recherche et des doctorants qui y préparent une thèse, ainsi le laboratoire s'investit fortement dans la formation par la recherche, notamment au sein de l'Ecole Doctorale « Physique, Modélisation et Sciences pour l'Ingénieur » et de l'ECM. Les membres du LMA participent activement à l'enseignement de la mécanique et de l'acoustique en France.

Partenaire de plusieurs contrats européens, développant de nombreuses coopérations à travers le monde, le laboratoire a également vu confier à plusieurs de ses membres des responsabilités importantes dans l'organisation de groupements de recherche ou de programmes internationaux de coopération scientifique (Groupements de Recherche français et européens, Actions Concertées Incitatives, Plan-Pluri Formation, Equipe de Recherche Technologique,...). Dans ce cadre, il accueille chaque année plusieurs visiteurs étrangers (de niveau post-doctoral ou seniors) et de nombreux scientifiques pour des séjours de courte et longue durée. Il développe des collaborations régulières avec les grands groupes industriels (dans le domaine des transports, de la sécurité nucléaire...) ainsi qu'avec les PME/PMI locales.

L'équipe d'accueil

L'équipe « Modélisation, Synthèse et Contrôle des Signaux Sonores et Musicaux » du LMA est composée de :

- un responsable : Richard Kronland-Martinet (directeur de recherche).
- cinq participants : Jean Kergomard (directeur de recherche), Philippe Guillemain, Solvi Ystad, Christophe Vergez (chargés de recherche), Thierry Voinier (ingénieur de recherche).
- 6 doctorants : Mathieu Barthet, Fabrice Silva, Charles Verron, Adrien Merer, Jean-François Sciabica, Thibault Necciari.

L'équipe s'appuie aujourd'hui sur deux opérations de recherche en forte interaction : *Modélisation, Synthèse et Contrôle des Signaux Sonores et Musicaux* et *Physique des Instruments de Musique*. La collaboration entre ces deux orientations s'articule autour de trois axes fondamentaux de recherche :

- la création des sons (analyse, instruments numériques et mécaniques, spatialisation, codage...).
- le contrôle des sons (synthèse, temps réel, geste, interprétation...).
- la perception et la cognition sonore (réalité virtuelle sonore, timbre, sémiotique de sons...).

et de deux projets ANR :

- projet CONSONNES (CONtrôle de SONs instrumentaux Naturels Et Synthétiques).
- projet senSons (vers le sens des sons).

Les actions menées dans le cadre de ces opérations visent à la construction d'outils numériques pour la synthèse et le traitement des sons et de la musique. Elles poursuivent les travaux sur l'analyse-synthèse menés au laboratoire depuis une quinzaine d'années, tout en les étendant aux processus de diffusion et de

spatialisation, ainsi qu'à la prise en compte des relations entre la nature des sons, la perception et la cognition. L'essentiel de ces travaux donne lieu à des réalisations fonctionnant en temps réel sous environnement Max/MSP.

Perspectives

Les travaux sur la synthèse temps réel d'instruments de musique seront poursuivis. L'installation au laboratoire d'une bouche artificielle devrait permettre, après plusieurs années de développement d'algorithmes de synthèse, c'est à dire le problème direct, de s'intéresser au problème inverse, consistant à déterminer les paramètres du modèle par l'analyse de sons produits dans des conditions connues et reproductibles. Ce travail s'inscrira naturellement dans le cadre du projet ANR CONSONNES. La synthèse des sons en vue d'applications à la réalité virtuelle sera également au centre des préoccupations de l'équipe. Il s'agira notamment d'étendre les possibilités de la plate-forme d'analyse-synthèse à la prise en compte de l'interaction entre excitateur et structure, et notamment du geste (gratter, frotter, souffler,...). L'ouverture vers les sciences cognitives et la prise en compte de la perception dans la définition des processus de génération sonore seront également poursuivies. L'étude du masquage temps-fréquence sera poursuivie dans le cadre de la thèse de T. Necciari en collaboration avec l'OR « Acoustique Perceptive ». La sémiotique sonore, qui est au coeur des préoccupations des chercheurs en acoustique audio et neurosciences devrait être un point d'attraction susceptible de permettre la structuration d'une véritable équipe pour l'essentiel décrit dans les objectifs du projet senSons.

Une présentation plus détaillée se trouve sur le site internet du Laboratoire de Mécanique et d'Acoustique du CNRS de Marseille (LMA) : `http://www.lma.cnrs-mrs.fr`

II La problématique

Contexte

L'explosion des bases de données audio nous pousse à élaborer des méthodes de navigation rapide et intuitive. Le signal temporel (suite d'échantillons) est la description la plus complète de l'information, mais écouter un à un chaque son pour le classifier n'est pas la manière la plus optimale et intelligente de procéder. On se base donc sur l'estimation de descripteurs qui permettent de réduire l'information contenue dans le signal et de proposer une méthode de classification en fonction des valeurs de ces descripteurs. En pratique, dans le cadre de la norme standardisée MPEG-7[1], une large palette de descripteurs (temporels, spectraux, spectro-temporels) a pu être définie. Cette norme se base sur le calcul de milliers de descripteurs, toutefois cela reste moins coûteux que de considérer le signal lui-même. Ainsi, au travers des calculs de descripteurs, le signal peut être classifié selon des taxonomies pré-définies telles que les genres musicaux (jazz,...), les sons d'environnements (explosion, impact, liquide,...), la parole (conversation...) etc. De manière plus générale, le calcul de descripteurs est très largement utilisé dans les domaines où une masse importante de

[1] Le *Moving Picturee Experts Groupe (MPEG)* est un groupe de travail du l'*International Standard Organization/International Electrotechnical Committee (ISO/IEC)* en charge de développer des normes de représentations codées de signal audio et vidéo, MPEG-7 est la norme qui réalise des outils de description du contenus multimédia pour la recherche et le classement de signaux audio et vidéo [KMS05]

données ont besoin d'être classifiées, en anglais on parle alors de *data retrieval* ou encore d' *information retrieval*.

Notre ambition est de proposer une méthode de classification basée sur un nombre réduit de descripteurs. En outre, on souhaite que cette méthode classifie les sons sur la base de critères perceptifs et non physiques. Le choix des descripteurs est donc déterminant pour la validité de la méthode, il en existe une infinité, mais nous choisirons ceux dont on connaît la pertinence d'un point de vue perceptif. Dans le cadre de ce stage, nous nous limiterons à une classe spécifique de sons d'environnements : les sons d'impacts, et à la classification des différents matériaux perçus.

En pratique, des modèles prédictifs pour chaque catégorie de matériau ont été construits. L'algorithme final de classification a été ensuite mis au point sur la base des prédictions données par chaque modèle. La calibration de ces modèles a été effectuée sur une banque de sons qui a d'abord été créée, puis évaluée par un test perceptif. L'étape de validation, pas encore effectuée lors de la rédaction de ce rapport, consistera à générer une nouvelle banque de sons évaluée de nouveau par un test perceptif, et de confronter les prédictions du modèle aux résultats observés par ce test.

Ce stage s'intègre directement dans un projet plus large de l'équipe que j'ai intégré portant sur la synthèse des sons d'impact. En effet, le LMA à développé un outil temps réel permettant de synthétiser des sons d'impacts, intitulé *Tapong*, selon un modèle décrit dans [AKM06]. *Tapong* nécessite en entrée une centaine de paramètres de synthèse. Une des questions qui s'est posé à l'équipe du LMA fut de savoir si l'on pouvait définir une stratégie de contrôle de ces paramètres qui soit à la fois simple, intuitive et robuste. En s'appuyant sur des études de psychoacoustique et de neuroacoustique [AKMVY06] on montre que pour déterminer à quelle classe appartient un son de matériau frappé (bois, verre ou métal), deux paramètres semblent prépondérants : l'amortissement (en fonction de la fréquence) et la rugosité. Sur la base de ces résultats, un contrôle de l'amortissement à été proposé [AKMVY06] et le contrôle de la rugosité est en cours d'évaluation, ces deux paramètres ont été également pris en compte en tant que descripteurs pertinents pour la construction de modèles prédictifs.

Intérêt du stage

Les prédicteurs perceptifs du signal sont utiles pour l'indexation par le contenu de bases de donnée audio ou multimédia, par exemple avec le E-commerce : plutôt que rechercher une musique par le biais du titre ou de l'auteur, nous pourrons utiliser des systèmes de recherche utilisant un « langage naturel » en indiquant que l'on souhaiterait télécharger un son qui « sonnerait comme tel autre » ou qui nous ferait « ressentir telle émotion » [Lem02]. Un autre intérêt est celui pour la synthèse sonore : si nous comprenons mieux quelles sont les caractéristiques perceptives propres aux catégories (d'un point de vue acoustique, catégoriser des sons avec un nombre limité de descripteurs qui soient pertinents au niveau perceptif peut nous donner des pistes de compréhension sur la façon dont nous percevons les signaux sonores), nous pourrons synthétiser des sons de manière plus intuitive en reproduisant ces caractéristiques. Enfin, en établissant des méthodes de classification basées sur l'analyse statistique de données, nous construisons une méthodologie généralisable à n'importe quel type de données sonores nécessitant une classification.

Chapitre 3

Phase expérimentale

I Constitution d'une base de données sonores

La première étape de l'étude fut d'enregistrer un maximum de sons d'impacts sur tout type de matériaux de notre environnement quotidien (verre, métal, bois, pierre, plastique,...). Les prises de sons ont été réalisées au format stéréophonique au moyen d'un enregistreur numérique Zoom H4 (fréquence d'échantillonnage 48 kHz). Nous avons enregistrés deux types d'impact pour chaque objet : dur (petit maillet) et mou (gros maillet). Nous avons également profité de cette campagne de mesures pour effectuer différents types d'excitation : brossage et grattage, qui n'ont pas été exploités lors de cette étude. L'ensemble du matériel utilisé est présenté sur la fig. 3.1. Au total, 325 sons ont été enregistrés et répertoriés dans une base de données globale, comportant une photo de l'objet excité montrant les différents points d'excitation, la nature du matériau excité et le type d'impacteur.

FIG. 3.1 – Matériel utilisé pour les prises de sons

II Test perceptif

II.1 Protocole

Dans le cadre de notre étude, il est fondamental que les catégories de sons avec lesquelles les modèles prédictifs seront calibrés soient déterminées sur des critères perceptifs et non pas physiques. Cela nécessite d'obtenir, pour chaque son, un jugement perceptif d'appartenance à telle ou telle classe de matériau. Ceci à été obtenu grâce à un test de catégorisation effectuée sur un groupe de sujets. La figure 3.2 présente l'interface qui a été développée pour ce test. Le sujet devait classer chaque son suivant le type de matériau perçu : *métal, bois, pierre, plastique, ou verre.* On a également laissé au sujet la possibilité de répondre *autre* et ainsi de proposer (ou non) un matériau différent. Chaque son pouvait être ré-ecouté autant de fois qu'il le désirait. Les sons étaient présentés dans un ordre aléatoire à travers les sujets afin d'éviter que le jugement d'un son n'influence le suivant de manière systématique. Un total de 525 sons ont été présentés aux auditeurs, 325 d'entre eux ont servi à la calibration du modèle et les 200 autres à la validation. La présentation des 325 premiers sons à été divisé en 5 sessions de 65 sons, soit environ un test de 10 minutes par session, la présentation des 200 sons suivants à été divisé en 4 sessions de 50 sons, soit environ un test de 8 minutes par session.

FIG. 3.2 – Interface développée pour le test perceptif

Concernant les sons qui ont servi à la calibration du modèle : 13 sujets ont répondu à la première session, 11 sujets aux deuxième, troisième et quatrième session et 9 sujets à la cinquième session. Au total ces 325 sons ont été écoutés par 20 sujets différents, certains d'entre eux ayant passé plusieurs sessions. Concernant les sons qui ont servi à la validation du modèle : 8 sujets ont répondu à chacune des 4 sessions. Au total ces 200 sons ont été écoutés par 13 sujets différents, certains d'entre eux ayant passé plusieurs sessions.

II.2 Résultats obtenus pour les sons destinés à la calibration

Pour chacun des 325 sons de la calibration, nous obtenons un pourcentage de classification dans chaque catégorie proposée :

	Observé					
	% mét	% boi	% pie	% pla	% ver	%aut
son 1	0	53.86	30.77	7.69	7.69	0
...	...	...	...	...	...	...
son 325	92.30	0	0	0	0	7.69

Nous avons ensuite estimé les taux de confusion en mettant en regard les matériaux réellement frappés et les pourcentages de réponses attribués dans chaque catégorie perçue. On met ainsi en évidence la façon dont chaque matériau a été perçu (voir fig 3.3).

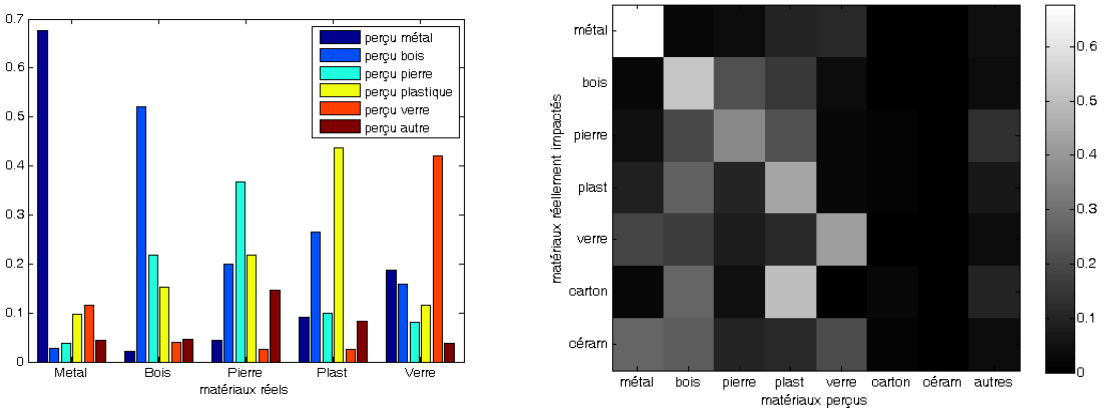


FIG. 3.3 – Sons destinés à la calibration. Taux de confusion (normalisé à 1) entre matériaux réellement impactés et matériaux perçus ; (à gauche) : représentation sous forme d'un diagramme en barres ; (à droite) : représentation sous forme de matrice, dans ce cas, les catégories « carton » et « céramique » ont été spécifiées car des enregistrements d'impacts sur des objets en carton et en céramique ont été soumis aux sujets.

Avec la matrice de confusion (bas de la fig 3.3), nous avons considéré deux catégories supplémentaires « carton » et « céramique » car des objets en carton et en céramique ont réellement été impactés. On observe alors que le carton à été souvent confondu avec le plastique (surtout les impacts sur des boîtes), mais aussi avec le bois. En ce qui concerne la céramique, elle a été principalement perçue comme du métal, du bois et du verre. Les sujets n'ont jamais pris l'initiative d'écrire en toute lettre « céramique » à l'écoute d'un son, en revanche ils l'ont fait pour le carton en référence à des sons de cartons, de céramique, de pierre et de plastique.

En analysant les données catégorie par catégorie, nous observons les choses suivantes :

1 le métal est le matériau le mieux reconnu (environ 70% des matériaux réellement métalliques ont été reconnus comme tels), il est quelquefois confondu avec le verre à hauteur de 10%.

2 le bois est assez bien reconnu (plus de 50%). On constate qu'il a majoritairement été confondu avec le plastique et la pierre.

3 le plastique est le troisième matériau le mieux reconnu (plus de 40%), il est le plus souvent confondu avec la pierre.

4 le verre est un matériau pour lequel les réponses obtenues sont variées. Le taux de confusion dépend fortement de la nature de l'objet : les impacts sur un verre (l'objet) ou un goulot de bouteille ont été unanimement catégorisés dans « verre », en revanche, les impacts sur une vitre en verre par exemple ont été moins bien reconnu et généralement confondu avec le bois et le plastique.

5 enfin la pierre est l'un des matériaux les plus difficiles à reconnaître (moins de 40%). Elle a été souvent confondu avec le bois et le plastique.

II.3 Résultats obtenus pour les sons destinés à la validation

De la même manière on met en évidence la façon dont chaque matériau à été perçu avec les 200 sons destinés à la validation par le diagramme de dispersion et la matrice de confusion présenté sur la figure 3.4. Ces 200 sons ont été divisée en 4 sessions de 50 sons, chaque session a été écouté par 8 sujets, au total 12 sujets différents ont été interrogé.

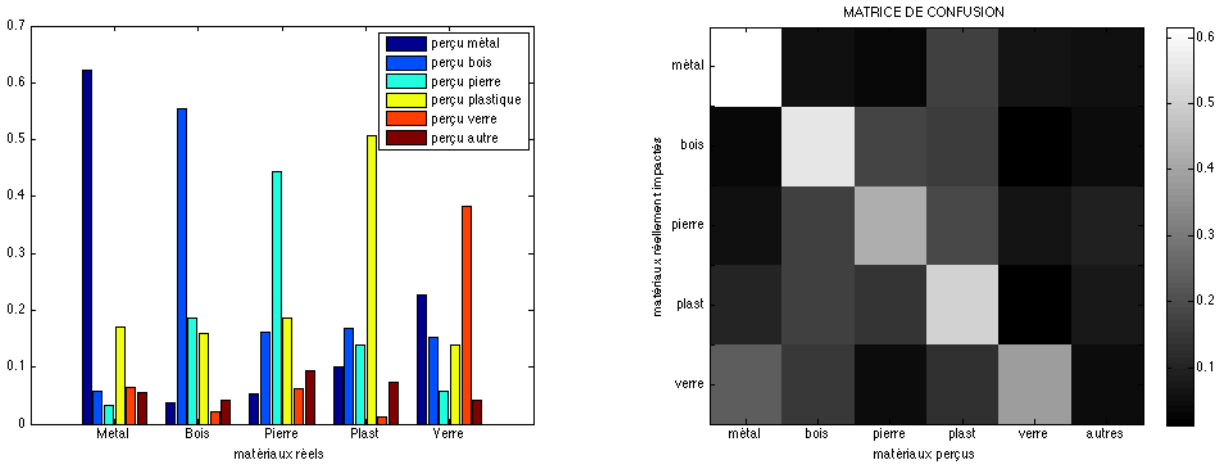


FIG. 3.4 – Sons destinés à la validation. Taux de confusion (normalisé à 1) entre matériaux réellement impactés et matériaux perçus ; (à gauche) : représentation sous forme d'un diagramme en barres ; (à droite) : représentation sous forme de matrice.

En analysant les données catégorie par catégorie, nous observons les choses suivantes :

1 le métal est de nouveau le matériau le mieux reconnu (à plus de 60%) et est à un peu moins de 20% confondu avec le plastique.

2 le bois est assez bien reconnu (plus de 50%). Comme précédemment, il a majoritairement été confondu avec le plastique et la pierre.

3 le plastique est le troisième matériau le mieux reconnu (plus de 50%), il est toujours confondu avec la pierre mais également avec le bois.

4 la pierre est cette fois-ci reconnu à plus de 40%, il a été souvent confondu avec le bois et le plastique.

5 enfin c'est le verre, qui pour cette banque de sons, est le matériau le plus ambigu avec une reconnaissance de moins de 40% et une forte confusion avec le métal, le bois et le plastique.

Notes sur les impressions des sujets

Voici 3 observations intéressantes que m'ont confié certains auditeurs après le test :

« Dans les cas ambigus de bruits d'impacts réverbérés, j'ai souvent eu tendance à répondre plastique car je lui associe plus facilement les formes complexe type boite creuse plutôt qu'au bois pour lequel je m'attend d'avantage à un bruit compact. »

« Lorsqu'il n'y a aucune réverbération j'ai souvent eu tendance à répondre pierre, bois ou plastique. Je ne perçois pas de son métal ou verre sur les bruits compacts. »

« La distinction métal / verre est assez facile lorsqu'il s'agit d'un tintement sur la tranche d'un verre ou d'un goulot de bouteille, ce bruit semble plus aigu qu'un impact de métal. »

Ces observations sont à prendre en compte, elles peuvent nous aider lors de la recherche de descripteurs, il semble d'ores et déjà qu'un descripteur d'ordre spectral jouerait un rôle important dans la distinction métal / verre, et que l'amortissement favoriserait la distinction [métal verre] / [bois plastique pierre].

III Conclusion sur la phase expérimentale

Dans cette phase expérimentale, nous avons mené une campagne de mesures afin de constituer deux banques de sons différentes, la première pour calibrer les modèles statistiques de classification et la deuxième pour les valider. Ainsi nous avons enregistré 325 sons d'impacts pour la calibration, puis 200 sons supplémentaires pour la validation. Nous avons ensuite demandé à plusieurs auditeurs de classer chacun des 525 sons selon le matériau perçu. Sur la base des résultats de ce test perceptif, nous allons pouvoir construire les modèles qui, pour un son d'impact donné, nous prédiront un pourcentage d'appartenance au métal, au bois, à la pierre, au plastique et au verre, ceci en accord avec notre perception et non avec la physique de l'objet.

Chapitre 4

Caractérisation acoustique des sons

L'analyse acoustique consiste à extraire les caractéristiques fondamentales des signaux. Afin de décrire, d'un point de vue acoustique, tous les sons soumis au test perceptif, nous avons considéré un ensemble de descripteurs connus pour leur pertinence en terme de perception du timbre et du matériau.

La définition du timbre énoncée par l'association américaine de normalisation (American National Standards Institute, ANSI, 1960) reste encore très floue : « *le timbre est la qualité perceptive utilisée par l'auditeur pour estimer la différence entre deux stimuli présentés dans les mêmes conditions avec la même sonie, la même hauteur et la même durée* ». Cette définition est accompagnée d'une note : « *le timbre dépend premièrement du spectre fréquentiel du stimulus, mais également de la forme d'onde, de la pression acoustique, de la disposition des fréquences à l'intérieur du spectre et des caractéristiques temporelles du stimulus* ». Ici, le timbre est défini par ce qu'il n'est pas : ni dynamique, ni hauteur, ni durée. Malgré cette non-définition, la notion de timbre est communément rentrée dans notre langage courant et s'applique à tous les types de sons : les sons musicaux (timbre d'un instrument), les sons de parole (timbre de la voix) et les sons de l'environnement (lié aux caractéristiques de la source physique).

Nous nous intéresserons donc aux descripteurs spécifiques du timbre. La note de la définition de l'ASA souligne que plusieurs paramètres physiques sont susceptibles de caractériser le timbre, on dit alors que le timbre est un attribut perceptif multidimensionnel. Beaucoup d'études psychoacoustiques ont pour objectif de déterminer les paramètres acoustiques ou physiques sur lesquels est fondé le calcul du timbre. Pour rendre compte de l'aspect multidimensionnel du timbre, on utilise classiquement la notion d'*espace de timbre*. Les axes d'un tel espace sont les différentes dimensions du timbre, que l'on cherche à relier aux caractéristiques du stimulus [Cac04]. Ainsi, Pols et coll. (1969) a créé un espace permettant d'isoler des sons de voyelles, cette technique à été appliquée aux sons d'instruments de musique par [Gre77], plus récemment de nouveaux espaces furent obtenus à partir de stimuli de synthèse [McA99]. Ces études ont mis en avant l'importance perceptive de descripteurs comme le logarithme du temps d'attaque, le centre de gravité spectral et le flux spectral [Mar04].

Nous nous intéresserons également à un descripteur spécifique à la perception du matériau : l'amortissement. Une large littérature à montré que l'amortissement est un indice perceptif fondamental [WR88], [GM06], [KPK00], [CD97], [LO97]. En effet, d'un point de vue physique, un son percussif résulte généralement d'un impact sur un objet vibrant dont le comportement vibratoire peut être modélisé par un

système mécanique masse-ressort-amortissement. La solution de l'équation différentielle correspondante peut donc s'exprimer sous la forme d'une somme de modes propres :

$$s(t) = \sum_{k=1}^{K} s_k(t) \text{ avec } s_k(t) = A_k e^{i(2\pi f_k t + \phi_k)} e^{-\alpha_k t} \tag{4.1}$$

On a donc une somme de K sinusoïdes exponentiellement amorties (modes propres) $s_k(t)$ dont les paramétres A_k, f_k, α_k=$\alpha(f_k)$, ϕ_k représentent respectivement l'amplitude, la fréquence propre, le coefficient d'amortissement et la phase à l'origine. La représentation fréquentielle de $s_k(t)$ est alors donnée par :

$$\hat{s}_k(f) = \frac{A_k}{\alpha(f_k) + 2i\pi(f - f_k)} \tag{4.2}$$

Ainsi la partie décroissante d'un son d'impact suit une loi exponentielle. De plus, d'après l'équation 4.1, le α_k est différent suivant les composantes fréquentielles, on parle alors de loi d'amortissement, c'est cette loi qui peut caractériser le matériau [CD97], [CL01].

I Les descripteurs fréquentiels

Le premier type de critères acoustiques pour décrire le timbre est d'ordre spectral. En effet, lorsque deux signaux ont la même hauteur tonale et la même sonie, un des paramètres acoustiques que l'on peut faire varier est l'enveloppe spectrale. Il parait donc intéressant d'étudier ses caractéristiques [Meu].

I.1 Le Centre de Gravité Spectral : *CGS*

Le centre de gravité spectral est une valeur en Hz qui correspond au centre de gravité du spectre du signal (voir fig 4.1). Nous nous sommes basés sur la définition de [Bea82] :

$$CGS = \frac{F_e}{N} \frac{\sum_{k=1}^{K} k a_k}{b_0 + \sum_{k=1}^{K} a_k} \tag{4.3}$$

où les a_k représentent les amplitudes linéaires (module de la transformée de Fourier) des fréquences k, $\frac{F_e}{N}$ est le rapport entre la fréquence d'échantillonnage et le nombre de points de calcul de la transformée de Fourier. Cette définition inclut également un seuil b_0 forçant le *CGS* à décroître pour les amplitudes très faible généralement dominées par du bruit.

D'un point de vue perceptif, le *CGS* est communément associé à la *brillance* du son (Grey and Gordon, 1978) : plus le son est brillant, plus son *CGS* est élevé [KMS05].

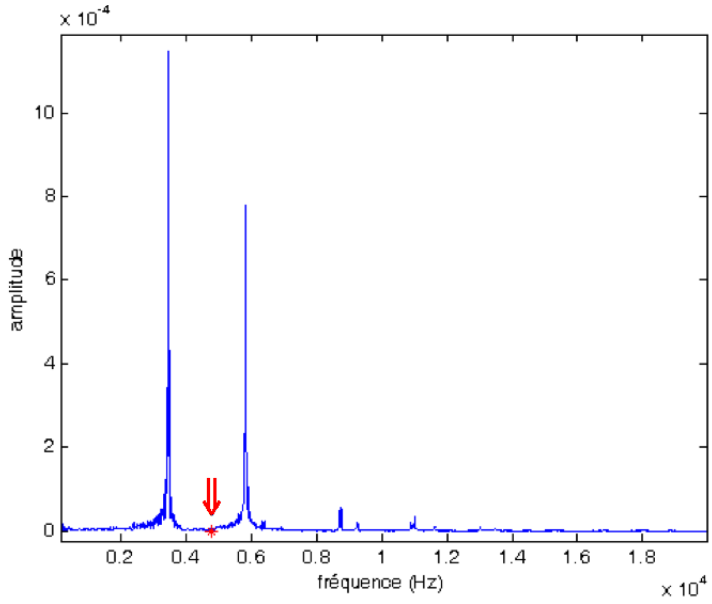


FIG. 4.1 – CGS d'un son d'impact de verre

I.2 Le point de roll-off : *ROF*

Le point de roll-off (traduction du terme anglais *spectral roll-off point* proposé par [RDR+08]) est la fréquence au dessous de laquelle 95% (selon [Pee04]) ou 85% (selon [KMS05]) de l'énergie spectrale est contenue (voir fig 4.2). Considéré comme un indice de répartition du spectre, le point de roll-off est plus grand pour les signaux ayant un spectre important au niveau des hautes fréquences [RDR+08]. Généralement il permet de distinguer les signaux bruités des signaux harmoniques [Pee04]. Nous utiliserons la définition de [KMS05] :

$$\sum_{0}^{ROF} a^2(k) = 0.85 \sum_{0}^{\frac{Fe}{2}} a^2(k) \tag{4.4}$$

où ROF et F_e sont respectivement le point de roll-off et la fréquence d'échantillonnage.

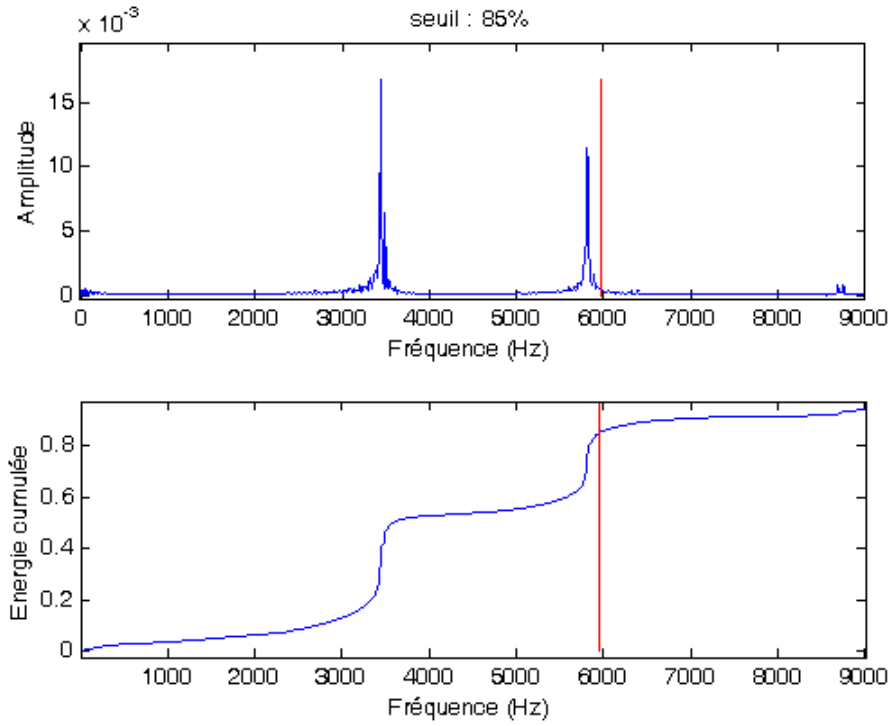


FIG. 4.2 – Amplitude (en haut) et énergie cumulée (en bas) du spectre d'un impact de verre. Le *ROF* avec un seuil de 85% est représenté sur les deux graphes par une ligne rouge verticale.

L'intérêt perceptif des descripteurs décrits en I.3, I.4, I.5 et I.6 est difficile à déterminer, néanmoins ils nous informent sur les caractéristiques de l'enveloppe spectrale.

I.3 L'étalement spectral : *SSP*

L'étalement spectral (en anglais *spectral spread*) est une largeur de bande en Hz, fonction du *CGS*. Il s'apparente à l'écart type d'une distribution statistique :

$$SSP = \sqrt{\frac{\sum_k (f_k - CGS)^2 \times a_k}{\sum_k f_k}} \tag{4.5}$$

où f_k, a_k, et CGS représentent respectivement la fréquence et l'amplitude de la k-ième composante et le centre de gravité spectral.

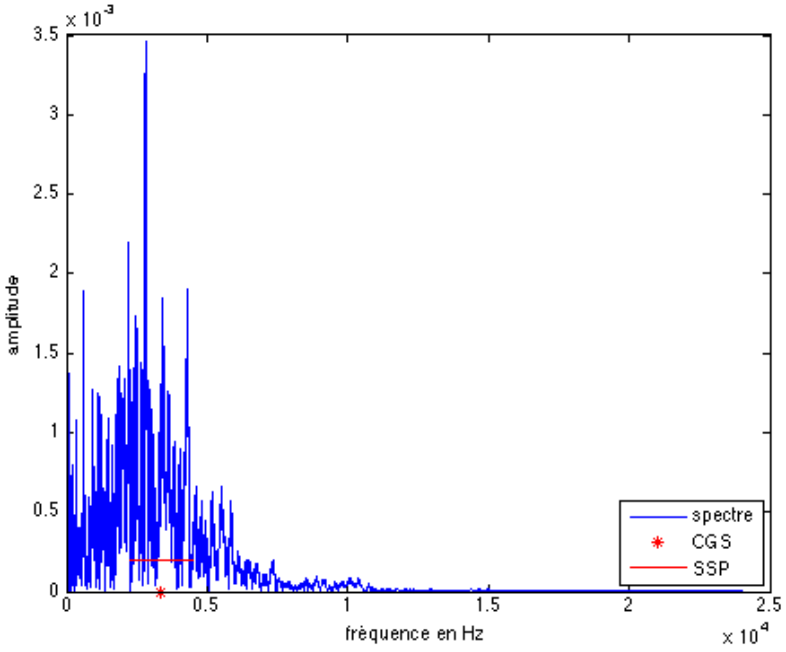


FIG. 4.3 – *SSP* d'un son d'impact de verre

I.4 Le coefficient de dissymétrie : *SKN*

En anglais *skewness* ; il s'agit d'une mesure de l'assymétrie de la distribution spectrale autour de sa valeur moyenne. Un coefficient positif indique une queue de distribution étalée vers la droite. Un coefficient négatif indique une queue de distribution étalée vers la gauche. Dans le cas d'une distribution normale, le coefficient est nul. Le coefficient de dissymétrie d'une distribution spectrale x est défini par :

$$SKN = \frac{E(x-\mu)^3}{\sigma^3} \tag{4.6}$$

où σ est la variance de la distribution spectrale x, μ sa valeur moyenne et $E(t)$ est l'espérance mathématique de la quantité t.

I.5 Le kurtosis : *KRT*

Il s'agit d'une mesure de l'aplatissement (ou a contrario de la pointicité) de la distribution spectrale. Le kurtosis d'une distribution suivant la loi normale est de 3, inférieur à 3 pour une distribution plus aplatie et supérieur à 3 pour une distribution plus pointue. Le kurtosis d'une distribution spectrale x est défini par :

$$KRT = \frac{E(x-\mu)^4}{\sigma^4} \tag{4.7}$$

où μ, σ et $E(t)$ représentent les mêmes quantités que dans l'équation 4.6.

I.6 La décroissance spectrale : *SDC*

Ce descripteur permet de décrire la décroissance de l'amplitude spectrale, d'après [Pee04] il serait fortement corrélé à la perception humaine. Le *SDC* est défini par :

$$SDC = \frac{1}{\sum_{k=2}^{K} a(k)} \times \sum_{k=2}^{K} \frac{a(k) - a(1)}{k-1} \tag{4.8}$$

où a(k) est l'amplitude des composantes fréquentielles k allant de 1 à K.

I.7 Le taux de croisement du spectre : *SCR*

Sur le même principe que le taux de passage à zero (décrit plus bas au paragraphe II.6 page 31), ce descripteur consiste à compter le nombre de fois où le spectre dépasse un seuil fixé à 50% de l'amplitude fréquentielle maximale et à normaliser cette valeur par la fréquence de Nyquist. L'idée première à l'origine de ce descripteur (pour lequel nous n'avons pas de références bibliographiques à notre connaissance) est de vouloir distinguer les sons spectralement riches des sons plus purs.

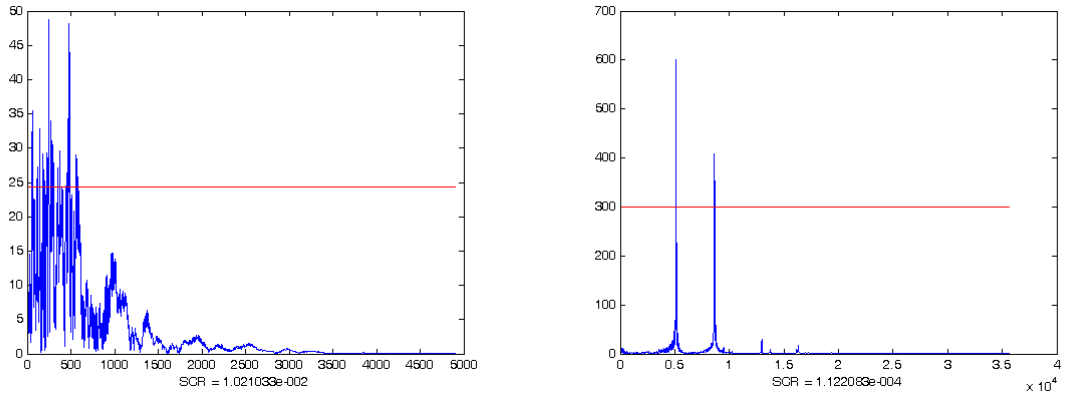


FIG. 4.4 – *SCR* d'un son d'impact de bois (à gauche) et de verre (à droite).

II Les descripteurs temporels

On met en évidence le facteur temporel lorsqu'on reproduit le son à l'envers. Prenons un son de cymbale, il semble évident qu'à l'écoute de ce son on lui associe un geste percussif et un matériau métallique. Maintenant écoutons le même son à l'envers, le matériau métal est éventuellement encore reconnu, mais il est fort probable que l'on ne reconnaisse pas la nature percussive du son. Pourtant, dans les deux cas, le spectre est identique, les seuls changements interviennent dans la structure temporelle du son, d'où l'idée d'étudier un certain nombre de descripteurs temporels.

II.1 Le logarithme du temps d'attaque : *LAT*

Le temps d'attaque est défini comme la durée nécessaire au signal pour croître de 2% à 100% de son pic maximal en amplitude [KMS05]. Nous en calculerons son logarithme car certaines études, notamment [MWD$^+$95], montrent que le logarithme du temps d'attaque est plus corrélé à la perception que le temps d'attaque lui même. Nous articulerons l'algorithme de calcul du temps d'attaque de la manière suivante :

estimation de l'enveloppe du signal sur la partie croissante

Classiquement l'enveloppe d'un signal est calculé en appliquant un filtre (type Butterworth) au module du signal analytique. Mais dans le cas des sons percussifs, les temps d'attaque sont très courts (généralement de l'ordre d'une dizaine d'échantillons), un filtre type Butterworth est inefficace pour estimer correctement l'enveloppe du signal dans cette zone. Nous nous sommes donc inspirés de la technique du *crête-mètre* (en anglais *peakmeter*) qui consiste à suivre de manière constante les amplitudes des crêtes du signal redressé double alternance. L'enveloppe en crête-mètre est calculée sur la partie croissante définie par une zone de 20 ms en amont de l'amplitude maximale, (on suppose qu'aucun temps d'attaque de son percussif ne dépassera 20 ms).

approximation de l'enveloppe par une droite

On approxime l'enveloppe ainsi crée par une droite grâce à une méthode de régression linéaire classique.

repérage des temps T_{start} et T_{stop}

Nous avons défini T_{start} comme étant l'´échantillon pour lequel le segment fraîchement estimé atteint 50% de sa valeur maximale et T_{stop} l'échantillon pour lequel il atteint 100% de sa valeur maximale. Nous définirons le temps d'attaque comme le double de la durée séparant les deux points.

logarithme du temps d'attaque

Ainsi le logarithme du temps d'attaque est défini comme :

$$LAT = log_{10}(2 * (T_{stop} - T_{start})) \tag{4.9}$$

D'un point de vue perceptif, le *LAT* est un descripteur pertinent pour la distinction de différentes classes d'instruments. Il permet par exemple de différentier les percussions (temps d'attaques courts) des vents (temps d'attaques plus longs).

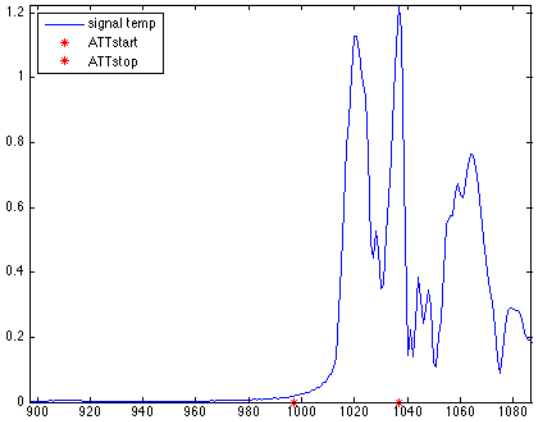


FIG. 4.5 – Etapes de l'estimation du temps attaque sur un impact de métal.

FIG. 4.6 – Estimation du temps d'attaque d'un impact de bois, le *LAT* est le logarithme du temps qui s'écoule entre les deux points rouges.

II.2 Le centre de gravité temporel : *CGT*

Il s'agit de la valeur (en seconde) qui sépare l'enveloppe temporelle en deux parties d'égales énergies. D'après [KMS05], le *CGT* est défini par :

$$CGT = \frac{\sum_{k=0}^{N-1} kEnv(k)}{\sum_{k=0}^{K-1} Env(k)} \tag{4.10}$$

où *Env(k)* est l'enveloppe du signal temporel défini par :

$$Env(k) = \sqrt{\frac{1}{N_w} \sum_{n=0}^{N_w-1} s^2(k.N_{hop}+n)} \quad \text{avec } (0 \leq k \leq K-1), \tag{4.11}$$

où K est le nombre total de fenêtres, N_{hop} est le nombre d'échantillons de chevauchement (*overlap*) et N_w est le nombre d'échantillons de la fenêtre. Nous utilisons des fenêtres de 0.4 ms (meilleure valeur obtenue de manière empirique).

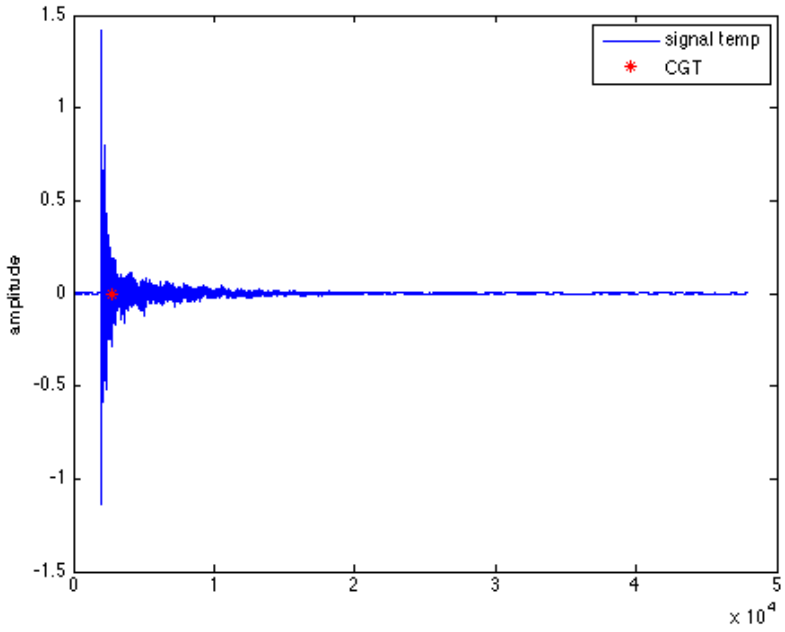


FIG. 4.7 – *CGT* d'un son d'impact de verre

II.3 La valeur efficace : *RMS*

La valeur efficace (en anglais *root mean square*) est calculée sur le signal temporel *s(t)* sur *N* échantillons :

$$RMS = \sqrt{\frac{\sum_{k=1}^{N} s(k)^2}{N}} \tag{4.12}$$

Cette grandeur est directement corrélée à l'intensité du signal. Pour des signaux impulsionnels, le *RMS* dépend du nombre d'échantillon sur lequel il a été calculé. Ici *N* correspond au nombre d'échantillons entre l'instant T'_{start} et T'_{stop} qui sont respectivement les instants où l'enveloppe *Env(k)* (définie par l'équation 4.11) du signal atteint pour la première et la dernière fois les 2% de sa valeur maximale.

II.4 La durée perceptive : *TED*

La durée perceptive (en anglais : *Time Effective Duration*) est la durée pendant laquelle le signal est perceptivement significatif. On l'approxime par la durée pendant laquelle l'enveloppe spectrale (équation 4.11) est supérieure à un seuil, généralement défini à 40% de sa valeur maximale [Pee04].

Ce descripteur est couramment utilisé pour distinguer les sons percussifs des sons entretenus [Pee04].

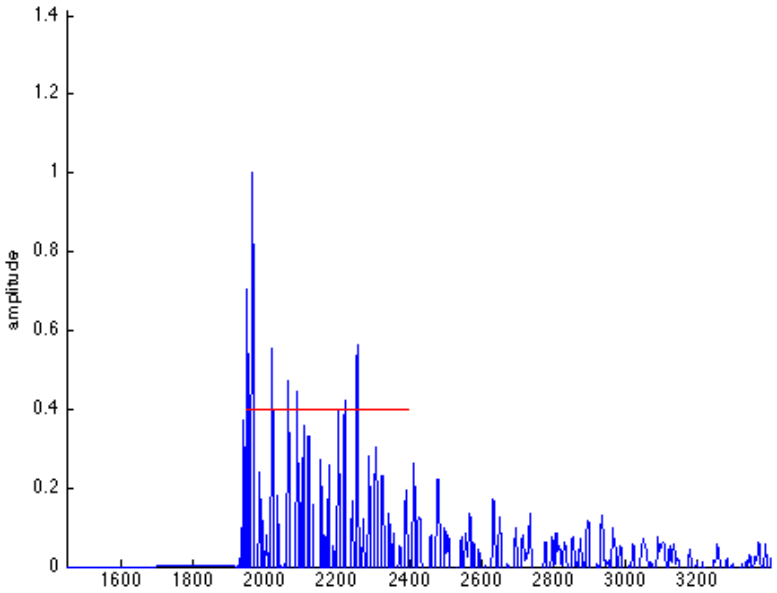


FIG. 4.8 – *TED* d'un son d'impact de verre (calculé sur le module du signal) : durée pendant laquelle le signal est au dessus de 40% de sa valeur maximale.

II.5 L'amortissement : *AMO*

Ce descripteur est spécifique à la classe des sons d'impact dans la mesure où il ne peut être calculé que sur des sons qui présentent une décroissance temporelle exponentielle. Nous nous limiterons à estimer un amortissement global α sur l'enveloppe du signal temporel (et non sur chacun des α_k définis dans l'équation 4.1). En prenant le logarithme de cette décroissance on s'attend donc à obtenir une droite de pente α qui, normalisée par le *CGS*, deviendra notre valeur *AMO*.

Calcul de l'enveloppe du signal temporel : l'amortissement est donc estimé en approximant par une droite le logarithme de l'enveloppe du signal. Celle-ci est obtenu en filtrant le module du signal analytique par un filtre de Butterworth. (voir fig 4.9).

Calcul de l'amortissement : l'approximation par une droite du logarithme de l'enveloppe de la partie décroissante du signal est calculé avec la fonction matlab *polyfit*. Le coefficient directeur de cette droite correspond à l'amortissement global α (voir fig 4.9).

Normalisation de l'amortissement : d'un point de vue physique, l'amortissement est fonction de la fréquence, son taux de décroissance dépend du contenu spectral du son (équation 4.1). Dans notre cas, les sons que nous étudions présentent une grande disparité dans le contenu spectral. Ainsi, pour pouvoir comparer ces valeurs, nous les normalisons par le *CGS* (qui rend compte de la localisation spectrale de l'énergie) du son étudié.

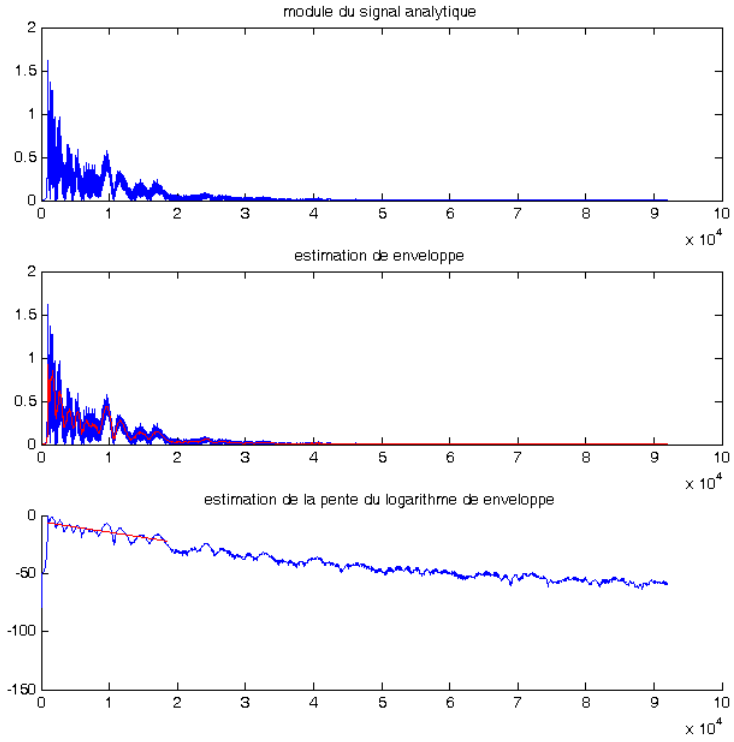


FIG. 4.9 – (en haut) : module du signal analytique calculé à partir de la transformée de Hilbert (décroissance exponentielle), (au milieu) : enveloppe du signal analytique estimée avec un filtre de butterworth, (en bas) : logarithme de l'enveloppe (comportement linéaire) et son approximation par une droite avec la fonction *polyfit* de Matlab.

On définit ainsi un amortissement normalisé *AMO* par :

$$AMO = \frac{\alpha}{CGS} \tag{4.13}$$

II.6 Taux de passage par zéro : *ZCR*

Le taux de passage par zero (en anglais *zero-crossing rate*) est un paramètre déduit du nombre de fois où le signal change de signe, cette mesure est corrélée au centroïde spectral [Ver03].

$$ZCR = \sum_{n=m}^{M} |sign(s(n)) - sign(s(n-1))| \tag{4.14}$$

Le ZCR est calculé sur une portion de signal de 5 ms (valeur permettant un bon compris entre rapidité de calcul et nombre d'échantillons minimum nécessaire pour le réaliser) entre l'échantillon m et M où m est l'échantillon correspondant à l'amplitude maximale du signal. *sign(x)* est défini comme suit :

$$sign(x) = \begin{cases} 1 \text{ si x} > 0 \\ 0 \text{ si x} = 0 \\ \text{-1 si x} < 0 \end{cases} \tag{4.15}$$

Le *ZCR* est couramment utilisé pour distinguer les signaux vocaux des signaux musicaux (Sheirer and Slaney, 1997) ou pour la classification des musiques par genre (Tzanetakis and Cook, 2002 ; Burred and Lerch, 2004). Les sons périodiques tendent à avoir une faible valeur de *ZCR*, alors que les sons bruités approchent des valeurs beaucoup plus importantes.

III Les descripteurs spectro-temporels

En complément à l'études des caractéristiques fréquentielles et temporelles du signal, il est classique de s'intéresser à des descripteurs d'ordre spectro-temporels.

III.1 Le flux spectral : *FSP*

Il s'agit d'une mesure de la variation du spectre au cours du temps calculée selon la définition [Cac04] :

$$FSP = \frac{1}{M} \sum_{p=2}^{M} |r_{p,p-1}| \text{ avec } M = \frac{T}{\Delta t} \tag{4.16}$$

où T est la durée totale du son, Δt = 16 ms (idéal selon [Cac04]) et $r_{p,p-1}$ est le coefficient de corrélation de Pearson entre les spectres instantanés aux temps p et p-*1*.

Pour deux fenêtres d'analyses considérées, le coefficient de corrélation vaut 1 si les spectres instantanés sont identiques, et tend vers 0 si les spectres instantanés présentent de fortes dissemblances (voir 4.10).

Le *FSP* est un descripteur qui a été utilisé pour la séparation voix/musique (Scheirer and Slaney, 1997 ; Burredand Lerch, 2004). On sait aussi que les sons environnementaux ont des *FSP* plus élevés que les sons musicaux et qu'il permet de distinguer les instruments d'orchestre [BDKMY08].

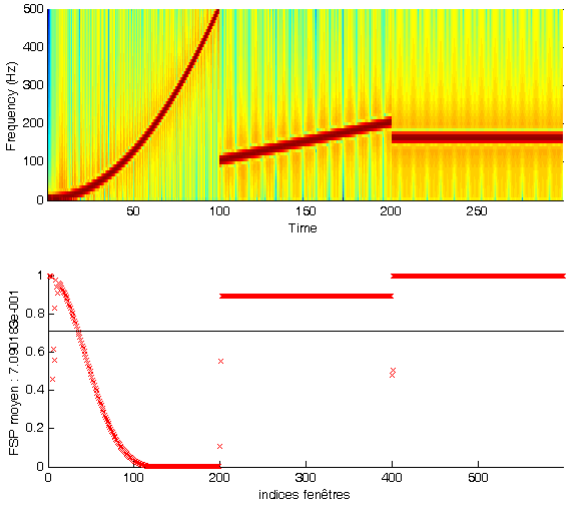


FIG. 4.10 – (en haut) : représentation temps fréquence d'un signal composé d'un chirp quadratique (0-100), d'un chirp linéaire (100-200) et d'un signal stationnaire sinusoïdal (200-300), (en bas) : les coefficients de corrélation de Pearson correspondant dont la somme normalisée par $\frac{T}{\Delta t}$ (equation 4.16) nous donnera le *FSP*.

III.2 La rugosité : *RUG*

La rugosité est un descripteur dont on ne connaît que peu de choses et les modèles pour la calculer sont variés. Afin de mieux comprendre ce que nous entendons par rugosité, ce descripteur fera l'objet d'une section plus importante.

Définition de la rugosité

La rugosité est une grandeur subjective de la perception. Donner une définition simple de la rugosité n'est pas évident car ce percept n'est pas aussi familier que les autres grandeurs (hauteur tonale, sonie...) et peut être approchée de plusieurs façons. Dans la musique européenne le concept de rugosité a longtemps été lié à celui de la *dissonance*. Dans d'autres traditions au contraire, la rugosité est un effet recherché, on peut citer le *tamboura*, instrument essentiel de la musique traditionnelle indienne qui sert de bourdon harmonique, ou encore les chants de *ganga* traditionnels du sud de la Croatie et de la Bosnie-Herzégovine, qui reposent en grande partie sur des effets de rugosité.

Bien que formalisé assez tard, la rugosité est un phénomène connu depuis longtemps. La première définition de la rugosité fut donnée par Helmholtz (1863), il utilisa ce terme pour caractériser la texture d'un son, en terme d'*impureté* ou de *sensation désagréable* [Lem00]. En 1877, il affine sa définition en étudiant l'effet produit par deux sons purs présentés simultanément. En faisant varier l'écart de fréquence entre les deux sons purs il distingue trois zones. Si les sons sont de fréquences très proches, nous perçevons un son pur à la fréquence moyenne des sons originaux dont l'amplitude fluctue lentement. On parle alors de phénomène de battements. En augmentant l'écart fréquentiel, les battements s'accélèrent et au delà d'une dizaine de Hz leur perception s'estompe et on observe alors un phénomène perceptif différents des battements, ce que Helmholtz définit comme étant la rugosité. Si l'écart fréquentiel entre les sons est encore

augmenté, la perception de rugosité diminue jusqu'à disparaître, il apparaît alors la perception de deux sons purs à des fréquences différentes, aux amplitudes constantes (voir fig 4.13).

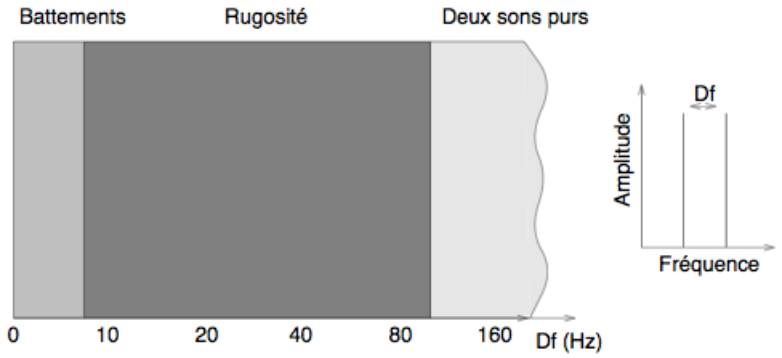


FIG. 4.11 – Les qualités perceptives de deux sons purs en fonction de leur écart fréquentiel. Les frontiéres entre les régions ne sont pas abruptes, notamment la frontière supérieure entre rugoisté et deux sons distincts dépend du registre, d'aprés Helmholtz (1877).

Le calcul de la rugosité peut s'effectuer selon plusieurs modèles, celui que nous utilisons est le modèle de Leman explicité dans [Lem00]. L'idée de Leman est d'assimiler la rugosité à l'énergie de la synchronisation neuronale avec la fréquence de battement.

Afin d'expliquer le procedé calculatoire de la rugosité il faut auparavant définir deux notions, celle du *Auditory Peripheral Module* et celle du *Syncrhonisation Index Model* définies dans [LLT].

La modélisation de l'oreille externe et moyenne (*Auditory Peripheral Module, APM*)

La première étape du modèle de Leman consiste à transformer l'information sonore (pression au cours du temps) en information nerveuse, c'est à dire en quantité de décharges au cours du temps, Leman modélise ainsi le fonctionnement de l'oreille. Plus exactement, il applique un filtre passe-bas pour modéliser l'influence de l'oreille externe et moyenne, puis applique un banc de filtres passe-bande pour modéliser le filtrage mécanique de la cochlée par bandes critiques.

La décomposition en sous-bandes de fréquences est exprimée comme suit :

$$APM : s(t) \rightarrow \tilde{\mathrm{d}} =< d_c(t) >_{c=1...C} \tag{4.17}$$

où $d_c(t)$ est la probabilité de décharge neuronale par intervalle de 0.4 ms de la fibre nerveuse (ou canal auditif) c. Un canal auditif est considéré comme un filtre de bande passante égale à la bande critique correspondante. C filtres sont alors considérés.

Leman applique ensuite dans chaque sous-bande un algorithme modélisant les cellules ciliées (modèle issu du travail de Van Immerseel). Ce modèle est composé d'un redresseur et d'une amplification dont le gain varie en fonction de l'entrée, filtrée par un passe-bas.

En sortie de ce traitement, on obtient pour chaque sous-bande, un vecteur donnant au cours du temps, le nombre de décharges neuronales par milliseconde (*spikes/ms*).

L'indice de synchronisation (*Synchronisation Index Module, SIM*)

Une analyse fréquentielle du taux de décharges neuronales dans les chaînes auditives donne une information sur la synchronisation des neurones à une fréquence particulière. La transformée de Fourier rapide de la synchronisation neuronale du canal c est :

$$D(t,f,c) = \int_{t'=-\infty}^{+\infty} e(t,c)(t)w(t'-t)e^{-j2\pi ft'}dt' \tag{4.18}$$

où *e(t,c)* est la synchronisation neuronale du canal c et *w(t'-t)* est une fenêtre de hamming. Cette formule calcule le taux de synchronisation pour chaque fréquence. L'amplitude spectrale est alors définie par $|$ *D(t,f,c)* $|$.

L'indice de synchronisation sera calculé en divisant cette dernière valeur par le taux à fréquence nulle :

$$I(t,f,c) = |\frac{D(t,f,c)}{D(t,0,c)}| \tag{4.19}$$

Différentes approches basées sur le fait que l'information de synchronisation est représenté selon trois axes : le temps, la fréquence et le canal auditif, sont maintenant envisageables. Leman lie la rugosité à l'énergie de la synchronisation neuronale. Nous calculerons d'abord l'énergie de la synchronisation neuronale individuellement pour chaque canal auditif, puis nous considérerons que l'énergie totale est la somme de ces énergies.

Les définitions suivantes expliquent deux concepts utilisés :

- *l'énergie spectrale* de la synchronisation neuronale dans le canal c est :

$$D(t,f,c) = |I(t,f,c)|^2 \tag{4.20}$$

  Chaque composante fréquentielle f fourni le taux de synchronisation en terme d'énergie pour cette fréquence particulière. Chaque composante est appelée *index de synchronisation*.
- l'énergie spectrale de synchronisation neuronale pour tous les canaux c est :

$$D(t,f) = \sum_{c=1}^{C} D(t,f,c) \tag{4.21}$$

  où C est le nombre total de canaux auditifs.

- l'*énergie* de la synchronisation neuronale dans le canal c est défini comme la somme des indices de synchronisation neuronale.

$$E_D(t,c) = \int D(t,f,c)df \tag{4.22}$$

De ces différentes définitions s'en suit que l'énergie totale du taux de synchronisation pour tous les canaux peuvent être exprimés de différentes façons :

$$E_D(t) = \int D(t,f)df = \int \sum_{c=1}^{C} D(t,f,c)df = \sum_{c=1}^{C} E_D(t,c) = \sum_{c=1}^{C} \int D(t,f,c)df \tag{4.23}$$

L'équation 4.23 montre qu'il est possible de visualiser deux aspects de l'énergie : la visualisation de l'énergie spectrale $D(t,f)$ donne l'énergie de la synchronisation neuronale le long d'un axe fréquentiel. La visualisation de l'énergie E_D(t,c) donne l'énergie de la synchronisation neuronale le long d'une échelle de bande critique. Ces deux représentations seront utilisées pour la représentation de la rugosité.

Modélisation de la rugosité

L'approche SIM calcule la rugosité en terme d'*énergie* de synchronisation neuronale avec les fréquences de battements. Cette *énergie* fait réference à une quantité provenant de l'amplitude du spectre. Comme la fréquence de battement est plutôt en zone basse fréquence, la partie spectrale qui nous intéresse est définie par :

$$B(t,f,c) = F(f,c)D(t,f,c) \tag{4.24}$$

où F(f,c) est un filtre dont l'amplitude spectrale dépend du canal c. Afin d'être en accord avec les données psychoacoustiques connues, les filtres sont plus étroits pour les canaux auditifs dont la fréquence centrale est inférieure à 800 Hz, et plus élargis pour les hautes fréquences centrales (voir la figure 4.12). Cependant, on peut dire qu'en général *B(t,f,c)* représente le spectre de la synchronisation neuronale pour la fréquence de battement dans le canal c. *L'index de synchronisation* de la fréquence de battement est alors défini comme l'amplitude spectrale normalisée :

$$I(t,f,c) = |\frac{B(t,f,c)}{D(t,0,c)}| \tag{4.25}$$

La rugosité est calculée individuellement dans chaque canal et la rugosité totale est la somme des rugosités de chaque canal. Considérant cette approche, il devient possible de visualiser la contribution de l'énergie de synchronisation le long de l'axe des canaux auditifs, aussi bien que le long de l'axe des fréquences de battements. Le terme d'*énergie spectrale* de la synchronisation neuronale avec des fréquences de battements est, pour le canal auditif c, défini comme :

$$B(t,f,c) = I(t,f,c)^{\gamma} \tag{4.26}$$

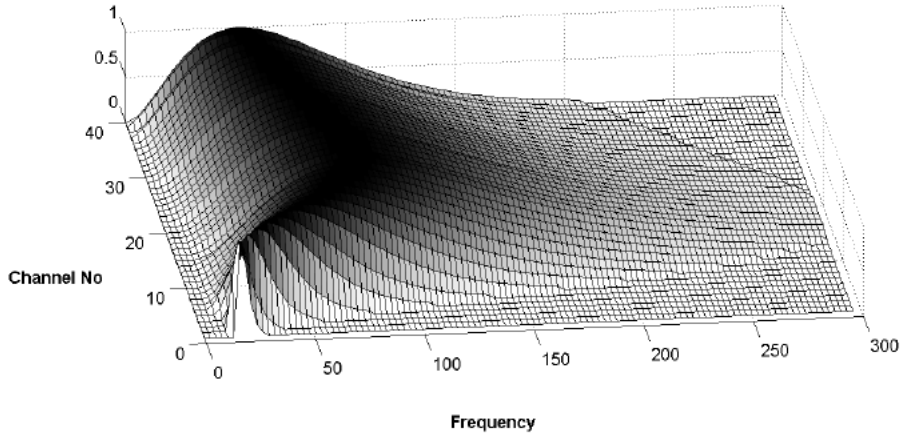


FIG. 4.12 – Les filtres deviennent plus étroits en basse fréquence. Figure extraite de [LLT]

où γ est un paramétre compris entre 1 et 2. En analogie avec l'équation 4.23, nous obtenons les différentes relations pour le calcul de la rugosité :

$$R_B(t) = \int \sum_{c=1}^{C} B(t,f,c)df = \sum_{c=1}^{C} \int B(t,f,c)df = \sum_{c=1}^{C} E_B(t,c) \tag{4.27}$$

Ces expressions permettent la visualisation de la rugosité le long de l'axe des canaux auditifs et le long de l'axe des fréquences de battements.

En pratique

La toolbox Matlab *IPEMToolbox1.01* de Marc Leman nous permet d'obtenir la rugosité d'un son en fonction du temps (voir fig 4.13). Pour pouvoir comparer les rugosités entre différents sons, il est préférable de considérer une seule valeur de rugosité par son. Nous avons donc défini une rugosité globale à partir de l'équation 4.27 :

$$RUG = \int R_B(t)dt \tag{4.28}$$

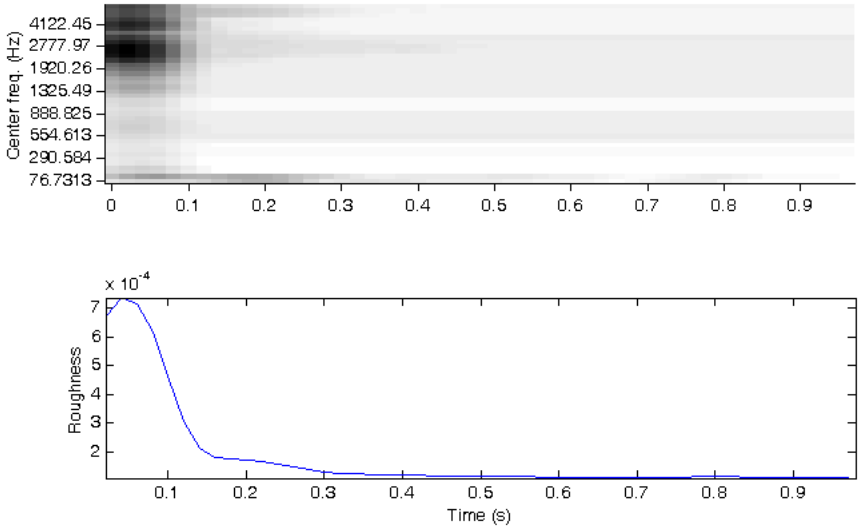


FIG. 4.13 – Rugosité d'un impact de verre, (en haut) : en fonction du temps et de la fréquence, plus la couleur est foncée plus la rugosité est importante, (en bas) : rugosité de l'impact au cours du temps.

III.3 Le centroïde de rugosité : *CGR*

Le calcul du centroïde de rugosité consiste à estimer la fréquence moyenne où le plus de rugosité à été observée.

$$CGR = \frac{\sum_k k.r(k)}{\sum_k r(k)} \tag{4.29}$$

où k prend les 40 valeurs de fréquences centrales des canaux ($C = 40$, voir eq. 4.17) utilisées par Leman dans le calcul de rugosité, et *r(k)* représente la rugosité accumulée dans la bande de fréquence centrale k (haut de la figure 4.13).

III.4 Le rapport CGR/CGS : *RSF*

Il s'agit de la valeur de *CGR* normalisée par le centroïde spectral (*CGS*).

$$RSF = \frac{CGR}{CGS} \tag{4.30}$$

Ceci nous permet d'obtenir une grandeur adimensionnée.

IV Conclusion sur la caractérisation acoustique

Sur la base de ces descripteurs, un son n'est plus caractérisé par sa forme d'onde mais globalement par 17 valeurs de descripteurs, choisis pour leur pertinence d'un point de vue perceptif. Ces descripteurs permettent de quantifier des comportements temporels, spectraux ou spectro-temporels du signal acoustique. Les modèles que l'on veut établir vont se baser sur les valeurs de descripteurs les plus pertinents pour prédire les pourcentages d'appartenance à une catégorie donnée. Désormais nous avons tous les éléments pour construire nos modèles, mais avant cela, il est important de mieux connaître les relations qui existent entre ces descripteurs, ceci en passant par une analyse statistique descriptive.

Chapitre 5

Analyse statistique descriptive

I Mise en forme des données

Afin de pouvoir analyser toutes les données résultantes aussi bien des tests perceptifs que des calculs de descripteurs, nous rassemblons tous les résultats nécessaires à la conception des modèles prédictifs dans une matrice.

Cette matrice regroupe les résultats correspondant à 5 sessions de 65 sons écoutés par 13 sujets pour la première session, 11 sujets pour les deuxièmes, troisièmes et quatrièmes sessions et 9 sujets pour la dernière session.

Désormais, nous ferons la distinction entre sons perçus et sons absolus. Un son absolu engendre plusieurs sons perçus car écoutés par plusieurs personnes différentes. Par exemple, le son absolu n°1 de la première session, engendre les sons perçus 1.1, 1.2, ... , 1.13 car il a été écouté par 13 sujets.

Au total, la matrice totale est constituée de 3575 lignes (65*(13+3*11+9) = 3575) et de 33 colonnes qui sont décrites ci-dessous.

colonnes 1 à 3 : labelisation

La première colonne indique le numéro du son (de 1 à 325), la deuxième colonne indique le numéro du sujet qui à écouté ce son (de 1 à 20 : 20 sujets différents ont passé au moins l'une des 5 sessions), la troisième colonne est un label fabriqué à partir du numéro du son et du numéro du sujet qui l'a écouté afin de rendre la ligne différentiable des autres.

colonnes 4 à 16 : résultats des tests perceptifs

Les colonnes 4 à 9, respectivement *perçuMetal*, *perçuBois*, *perçuPierre*, *perçuPlastique*, *perçuVerre* et *perçuAutre*, correspondent aux réponses données par les sujets. Par exemple, lorsque le sujet à répondu

métal à l'écoute d'un son, alors on note 1 dans la colonne *perçuMetal* et 0 dans les autres.
La colonne 10 correspond au type d'impacteur utilisé : 1 pour un impacteur dur, 0 pour un impacteur mou.
Les colonnes 11 à 16 indique le véritable matériau impacté. Ainsi on a respectivement les colonnes *réelMetal*, *réelBois*, *réelPierre*, *réelPlastique*, *réelVerre* et *réelAutre*.

colonnes 17 à 33 : les descripteurs de timbre

Les descripteurs de timbres sont ceux cités dans le chapitre 4 page 21. Ils sont labelisés de la manière suivante :

- RUG = la rugosité.
- CGS = le centre de gravité spectral.
- FSP = le flux spectral.
- TED = la « durée perceptive ».
- AMO = l'amortissement.
- ZCR = le taux de passage par zéro
- SSP = l'étalement spectral, en Hz.
- LAT = le log du temps d'attaque.
- CGT = le centre de gravité temporel.
- ROF = la « point de roulement spectral ».
- RMS = la valeur efficace.
- CGR = le centre de gravité de la rugosité.
- RSF = $\frac{CGR}{CGS}$
- SCR = taux de croisement du spectre avec un seuil.
- SKN = l'indice de dissymétrie spectrale.
- KRT = le kurtosis.

C'est cette matrice qui va nous servir pour concevoir les modèles prédictifs. Ces modèles seront déterminés à l'aide d'un outil informatique : le logiciel *SPSS*.

II Présentation du logiciel *SPSS*

Les données ont été analysées avec le logiciel *SPSS* (*Statistical Package for the Social Sciences*). Développé essentiellement pour l'analyse statistique, *SPSS* est utilisé par des chercheurs en économie, en sciences de la santé, par des compagnies d'études, par le gouvernement, des chercheurs de l'éducation nationale etc.

SPSS va nous permettre d'effectuer deux types d'analyses : une analyse descriptive (analyse en composante principale) qui nous permettra d'étudier les corrélations entre descripteurs, et une analyse prédictive (régression logistique binaire) qui va nous permettre de concevoir un modèle pour catégoriser de manière probabiliste et automatique nos différents sons d'impacts.

III Analyse de corrélation

Cette analyse descriptive à pour objectif d'évaluer la corrélation des descripteurs entre eux. En effet, si deux descripteurs se trouvent être trés corrélés, cela signifie qu'ils apportent la même information et donc on peut se priver de l'un des deux pour catégoriser. Cependant la méthode prédictive employée par la suite, (voir I page 49) tient compte des liens de corrélations entre descripteurs et ne conserve que ceux qui sont les moins corrélés entre eux, inutile donc d'en éliminer, nous nous servirons de cette analyse uniquement pour vérifier la cohérence des descripteurs.

La matrice de corrélation des descripteurs de signaux à été obtenue avec *SPSS*, mais nous pourrions écrire un code de programmation menant au même résultat en suivant l'algorithme décrit ci-dessous.

Echantillon

Nous considérons un ensemble de N descripteurs X_1,...,X_N connus à partir d'un échantillon de K réalisations conjointes ces variables. Dans notre cas, K = 325.

$$\mathrm{M} = \begin{bmatrix} X_{1,1} & \dots & X_{N,1} \\ \dots & \dots & \dots \\ X_{1,K} & \dots & X_{N,K} \end{bmatrix}$$

L'ensemble des points qui représentent les variables est communément appelé « nuage des variables », [DS06]. Chaque variable aléatoire $X_n = (X_{n,1},...,X_{n,K})$' a une moyenne μ_{X_n} et un écart-type σ_{X_n}.

Transformations de l'échantillon

La matrice M est centrée sur le centre de gravité :

$$\bar{M} = \begin{bmatrix} X_{1,1} - \mu_1 & \dots & X_{N,1} - \mu_N \\ \dots & \dots & \dots \\ X_{1,K} - \mu_1 & \dots & X_{N,K} - \mu_N \end{bmatrix}$$

Elle est ensuite réduite :

$$\tilde{\mathrm{M}} = \begin{bmatrix} \frac{X_{1,1}-\mu_1}{\sigma(X_1)} & \dots & \frac{X_{N,1}-\mu_N}{\sigma(X_N)} \\ \dots & \dots & \dots \\ \frac{X_{1,K}-\mu_1}{\sigma(X_1)} & \dots & \frac{X_{N,K}-\mu_N}{\sigma(X_N)} \end{bmatrix}$$

Calcul de corrélation

Une fois la matrice $\tilde{M}$ obtenue, on la multiplie par sa transposée pour obtenir la matrice de corrélation des X_1,...X_N (voir fig 5.1).

$$C = \tilde{\mathrm{M}}^T.\tilde{\mathrm{M}} \tag{5.1}$$

où C est la matrice de corrélation des descripteurs de signaux et $\tilde{M}$ la matrice d'échantillons centrée réduite.

	rug	cgs	fsp	ted	amo	zcr	ssp	lat	cgt	r85	rms	cgr	rsf	sdc	scr	krt	skn
rug	1,000	-,047	,244	,191	,317	-,092	,060	,042	,378	,010	-,216	-,468	-,126	-,297	-,172	,150	,161
cgs	-,047	1,000	-,155	,002	,501	,854	,720	-,494	-,073	,935	-,410	,410	-,765	-,571	,185	-,051	-,211
fsp	,244	-,155	1,000	,213	-,034	-,163	-,014	,134	,260	-,098	,241	-,277	,189	,103	-,455	,277	,453
ted	,191	,002	,213	1,000	,065	,022	,016	,234	,675	-,005	,199	-,242	-,002	-,121	-,155	,750	,690
amo	,317	,501	-,034	,065	1,000	,382	,424	-,278	,183	,493	-,618	-,049	-,826	-,743	,048	,035	-,089
zcr	-,092	,854	-,163	,022	,382	1,000	,420	-,409	-,010	,690	-,368	,426	-,590	-,532	,223	-,028	-,193
ssp	,060	,720	-,014	,016	,424	,420	1,000	-,301	-,021	,860	-,256	,134	-,651	-,305	,015	-,035	-,123
lat	,042	-,494	,134	,234	-,278	-,409	-,301	1,000	,212	-,436	,524	-,322	,397	,293	-,171	,138	,296
cgt	,378	-,073	,260	,675	,183	-,010	-,021	,212	1,000	-,065	-,058	-,476	,032	-,180	-,154	,652	,693
r85	,010	,935	-,098	-,005	,493	,690	,860	-,436	-,065	1,000	-,367	,292	-,751	-,503	,103	-,070	-,203
rms	-,216	-,410	,241	,199	-,618	-,368	-,256	,524	-,058	-,367	1,000	-,019	,506	,507	-,265	,092	,335
cgr	-,468	,410	-,277	-,242	-,049	,426	,134	-,322	-,476	,292	-,019	1,000	-,251	-,027	,228	-,210	-,301
rsf	-,126	-,765	,189	-,002	-,826	-,590	-,651	,397	,032	-,751	,506	-,251	1,000	,659	-,154	,049	,233
sdc	-,297	-,571	,103	-,121	-,743	-,532	-,305	,293	-,180	-,503	,507	-,027	,659	1,000	-,039	-,034	,113
scr	-,172	,185	-,455	-,155	,048	,223	,015	-,171	-,154	,103	-,265	,228	-,154	-,039	1,000	-,133	-,332
krt	,150	-,051	,277	,750	,035	-,028	-,035	,138	,652	-,070	,092	-,210	,049	-,034	-,133	1,000	,865
skn	,161	-,211	,453	,690	-,089	-,193	-,123	,296	,693	-,203	,335	-,301	,233	,113	-,332	,865	1,000

FIG. 5.1 – Matrice de corrélation des descripteurs de signaux

Observations sur la matrice de corrélation

Cette matrice de corrélation aurait été idéale si aucun des descripteurs n'avait été très corrélés à un autre, ici on observe quelques corrélations fortes (supérieures à 0.7 en valeur absolue) mais la majorité des descripteurs sont peu corrélés à d'autres ce qui permet de valider nos choix de en soulignant que chaque descripteur apporte une information différente sur le signal. Certaines corrélations fortes sont également attendues et permettent de vérifier si les calculs de nos descripteurs sont cohérents. Ainsi on observe que le *ROF* est très corrélé au *CGS*, en effet le *ROF* est une borne supérieure du *CGS*. Le *CGS* est aussi correlé à l'étalement spectral (*SSP*) ce qui est cohérent également : un signal de grande étendue spectrale verra son *CGS* augmenter. Le *ZCR* est très correlé au *CGS* : en effet plus le signal croise l'axe des zeros, plus son spectre est riche en hautes fréquences.

IV L'Analyse en Composante Principale

L'Analyse en Composante Principale (ACP) est une méthode mathématique d'analyse de données qui consiste à rechercher les directions de l'espace qui représentent le mieux les corrélations entre n variables aléatoires. L'ACP est aussi connue sous le nom de transformée de Karhunen-Loève ou de transformée de Hotelling. A partir d'un ensemble de n sons dans un espace de p descripteurs, son but est de trouver une représentation dans un espace réduit de k dimensions ($k < p$) qui conserve « le meilleur résumé » (voir `http://fr.wikipedia.org/wiki/Analyse_en_composantes_principales`). C. Duby dans [DS06] rappelle que, comme toute méthode descriptive, réaliser une ACP n'est pas une fin en soi. L'ACP nous servira à mieux connaître les données sur lesquelles on travaille et à détecter éventuellement les valeurs suspectes. On pourra aussi se servir des représentations fournies par l'ACP pour illustrer certains résultats dans un but pédagogique.

Critère d'inertie

Le principe de l'ACP est de trouver un axe u, issu d'une combinaison linéaire des X_n, tel que la variance du nuage autour de cet axe soit maximale. Maximiser la variance expliquée par le vecteur u revient à maximiser l'inertie expliquée par u, c'est à dire minimiser l'inertie du nuage autour de u. Finalement, nous cherchons le vecteur u tel que la projection du nuage sur u aie une variance maximale. La projection de l'échantillon des X sur u s'écrit :

$$\pi_u(M) = M.u \tag{5.2}$$

La variance de $\pi_u(M)$ vaut donc :

$$\pi_u(M)^T.\pi_u(M) = u^T.C.u \tag{5.3}$$

Comme la matrice de corrélation C (décrite dans l'équation 5.1) est diagonalisable dans une base orthonormée, notons P le changement de base associé et Δ la matrice diagonale formée du spectre de C :

$$\pi_u(M)^T.\pi_u(M) = u^T.P^T\Delta.P.u = (P.u)^T.\Delta.(Pu) \tag{5.4}$$

$$Pu = v \tag{5.5}$$

Après cette réécriture, nous cherchons le vecteur unitaire v qui maximise $v'\Delta v$ où Δ=diag(λ_1,...,λ_N) est diagonale et dont les valeurs sont rangées par ordre croissant. On observe qu'il suffit de prendre le premier vecteur unitaire, on a alors :

$$v^T.\Delta.v = \lambda_1 \tag{5.6}$$

On continue la recherche du deuxième axe de projection w sur le même principe en imposant qu'il soit orthogonal à u. En appliquant une ACP sur les descripteurs, on obtient 4 composantes principales qui permettent d'expliquer 74% de la variance totale.

	Component			
	1	2	3	4
cgs	,903	,157	,304	,089
rsf	-,870	-,237	,056	-,037
r85	,861	,186	,252	,246
zcr	,767	,113	,317	-,098
ssp	,683	,217	,198	,379
amo	,679	,413	-,392	-,026
sdc	-,670	-,386	,277	,100
rms	-,637	-,082	,474	,283
lat	-,607	,121	,012	,034
cgt	-,200	,832	-,050	-,248
krt	-,248	,776	,331	-,276
ted	-,203	,775	,310	-,231
skn	-,459	,756	,322	-,051
rug	,007	,502	-,615	,138
cgr	,407	-,440	,553	-,079
scr	,291	-,309	,029	-,683
fsp	-,309	,428	,021	,587

FIG. 5.2 – Les axes extraits par l'ACP

On remarque que le *CGS*, le *RSF*, le *ROF*, le *ZCR*, le *SDC*, le *SSP* sont principalement portés par le premier axe (dominante fréquentielle), le *TED*, le *CGT*, le *KRT* et le *SKN* sont portés par le deuxième axe (mi fréquentiel, mi temporel), la *RUG* et le *CGR* par le troisième axe (dominante spectro-temporelle), le *SCR* et le *FSP* par le quatrième axe (mi fréquentiel, mi spectro-temporel).

Il est intéressant de voir comment sont distribués les sons dans l'espace des axes engendrés par l'ACP. La figure IV représente la dispersion des sons typiques sur les 2 premiers axes. On considère comme typique un son classé dans une catégorie par au moins 65% des personnes. Si un son n'est pas classé par au moins 65% des personnes dans une des cinq catégories, il sera alors considéré comme ambiguë et ne sera pas représenté ici.

À la lecture de ce graphique, il naît l'espoir de pouvoir dissocier certaines catégories entre elles. On voit que les descripteurs portés par l'axe 2 pourront distinguer le métal et le verre du bois et de la pierre, de même les descripteurs portés par l'axe 1 pourront quand à eux dissocier la pierre du bois. Enfin une combinaison linéaire des descripteurs portés par l'axe 1 et 2 pourrait isoler la pierre des autres matériaux. Ainsi on a l'impression que la pierre est isolée mais il est important de préciser que le nombre de sons typiques pour la pierre sont peu nombreux sur ce graphe, ces résultats sont donc à confirmer.

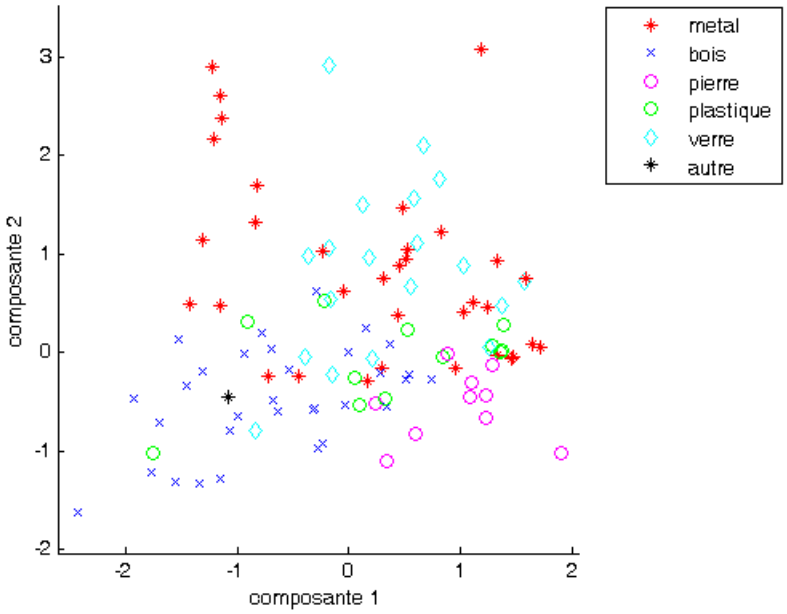


FIG. 5.3 – Projection des sons typiques dans l'espace engendré par les deux premiers axes de l'analyse en composante principale.

V Conclusion sur l'analyse descriptive

Cette analyse descriptive nous permet de mieux comprendre les relations entre nos descripteurs. Les inter-corrélations sont faibles dans l'ensemble et par conséquent ces résultats nous conforte dans notre choix de descripteurs. Ces informations ne sont pas directement utiles dans l'élaboration de nos modèles finaux, mais vont nous permettre de pondérer si nécessaire les interprétations possibles. En particulier, la représentation des sons typiques de chaque catégorie de matériau dans l'espace engendré par les deux premiers axes de l'ACP nous renseigne sur les limitations éventuelles de nos modèles. Ayant ces informations en tête, il faut maintenant établir un modèle de catégorisation pour chaque matériau, nous entrons alors dans l'analyse statistique prédictive.

Chapitre 6

Analyse statistique prédictive

I Principe de la régression logistique binaire

La régression logistique est un des modèles multivariables couramment utilisé dans la construction de modèles statistiques. Elle permet de mettre en relation une variable à expliquer Y avec une ou plusieurs variables explicatives [POP+05] , [DC]. Ici, nos variables explicatives sont les valeurs de nos descripteurs, ces valeurs sont continues. Notre variable à expliquer et, pour chaque catégorie de matériau, l'appartenance ou non à cette catégorie (variable dichotomique), on parle alors de régression logistique binaire (par opposition à la régression logistique multiple lorsque la variable à expliquer est continue).

Cette méthode à été choisie car elle est la forme la plus courante de régression logistique [DC].

On supposera que la variable Y à laquelle on s'intéresse est pour un son donné, l'appartenance ou non à une catégorie de matériau (Y=1 ou Y=0) avec une probabilité π donnée par :

$$\pi(x) = P(Y = 1/X = x) = f(L_{Cat}(x)) = \frac{e^{L_{Cat}(x)}}{1 + e^{L_{Cat}(x)}} \text{ avec } L_{Cat}(x) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \tag{6.1}$$

La fonction logistique $f(L_{Cat}(x))$ a la forme d'une sigmoïde. Les x_i représentent nos descripteurs (ils sont au nombre de n) et les β_i leur coefficients associés.

Calibrer un modèle consiste à estimer les paramètres β_i de la fonction $L_{Cat}(x)$ sur la base d'un échantillon de taille n. Pour cela, on utilise la méthode du maximum de vraisemblance qui vise à fournir une estimation des paramètres qui maximise la probabilité d'obtenir les valeurs réellement observées sur l'échantillon, cette méthode nous amène un système de deux équations que l'on ne peut résoudre que de manière itérative, voir [DC].

SPSS propose deux méthodes itératives : premièrement la méthode *backward* qui consiste à considérer tous les descripteurs au départ. A chaque étape, le descripteur le moins significatif est retiré du modèle et une matrice de contingence est calculée (pourcentage de non-détection, détection vraie, fausse alarme,

non-détection vraie, pourcentage de réussite globale). Les descripteurs sont donc retirés un à un jusqu'à atteindre le meilleur coefficient de détermination de Nagelkerke R^2.

A l'inverse, dans la méthode *forward*, on considère un seul descripteur au départ, c'est le plus significatif. A chaque étape, un descripteur supplémentaire est ajouté, il s'agit du descripteur qui décrit au mieux la variable dépendante et qui est le moins corrélé au descripteur précédent, puis une matrice de contingence est calculée. Ceci est répété jusqu'à obtenir le meilleur coefficient de détermination de Nagelkerke R^2.

Un bon modèle est un modèle stable donc comprenant le moins de descripteurs possible, c'est pourquoi nous préférerons utiliser la méthode *forward*.

II Modèles prédictifs pour chaque catégorie

II.1 Précautions d'usages

L'emploi de la régression logistique nécessite le respect d'un certain nombre de précautions. Celles-ci sont toutes détaillées dans [DC]. Nous allons vérifier celles qui correspondent à notre cas d'utilisation.

La taille *n* de l'échantillon doit être élevée, idéalement une centaine d'individus, dans notre cas $n = 3575$. Ce premier critère est donc vérifié.

La proportion d'individus codés 1 doit être au minimum de 5%. Dans notre cas, nous avons :

métal	19%
bois	24%
pierre	16%
plastique	21%
verre	12%

Ce critère est vérifié pour toutes les catégories.

De plus, cette proportion d'individus codés 1 est généralement évaluée en fonction du nombre de descripteurs : l'*EPV*[1], il s'agit du nombre de 1 par descripteur. D'après [CKHF96], un *EPV* minimum de 10 est souhaité. Dans notre cas, nous avons :

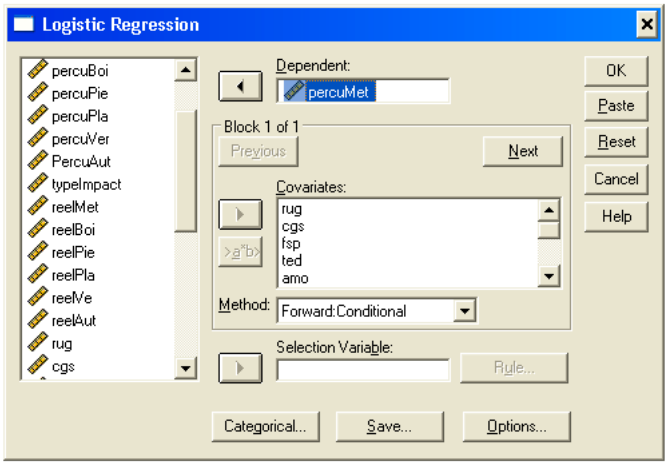
EPV métal	40
EPV bois	51
EPV pierre	33
EPV plastique	45
EPV verre	25

[1] events per variable

Le critère de l'*EPV* est vérifié pour toutes les catégories.

II.2 Calibration

Grâce au logiciel de statistique *SPSS*, l'élaboration d'un modèle de régression logistique se réalise en quelques clics. En pratique, un modèle prédictif basé sur la régression logistique binaire est construit pour chaque catégorie de matériau. La calibration de ce modèle est effectué sur les 325 sons (cf. page 16) et plus précisément la matrice globale 3575×33 qui suppose que chaque ligne est une observation indépendante. La première étape consiste à choisir notre variable dépendante (par exemple *percuMétal* pour le modèle de la catégorie « Métal », variable binaire), puis de sélectionner les variables indépendantes (ici nos descripteurs) et enfin de sélectionner la méthode pas-à-pas d'entrée des descripteurs (ici *forward method* comme vu dans la section I).



A chaque pas (i.e à chaque fois qu'un descripteur est intégré), le modèle est évalué par une matrice de contingence qui définit le pourcentage de détection vraie et de non-détection vraie. Ce que l'on veut, c'est détecter au mieux la catégorie d'un son, ainsi notre critère d'arrêt est de considérer l'étape où le pourcentage de détection vraie est maximal.

Chaque modèle prédictif est basé sur une probabilité d'appartenance exprimée par la fonction $\pi(x)$ dans l'équation 6.1. Les fonctions $L_{Cat}(x)$ pour chaque catégorie $\{Métal, Bois, Pierre, Plastique, Verre\}$ sont données ci-dessous :

Catégorie métal

$$L_{Metal} = 0.030CGT - 0.008CGR - 0.001KRT + 0.0001ROF + 0.026ZCR + 0.199LAT + 0.013TED + 8.091$$

Catégorie bois

$$L_{Bois} = -0.001CGS - 0.164SKN - 112.946SDC - 180.501RUG + 0.002KRT - 71405.1AMO - 0.015CGT + 0.0002SSP + 6.965$$

Catégorie pierre

$$L_{Pierre} = -25.369RMS - 0.053CGT + 131.297SCR + 0.616$$

Catégorie plastique

$$L_{Plastique} = -0.43ZCR + 0.0004SSP - 171.204RUG + 2.216$$

Catégorie verre

$$L_{Verre} = 0.055ZCR + 5.821FSP + 0.168SKN + 0.0003ROF - 0.002KRT - 1.511RSF + 0.004CGR - 11.025$$

Pour chaque modèle, les descripteurs apparaissent par ordre d'importance : le premier descripteur présenté est celui qui est en majeure partie responsable de l'identité du matériau.

II.3 Validation

L'étape de validation permet d'estimer la performance des modèles calibrés précédemment, en soumettant à ces modèles des nouveaux sons (les 200 sons de la validation cf page 16). Chacun des 200 sons ont été écoutés par 8 auditeurs (non pongistes), chaque auditeur à classifié chaque son dans une des catégories de matériau (réponse binaire). Nous comparons alors les réponses de chaque individu avec la prédiction du modèle. On considérera que le modèle à classifié le son dans la catégorie où le pourcentage d'appartenance à cette catégorie est supérieure à 50%. On génère alors, pour chaque modèle, une matrice de contingence :

Métal	prédit / observé	Non	Oui	% correct
	Non	1138	131	90
	Oui	150	181	55
	% global	88	58	82

Bois	prédit / observé	Non	Oui	% correct
	Non	1142	106	92
	Oui	226	126	36
	% global	83	54	79

Pierre	prédit / observé	Non	Oui	% correct
	Non	1299	36	97
	Oui	229	36	14
	% global	85	50	83

Plastique	prédit / observé	Non	Oui	% correct
	Non	1185	43	98
	Oui	319	53	14
	% global	79	55	77

Verre	prédit / observé	Non	Oui	% correct
	Non	1402	23	98
	Oui	118	57	33
	% global	92	71	91

Nous obtenons des seuils de non-détection vraie excellents pour chaque matériau. A l'inverse, les seuils de détection vraie sont plus faibles, particulièrement la pierre et le plastique qui présentent les plus faibles pourcentages de détection vraie.

II.4 Autres calibrations

Afin de nous assurer de la cohérence de nos modèles, nous avons calibré le smodèles avec les sons destinés cette fois-ci à la validation puis avec la totalité des sons. Pour comparer les différents résultats obtenus, on appellera le modèle A, l'ensemble des modèles calibrés avec les 325 premiers sons (élaboré précédemment), le modèle B sera l'ensemble des modèles calibrés avec les 200 derniers sons, et le modèle A+B sera l'ensemble des modèles calibrés avec la totalité des 525 sons. On peut ainsi supposer que les descripteurs mis en évidence dans plusieurs modèles seront les plus pertinents pour caractériser la catégorie de matériau.

catégorie métal

modèle A

$$L_{Metal} = 0.030CGT - 0.008CGR - 0.001KRT + 0.0001ROF + 0.026ZCR + 0.199LAT + 0.013TED + 8.091$$

modèle B

$$L_{Metal} = 0.054CGT - 0.135SKN + 0.018ZCR - 271.8SCR + 13.101RMS - 1.714FSP - 2.635$$

modèle A+B

$$L_{Metal} = 0.026CGT + 20.001RMS + 1.815RSF + 0.0002ROF + 505360.6AMO + 159.701RUG - 2.678FSP - 0.001KRT - 0.004CGR - 1.435$$

Il semble que le centre de gravité temporel (*CGT* présent dans les trois modèles) soit un facteur important dans la perception du métal. On peut expliquer ceci du fait que le son d'un impact de métal s'amortit moins vite, ainsi son *CGT* est en moyenne supérieur au sons d'impacts d'autres matériaux.

cas du bois

modèle A

$$L_{Bois} = - 0.001CGS - 0.164SKN - 112.946SDC - 180.501RUG + 0.002KRT - 71405.1AMO - 0.015CGT + 0.0002SSP + 6.965$$

modèle B

$$L_{Bois} = -0.039CGT - 0.002CGS - 246.38SDC - 0.201SKN + 0.003KRT + 0.0003ROF - 120.724SCR + 5.2$$

modèle A+B

$$L_{Bois} = -0.001CGS - 128.970SDC - 0.138SKN - 0.021CGT - 146240AMO - 119.119RUG + 0.001KRT - 0.0004SSP - 0.692RSF + 0.0002ROF + 0.009TED + 6.612$$

Il semble que perception du bois soit essentiellement due à des descripteurs fréquentiels : on note la présence dans les trois modèles du centre de gravité spectral (*CGS*), de la décroissance spectrale (*SDC*) et du kurtosis (*KRT*), également d'un descripteur temporel : le *CGT*.

cas de la pierre

modèle A

$$L_{Pierre} = -25.369RMS - 0.053CGT + 131.297SCR + 0.616$$

modèle B

$$L_{Pierre} = -26.679RMS + 0.015CGR + 123.578SCR - 23.582$$

modèle A+B

$$L_{Pierre} = -26.386RMS - 0.034CGT + 104.504SCR + 0.005CGR - 7.800$$

La pierre semble contenir autant de caractéristiques propres sur le plan temporel que fréquentiel, en effet, les trois modèles s'accordent pour s'appuyer sur la valeur efficace *RMS* et la décroissance spectrale *SCR*.

cas du plastique

modèle A

$$L_{Plastique} = -0.43ZCR + 0.0004SSP - 171.204RUG + 2.216$$

modèle B

$$L_{Plastique} = 0.001SSP - 0.028TED - 0.030ZCR + 1.058RSF - 3.937$$

modèle A+B

$$L_{Plastique} = 0.001SSP - 0.044ZCR - 127.495RUG - 0.12TED - 0.001KRT + 1.130$$

Les sons perçus comme du plastique semblent être caractérisés par le *ZCR* et le *SSP*. Ceci souligne les caractéristiques spectrales du plastique.

cas du verre

modèle A

$$L_{Verre} = 0.055ZCR + 5.821FSP + 0.168SKN + 0.0003ROF - 0.002KRT - 1.511RSF + 0.004CGR - 11.025$$

modèle B

$$L_{Verre} = 0.415SKN + 0.049ZCR - 0.004KRT + 0.013CGR - 25.289$$

modèle A+B

$$L_{Verre} = 0.072ZCR + 0.227SKN + 4.416FSP - 0.002KRT - 0.018CGT - 0.0001ROF - 6.466$$

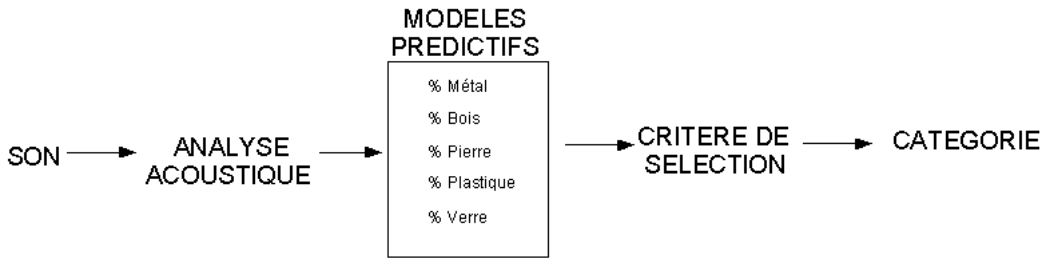
Il semble que la perception du verre relève essentiellement de descripteurs spectraux : on note l'omniprésence du zero-crossing rate (*ZCR*), du skewness (*SKN*) et du kurtosis (*KRT*).

Conclusion

Pour chaque catégorie, certains descripteurs restent présents quels que soient les modèles. Cela nous conforte dans le choix de nos descripteurs et dans la fiabilité de nos modèles. Mais l'on remarque également beaucoup de différences selon la banque de sons retenue pour la calibration. On peut penser que ces différences sont dues au nombre de sons utilisés pour la calibration, qui varie selon les modèles.

III Modèle prédictif global

L'idée consiste à proposer une méthode globale de classification basée sur les modèles prédictifs précédemment construits. Ainsi, cette méthode permettra de classer un son donné dans une des cinq catégories {*Métal, Bois, Pierre, Plastique, Verre*} en fonction de la probabilité d'appartenance donnée par chaque modèle :



Le critère de sélection de la catégorie la plus probable est définie selon deux conditions :
- la catégorie pour laquelle le pourcentage maximum à été obtenu (quelque soit sa valeur).
- on ajoute à la condition précédente le fait que le pourcentage d'appartenance à une catégorie doit également dépasser le seuil du hasard, ici 20% (1 sur 5).

Cette méthode globale à été évaluée avec les pourcentages de classification obtenus dans le test perceptif :

	Observé					Rep.	Prédit					Rep.
	% mét	% boi	% pie	% pla	% ver		% mét	% boi	% pie	% pla	% ver	
son 1	0	53.86	30.77	7.69	7.69	boi	11.31	44.87	12.83	11.55	6.18	boi
...	...	...	...	...	...	...	...	...	...	...	...	...
son 525	92.30	0	0	0	0	met	36.99	16.86	5.33	29.44	4.24	met

En particulier nous déterminons si les catégories prédites et observées (Rep.) sont similaires ou différentes.

Le taux de concordance des résultats de classification est donné :

catégorie métal	70%
catégorie bois	82%
catégorie pierre	49%
catégorie plastique	56%
catégorie verre	52%

On note de forts taux de réussite pour le bois et le métal. En revanche le verre, la pierre et le plastique peinent à être catégorisés correctement par rapport aux données expérimentales, ce qui est cohérent avec le fait que les modèles prédictifs pour ces trois catégories sont moins performants

IV Modèles prédictifs physiques

Pour chaque son, nous connaissions le matériau réel à l'origine de l'impact et le type d'impacteur (dur ou mou) utilisé. Il fut donc possible d'établir un modèle statistique permettant de prédire les attributs physiques de la source sonore, en particulier, le type d'impact (dur ou mou), ainsi que le matériau réellement impacté (métal, bois, pierre, plastique ou verre) sur les 325+200 sons enregistrés, cette fois-ci indépendamment de la manière dont ces sons ont été perçus par les auditeurs.

IV.1 Modèle prédictif pour le type d'impact : dur / mou

En effectuant une régression logistique sur les 325+200 sons étudiés, le *CGS* à lui seul détecte le type d'impact avec un pourcentage de vraie détection de 81.7% et de non détection vraie de 79.5%. Ce résultat est cohérent : plus l'excitateur percutant est mou moins l'énergie fournie au système vibrant est importante et moins le spectre est étendu (les composantes dans les hautes fréquences sont moins excitées que les composantes basses fréquences). La différence est d'ailleurs remarquable à l'écoute.

Catégorie excitation

$$L_{Excit} = 0.001CGS - 3.433$$

observé \ prédit	Mou	Dur	% correct
Mou	1552	261	86
Dur	318	1417	82
% global	83	84	83

IV.2 Modèle prédictif du matériau réellement impacté

De la même manière nous avons élaboré un modèle statistique permettant de prédire quel matériau a été réellement impacté. Dans ce cas, le classement observé correspond à la catégorie du matériau physique de l'objet impacté (cf. colonnes 11 à 16 de la matrice globale page 41).

Catégorie métal réel

$$L_{Reel-Metal} = 0.001SSP + 0.048ZCR - 4.823FSP + 833623{,}2AMO + 2.129RSF - 0.001KRT + 21.844RMS - 0.007CGR + 201.775RUG + 0.20CGT - 261.755SCR + 0.138LAT + 0.12TED + 0.910$$

Reel-Metal	prédit / observé	Non	Oui	% correct
	Non	4026	171	95.9
	Oui	463	480	50.9
	% global	89.7	73.7	87.7

Catégorie bois réel

$$L_{Reel-Bois} = -0.311SKN - 0.001SSP + 0.003KRT + 0.009CGR + 18.573RMS$$
$$-1.098RSF - 108.497SDC - 0.042ZCR - 0.0004CGS - 0.090LAT + 29.642$$

Reel-Bois	prédit / observé	Non	Oui	% correct
	Non	3894	257	93.8
	Oui	638	386	37.7
	% global	85.9	60.0	83.3

Catégorie pierre réelle

$$L_{Reel-Pierre} = -0.059CGT - 0.144ZCR - 39.834RMS + 914051.6AMO + 325.397SDC$$
$$+ 0.002CGS - 0.002SSP + 148.654SCR - 0.0004ROF - 0.204LAT - 0.004CGR + 10.837$$

Reel-Pierre	prédit / observé	Non	Oui	% correct
	Non	3938	197	95.2
	Oui	384	621	61.8
	% global	91.1	75.9	88.7

Catégorie plastique réel

$$L_{Plastique} = -0.002CGS - 386.319RUG + 0.001SSP - 180.040SDC + 19.123RMS$$
$$+ 0.001ROF - 0.036CGT - 131687AMO - 0.037TED - 0.087SKN + 17.093$$

Reel-Plastique	prédit / observé	Non	Oui	% correct
	Non	4044	100	97.6
	Oui	457	539	54.1
	% global	89.8	84.4	89.2

Catégorie verre réel

$$L_{Reel-Verre} = -0.001ROF + 616032.3AMO + 0.072ZCR - 624.385SCR + 0.127SKN$$
$$+ 195.540SDC - 0.001KRT + 1.990FSP + 0.001CGS - 0.010TED - 58.307RUG - 0.437RSF + 0.658$$

Reel-Verre	prédit / observé	Non	Oui	% correct
	Non	3982	171	95.9
	Oui	622	365	37
	% global	86.5	68.1	84.6

On remarque, d'une part, que les matrices de contingence sont bien meilleures pour les modèles prédictifs physiques que pour les modèles prédictifs perceptifs, et d'autre part que les descripteurs mis en évidences dans les modèles prédictifs physiques sont différents de ceux des modèles prédictifs perceptifs, ce qui montre une relation non directe entre la physique et la perception, ce qui d'ailleurs été mis en évidence par les résultats issus du test perceptif (matrices de confusion).

V Conclusion sur l'analyse prédictive

Grâce au principe de la régression logistique binaire nous avons établi un ensemble de modèles qui, pour un son d'impact, prédit les pourcentages d'appartenance à chacun des cinq matériaux d'un point de vue perceptif, ceci en fonction des valeurs des descripteurs calculés. Pour valider ces modèles, nous les avons soumis à des nouveaux sons et nous avons calculé les matrices de contingence. Pour évaluer la robustesse de ces modèles, nous avons comparé les descripteurs responsables de la catégorisation en calibrant d'autres modèles avec d'autres banques de sons, de cette comparaison nous avons pu mettre en évidence des descripteurs robustes dans la mesure où ils étaient présents dans tous les modèles. Enfin nous avons aussi établi des modèles prédictifs physiques permettant de prédire le type d'impact et la nature du matériau qui a été réellement impacté.

Chapitre 7

Application en temps réel

Durant les derniers jours du stage, nous avons développé une application Matlab, intitulée *Taping*. Après enregistrement d'un son d'impact grâce à un microphone relié à l'ordinateur, *Taping* calcule les 17 valeurs de descripteurs, et en fonction de ces valeurs, conformément aux modèles établis dans le chapitre précédent, nous fournit (voir fig 7.1) :

- le pourcentage d'appartenance perceptive à chaque catégorie de matériau.
- le pourcentage d'appartenance physique à chaque catégorie de matériau.
- le pourcentage de chance que l'impacteur utilisé soit un impacteur dur.
- les valeurs numériques des 17 descripteurs.
- une représentation graphique de 8 descripteurs : le temps d'attaque, l'*AMO*, le *CGT*, le *CGS*, le *SSP*, le *ROF*, le *CGR*, le *TED*.

Ces prédictions nécessitent un temps de calcul d'environ 30 secondes par son d'impact. Ce temps de calcul peut être optimisé en utilisant d'autres environnements d'implémentation couramment utilisés en informatique musicale tels que *MaxMSP*[1].

[1] `http://www.cycling74.com`

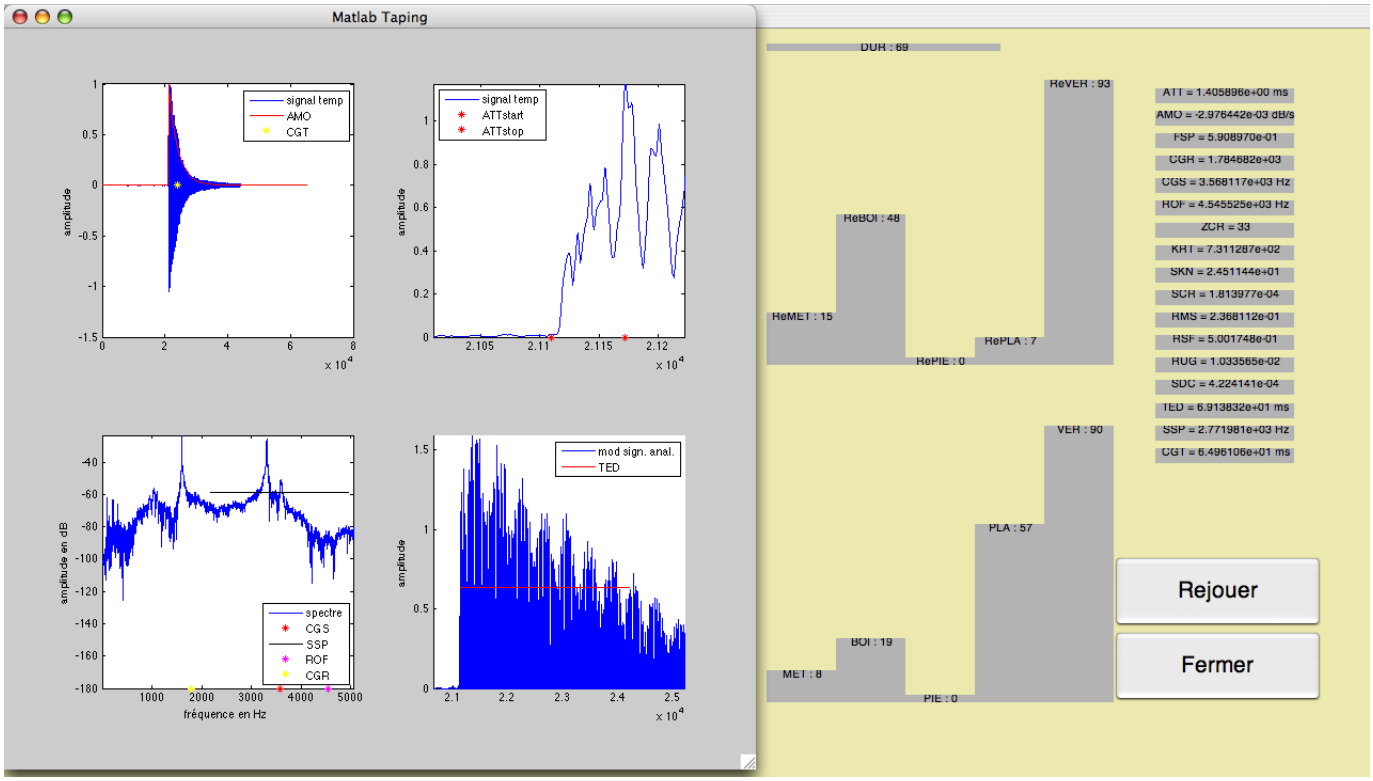


FIG. 7.1 – Interface de Taping : résultats suite à l'enregistrement d'un impact sur un verre avec un maillet dur, (à gauche) représentations de quelques descripteurs ; 1^{er} graphe : le *CGT*, l'*AMO*, 2^{e} graphe : l'*ATT*, 3^{e} graphe : le *CGS*, le *ROF*, le *SSP*, 4^{e} graphe : le *TED*, (à droite) : diagramme en barre du haut : type d'impact (ici le modèle estime qu'il y a 69% de chances que le maillet soit dur), diagrammes en barre du milieu : matériau réellement impacté (93% de chances que ce soit réellement du verre), diagrammes en barre du bas : matériau perçu (90% de chances que l'on entende du verre), sur la droite : valeurs des descripteurs correspondant au son enregistré.

Chapitre 8

Conclusion et perspectives

Cette étude nous a permis d'élaborer des modèles statistiques de prédiction permettant de classifier des sons d'impacts en fonction du matériau perçu. Sur la base de ces modèles, nous avons pu identifier quels descripteurs étaient en parti responsables de la caractérisation des matériaux d'un point de vue perceptif. Certains résultats connus ont été retrouvé comme l'importance du centre de gravité spectral pour la reconnaissance du bois et des descripteurs spectraux pour la reconnaissance du verre. En revanche, la rugosité et l'amortissement, qui ont été mis en évidence dans [ABKMY08], ne sont plus prépondérants dans notre cas. Ceci s'explique par le fait que les sons qui ont servi à la calibration de nos modèles sont issus d'objets impactés enregistrés dans notre environnement quotidien et qui ne sont généralement pas en oscillation libre. Ainsi ce sont d'autres descripteurs qui se sont présentés comme étant des prédicteurs classifiants comme la valeur efficace pour caractériser la pierre, le centre de gravité temporel pour caractériser le métal, ou le zero-crossing rate et l'étalement spectral pour caractériser le plastique. Nous pourrions alors compléter avec ces nouveaux descripteurs le système de synthèse sonore existant (*Tapong*) dans lequel seuls la rugosité et l'amortissement sont contrôlés à l'heure actuelle.

Nos taux de concordance entre prédictions et observations sont satisfaisants, ce qui permet de valider les modèles statistiques que l'on a mis au point dans cette étude. Toutefois certains modèles sont moins robustes que d'autres ce qui peut être expliqué par le fait que l'on avait moins de sons typiques dans ces catégories. Par ailleurs, ces modèles ont des limitations et ne visent pas à expliquer dans sa totalité notre perception des matériaux. Nous sommes conscient que les indices perceptifs utilisés par notre système auditif sont beaucoup plus complexes que les descripteurs que l'on a considéré, on peut en effet supposer que ces indices peuvent être approchés par une combinaison, certainement non linéaire, de tous ces descripteurs. Il serait envisageable de considérer d'autres descripteurs (nouveaux ou combinaison de ceux que nous avons utilisé) plus représentatifs de la manière dont notre système auditif fonctionne.

On note enfin que les modèles prédictifs physiques sont plus performants que les modèles prédictifs perceptifs, ceci conforte l'idée d'une relation non directe entre la physique et la perception. Si notre système auditif est très performant pour localiser un son dans l'espace 3D (le champ auditif est plus étendu que le champ visuel), il faut admettre qu'il est très limité pour reconnaître une grande variété de catégorie de matériaux (par rapport à nos autres systèmes sensoriels comme la vision ou le toucher). Du point de vue de la synthèse sonore, cette limitation de nos performances auditives simplifie la sonification d'une scène visuelle : en effet, au cinéma ou en réalité virtuelle, il suffirait de peu de catégories sonores de matériaux pour simuler une grande variété d'objets de notre vie quotidienne.

Bibliographie

[ABKMY08] Mitsuko Aramaki, Loïc Brancheriau, Richard Kronland-Martinet, and Solvi Ystad. Perception of impacted materials : sound retrival and synthesis control perspectives. *Computer Music Modeling and Retrieval*, 2008.

[AKM06] M. Aramaki and R. Kronland-Martinet. Analysis-synthesis of impact sounds by real-time dynamic filtering. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(2) :695–705, 2006.

[AKMVY06] Mitsuko Aramaki, Richard Kronland-Martinet, Thierry Voinier, and Solvi Ystad. A percussive sound synthesizer based on physical and perceptual attributes. *Computer Music Journal*, Summer 2006.

[BDKMY08] Mathieu Barthet, Philippe Depalle, Richard Kronland-Martinet, and Solvi Ystad. From performer to listener : an analysis of timbre variations. soumis à JASA, 2008.

[Bea82] J.W. Beauchamp. Synthesis by spectral amplitude and brigthness matching of analyzed musical instrument tones. *J. Audio Eng. Soc.*, 1982.

[Cac04] Anne Caclin. *Interactions et indépendances entre dimensions du timbre des sons complexes*. PhD thesis, Université Paris 6, 2004.

[CD97] A. Chaigne and V. Doutaut. Numerical simulations of xylophones. i. time-domain modeling of the vibrating bars. *Journal of the Acoustical Society of America*, 101(1) :539–557, 1997.

[CKHF96] P. Concato, J. Kemper, E. Holford, and T.R. Feinstein. A simulation study of the nulber of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 1996.

[CL01] A. Chaigne and C. Lambourg. Time-domain simulation of damped impacted plates : I. theory and experiments. *Journal of the Acoustical Society of America*, 109(4) :1422–1432, 2001.

[DC] F. Duyme and J.J Claustriaux. *La régression logistique binaire*.

[DS06] C. Duby and S.Robin. Analyse en composantes principales. Technical report, Institut National Agronomique Paris, 2006.

[GM06] B. L. Giordano and S. McAdams. Material identification of real impact sounds : Effects of size variation in steel, wood, and plexiglass plates. *Journal of the Acoustical Society of America*, 119(2) :1171–1181, 2006.

[Gre77] J. M. Grey. Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5) :1270–1277, 1977.

[KMS05] Hyong-Gook Kim, Nicolas Moreau, and Thomas Sikora. *MPEG-7 Audio and Beyond*. Wiley, 2005.

[KPK00] R. L. Klatzky, D. K. Pai, and E. P. Krotkov. Perception of material from contact sounds. *Presence : Teleoperators and Virtual Environments*, 9(4) :399–410, 2000.

[Lem00] M. Leman. Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (sim). *Procedings of the COST G-6 Conference on Digital Audio Effects (DAFW-00)*, 2000.

[Lem02] Marc Leman. Dealing with the data flood : Mining data, text and multimedia, 2002.

[LLT] Marc Leman, Micheline Leasaffre, and Koen Tanghe. *Toolbox for perception-based music analysis. Concepts, demos and reference manual*. Institute for Psychoacoustics and Electronic Music (IPEM), Ghent University, Blandijnberg 2, 9000 Ghent, Belgium.

[LO97] R.A. Lutfi and E.L. Oh. Auditory discrimination of material changes in a struck-clamped bar. *Journal of the Acoustical Society of America*, 102(6) :3647–3656, 1997.

[Mar04] J. Marozeau. L'effet de la fréquence fondamentale sur le timbre. Master's thesis, Université Pierre et Marie Curie, Paris VI, 2004.

[McA99] S. McAdams. Perspectives on the contribution of timbre to musical structure. *Computer Music Journal*, 23(3) :85–102, 1999.

[Meu] Sabine Meunier. Psychoacoustique musicale et psychoacoustique appliquée. Cours de Master Acoustique, EGIM.

[MWD+95] Stephen McAdams, Suzanne Winsberg, Sophie Donnadieu, Geert De Soete, and Jochen Krimphoff. Perceptual scaling of synthesized musical timbres : common dimensions, specificities, and latent subject classes. *Psychological Research*, 1995.

[Pee04] G. Peeters. *A Large Set Of Audio Features For Sound Description*, 2004.

[POP+05] P.M. Preux, P. Odermatt, A. Perna, B. Marin, and A. Vergnenègre. Qu'est ce qu'une régression logistique ? *Société de Pneumologie de Langue Française*, 2005.

[RDR+08] Asma Rabaoui, Manuel Davy, Stéphane Rossignol, Zied Lachiri, and Noureddine Ellouze. Sélection de descripteurs audio pour la classification des sons environnementaux avec des svms mono-classe. Technical report, Unité de recherche Signal, Image et Reconnaissance des formes (ENIT), Laboratoire d'Automatique, de Génie Informatique et Signal (INRIA), 2008.

[Ver03] Vincent Verfaille. *Effets audionumériques adaptatifs (théorie, mise en oeuvre et usage en création musicale numérique)*. PhD thesis, Université Aix-Marseille II, 2003.

[WR88] R. P. Wildes and W. A. Richards. *Recovering material properties from sound*, chapter 25, pages 356–363. W. A. Richards Ed., MIT Press, Cambridge, 1988.