Mémoire de stage de recherche
Master 2 S.A.R. ATIAM
Laboratoire d'accueil : Equipe Perception et Design Sonores, IRCAM
Durée : 5 mars au 3 août 2012
Encadrement : N. Misdariis[1], P. Susini[1], P. Esling[2]
[1] Equipe Perception et Design Sonores, IRCAM
[2] Equipe Représentations Musicales, IRCAM

Etude de la description morphologique des sons environnementaux

Evangelia Koliopoulou

20 août 2012

ATIAM
PARCOURS DU MASTER SCIENCES ET TECHNOLOGIES – UPMC / IRCAM / TELECOM

EQUIPE PERCEPTION ET DESIGN SONORES - IRCAM

So it seems that the world does reveal itself in its sounds, and not only by the way it reflects light [Van79]

Abstract

Ecological perception is concerned with the pickup of information about the environment, or about ourselves in relation to the environment. Organising the myriad of our auditory experiences during a variety of human activities, into clusters of situations that share important features, would help us understand this process.
This report's principal goals are to identify the acoustic features which could unfold the perceptive classification of environmental sounds according to morphological profiles and then to formalise the essential information, in terms of morphological description, conveyed by these clusters.
Relying on a wide range of audio descriptors, our model reveal those who can jointly explain the clusters of morphological profiles studied. Then a correlation analysis between these descriptors and the morphological profiles drawn by the subjects is presented. A statistical shape analysis is also applied to these morphological profiles, which appears adequate to form a prototype for each cluster. We finally present a protocol for a possible perceptual validation of the model suggested.

Résumé

L'approche écologique de la perception concerne l'acquisition d'information sur l'environnement ou sur notre propre relation avec cet environnement. La connaissance des processus d'organisation et de description de l'ensemble des expériences auditives nous permettrait de comprendre les principes sur lesquels est fondée notre perception des sons environnementaux.
Les objectifs principaux de ce rapport sont d'identifier les caractéristiques acoustiques permettant d'expliquer la classification perceptive de sons environnementaux selon des profils morphologiques, puis d'en extraire une formalisation objective permettant de décrire ces classes.
En s'appuyant sur une grande gamme des descripteurs audio, l'étude a permis de révéler ceux qui peuvent conjointement expliquer les différents classes de profils morphologiques étudiées. Ensuite, la corrélation de ces descripteurs significatives avec les profils morphologiques dessinés par les sujets est examinée. Puis, une technique de comparaison de formes sur ces profils morphologiques a été appliquée pouvant ainsi permettre de proposer un formalisme de ces profils. Enfin, nous présentons un protocole pour une éventuelle validation perceptive du modèle obtenu.

Mots clefs

Perception Écologique, Description Audio, Description Morphologique, Profils Morphologiques, Classificateur de sons environnementaux.

Table des matières

Introduction

La notion de la description morphologique des sons a été proposée par Pierre Schaeffer [Sch66]. « Les profils morphologiques ont pour but de décrire de manière pertinente l'évolution de certains paramètres du son au cours du temps et de proposer une structure d'indexation et de classification prenant en compte ces évolutions » [MMHS08].

Smalley presente les composantes principales de la pensée *spectromorphologique* [Sma97] en faisant une projection sur l'espace gestuel. Cette hypothèse, que « *Tout geste effecteur influence directement les caractéristiques du son produit* » [WDR99] permet d'accéder à une description morphologique des sons comme le propose, entre autres, Godøy avec ses « objets sonores-gestuels » [God06].

Cette étude s'inscrit dans le cadre de travail défini par le projet ANR Legos portant sur l'évaluation de l'apprentissage sensori-moteur dans des systèmes interactifs geste-son. Les résultats obtenus sur la formalisation d'une typologie de gestes associés à certains types de sons environnementaux pourront être exploités dans l'une des applications visées par le projet Legos : le Design Sonore Interactif (Sonic interaction Design) qui s'intéresse, entre autres, à l'utilisation du son pour l'amélioration de la manipulation d'objets tangibles ou d'interfaces du quotidien.

D'un point de vue général, on s'intéresse à l'étude de la morphologie d'une classe de sons dits environnementaux, que l'on définit comme : *Tout événement acoustique audible causé par des mouvements dans un environnement humain ordinaire* [Van79]. En effet, des travaux récents ont permis de caractériser certaines dimensions du timbre de différentes classes de sons environnementaux [MMS$^+$10]. Cependant, l'une des propriétés importantes de ces classes de sons provient de leur nature événementielle, c'est-à-dire leur relation à la cause qui a produit le son dont le profil temporel est un élément déterminent permettant de distinguer les différentes classes. L'objet principal de cette étude est alors l'exploration de données à partir de séries temporelles.

Cette problématique s'inscrit dans le prolongement de travaux effectués au sein de deux précédents projets ANR : Ecrins ([Der01], [Rio01]) et SampleORchestrator (SOR) [MMHS08]. Elle porte donc sur la description morphologique des sons environnementaux en considérant les catégories de profils représentatifs de ce type de sons (classification), et leurs représentations graphiques (description) associées aux contours (allure, dynamique) des signaux étudiés. Les derniers résultats obtenus dans ce domaine (projet SOR, [MMHS08]), ont mis en évidence la pertinence de cette démarche dans le cas des profils dynamiques : des classes de morphologies ont été formalisées puis associées à des profil prototypes, symbolisés graphiquement de manière empirique[MMHS10].

Dans le cadre d'une collaboration avec l'equipe RepMus et P. Esling, le paradigme de recherche MOTS (MultiObjective Time Series) sera utilisé pour concentrer le travail sur l'identification de prédicteurs acoustiques pertinents pour expliquer la classification ainsi que les relations pour chacune des six classes révélées lors de l'étude précédente dans le cadre du projet SOR. L'ensemble de sons étudiés provient du corpus des études perceptives précédentes [MMHS08].

Ce sujet de stage vise à proposer une approche en terme de description morphologique et de développer un outil de caractérisation des "objets environnementaux" en termes de description temporelle. Les objectifs du stage sont alors

- expliquer puis modéliser la classification de sons par des critères morphologiques,
- étudier la correlation entre profils perceptifs et descripteurs audio,
- proposer une formalisme de profils morphologiques et
- étudier les propriétés perceptives du modèle obtenu.

Le reste de ce document est organisé de la manière suivante. Nous commençons par présenter les travaux ayant servis de point de départ à ce travail de recherche. Les deux composantes principales sont d'une part le calcul exhaustif de descripteurs audio permettant de trouver ceux qui sont les plus appropriés à une catégorisation des sons environnementaux et d'autre part les études perceptives sur la description morphologique de cette classe de sons. Les sections suivantes s'attachent à développer le travail de recherche effectué. Dans un premier temps, on cherche à trouver la description qui nous permettra d'expliquer les classes proposées dans [MMHS10] (§2). D'autre part, les profils perceptifs sont étudiés et comparés à travers l'évolution temporelle de descripteurs spectraux (§3). Le but de cette analyse de corrélation effectuée sur les profils est de proposer une formalisation, c'est à dire des profils symboliques prototypes représentant chaque classe. Enfin, nous présentons les conclusions et perspectives à long terme qui se dégagent de cette étude (§4).

1. Description Audio - Morphologie - Sons environnementaux

Pierre Schaeffer était le premier à concevoir la notion de description morphologique des sons dans le *Traité des Objets Musicaux* [Sch66]. En s'appuyant sur son approche basée sur une description présentant trois dimensions indépendantes –la cause, la sémantique et la morphologie– Peeters et Deruty ([PD09], [PD08]) présentent une étude sur l'estimation automatique des descripteurs morphologiques pour des profils dynamiques ou mélodique, puis son application à des tâches d'indexation musicale. L'intérêt de cette étude, partie intégrante des projets Ecrins/ SOR ([Der01], [Rio01], [MMHS08]), était de proposer des descripteurs spécifiques, pouvant apporter une description de sons génériques, notamment ceux pour lesquels la relation avec la source ou cause est difficilement identifiable voir inconnue. Les auteurs s'appuient sur une description plus approfondie du contenu sonore. Dans cette étude, les idées de Schaeffer sont combinées avec l'habileté du designer sonore à illustrer cette morphologie des descripteurs. Concernant les profils dynamiques, les classes suggérées sont au nombre de 5 : croissant, décroissant, croissant/décroissant, stable et impulsif. Cette approche permet ainsi d'obtenir une description complémentaire aux résultats expérimentaux purement mathématiques.

D'autre part, des analyses expérimentales ont été effectuées [MMS+10] qui ont également permis d'étendre les connaissances sur la description de sons musicaux à des catégories de sons environnementaux, ainsi non-musicaux. Cette étude comprend des sons produits par des moteurs de voiture, des unités de climatisation, des klaxons et des fermetures de portes de voitures et propose qu'ils s'organisent en 3 méta-catégories perceptives : *motor, instrument-like, impact*. La caractérisation des différentes dimensions du timbre de sons qui appartiennent à une catégorie (définie comme un ensemble homogène et cohérent de similarité), a été faite en s'appuyant sur une technique de mise à l'échelle multidimensionnelle (multidimensional scaling technique - MDS) appliquée aux valeurs de dissimilarité perçues parmi les sons. Enfin, les auteurs proposent un outil de prédiction pour la classification de sons pour les méta-catégories proposées.

Projet SOR : Descripteurs audio de type morphologique pour les sons environnementaux

Dans le cadre de ce projet (SOR) [MMHS10], une étude expérimentale de la perception des profils morphologiques a été effectuée sur un ensemble de 55 sons environnementaux, un sous-ensemble du corpus représentatif formé dans ce même projet. Cette etude avait trois objectifs :

- la définition de classes de profils morphologiques (dynamiques et mélodiques) adaptées à cette catégorie de sons,
- la conception d'un formalisme symbolique pour la description de ces profils,
- l'implémentation d'un modèle de calcul de ces descripteurs temporels.

Au vu de l'ampleur de la tâche proposée, celle-ci a été divisée en deux phases distinctes :

- classification libre et tracé des profils correspondants aux classes identifiées, un moyen de contraindre les 19 participants à effectuer la classification selon des critères morphologiques,
- tracé de profils correspondants aux classes "moyennes" obtenues après avoir analysé les résultats de l'étape précédente (Analyse de cluster présentée en Figure 1.1).

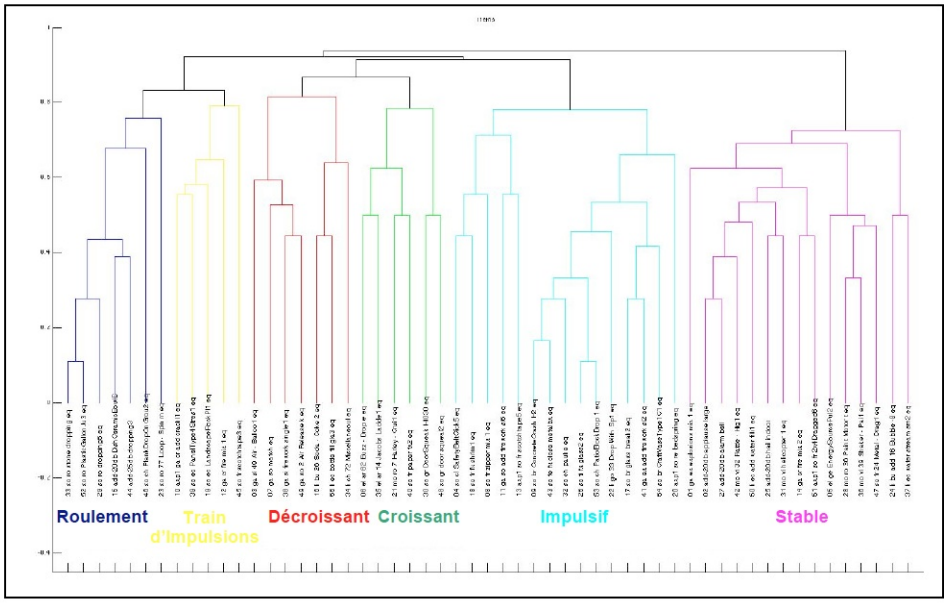

Figure 1.1 – Dendrogram résultant de l'analyse par clustering hiérarchique des resultats de la catégorisation libre ([MMHS10])

Les résultats de cette première phase sur les profils mélodiques s'étant avérés non concluants, notre étude ne se concentre que sur l'analyse de profils dynamiques.

Les résultats sur les profils temporels (classification libre et tracé selon des critères dynamiques) ont permis d'extraire 6 classes dynamiques (*stable, train d'impulsions, décroissant, impulsif, roulement* et *croissant*) (analyse de cluster présentée en figure 1.1). Enfin, ce travail a permis d'aboutir à un formalisme de symboles pour

chacune de ces classes (les profils prototypes) basés sur une observation des tracés individuels (les profils perceptifs) (figure 1.2).

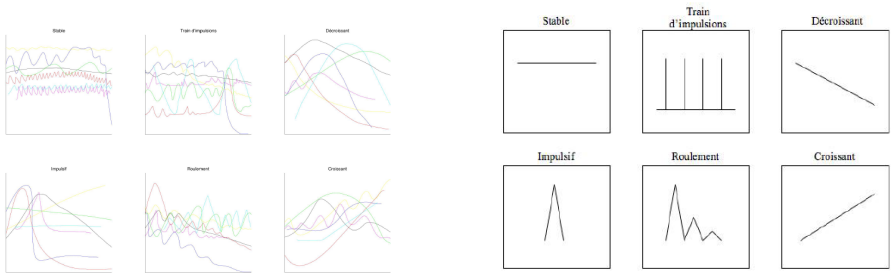

Figure 1.2 – Profils perceptifs tracés par les 19 sujets (gauche) et profils prototypes (droite) [MMHS10].

Il est important de noter que ce passage des profils individuels aux profils prototypes présenté en figure 1.2, c'est à dire l'étape de formalisation symbolique, a été faite manuellement et donc introduisant un biais subjectif. Une des perspectives de notre recherche a donc porté sur la manière d'affiner cette proposition des symboles par analyse mathématique puis de proposer une méthodologie de validation perceptive.

2. Classification

2.1 Description de la base de sons

Six classes moyennes ont été proposées sur un ensemble de 55 sons environnementaux grâce à un processus de classification libre mis en évidence dans [MMHS10]. Nous cherchons à caractériser cette répartition en classes *stable, train d'impulsions, décroissant, impulsif, roulement* à l'aide des connaissances sur l'exploration de données à partir de séries temporelles. Cette recherche s'appuie sur les techniques de représentation, de comparaison et sur les méthodes d'indexation de series temporelles [EA12b].

Inspiré du travail de Esling et Agon [EA12a], on calcule pour chacun des 55 sons, des descripteurs basés sur une transformée de Stockwell, une généralisation de la Transformée en ondelettes continue (figure 2.1) [SML96], et toute la gamme de IRCAMDescriptor, qui comprend différents ensembles de descripteurs sonores (*Energy, Harmonic & Noise, Perceptual, Spectral, Temporal*)[Pee04].

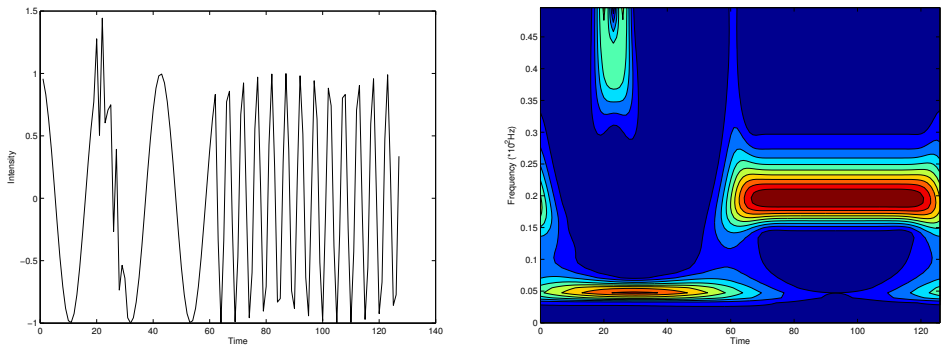

Figure 2.1 – Une série temporelle synthétique constituée d'un signal à large bande (gauche). L'amplitude de la transformée de Stockwell, de la série (droite). On peut remarquer la très bonne résolution simultanément fréquentielle ($f = 5Hz$) et temporelle ($f = 406Hz$).

Ces descripteurs décrivent alors l'évolution temporelle de diverses propriétés acoustiques et perceptives des sons. On calcule ensuite la moyenne et l'écart-type de chaque descripteur. Enfin, les séries temporelles sont ré-échantillonnées à une durée égale puis normalisées par la méthode zero-mean unit-variance. La méthode de représentation SAX (Symbolic Aggregate approximation) [LKLC03] est utilisée, qui est une manière compacte et efficace de stocker l'information temporelle (2.2) en utilisant un alphabet symbolique.

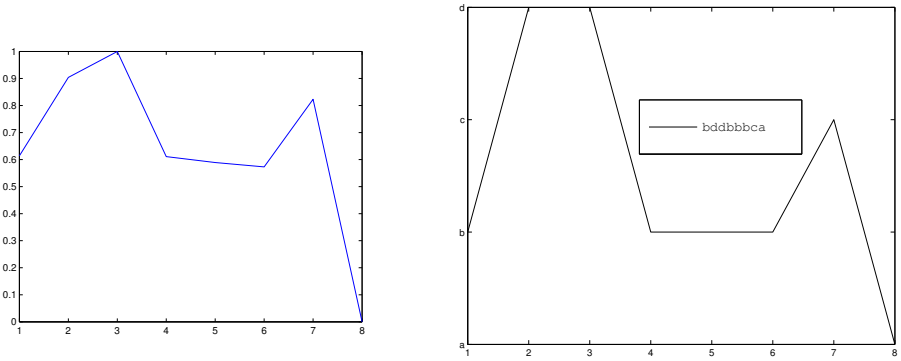

Figure 2.2 – La méthode de représentation SAX (Symbolic Aggregate approximation) appliquée à une série temporelle en utilisant l'alphabète {a, b, c, d}.

2.2 Modélisation

Mesure de similarité

Afin d'effectuer la classification des sons de manière à approcher au maximum la perception humaine, il est indispensable d'établir une notion de similarité entre séries temporelles qui permet d'identifier des objets perceptivement similaires même s'ils ne sont pas mathématiquement identiques. Une mesure de similarité telle que la distance Euclidienne ne permet pas un tel niveau d'abstraction car elle ne prend pas en compte la subtilité dont fait preuve la perception humaine, notamment au niveau des distortions non-linéaires de l'échelle temporelle. Des méthodes plus élaborées sont ainsi utilisées :

- la méthode Dynamic Time Warping (DTW) permet la gestion de distorsions locales de l'axe du temps via une dilatation temporelle non-uniforme ([BC94]). Le principe est donc de permettre la comparaison de points temporels différents, pouvant être contraints à appartenir à une certaine fenêtre de *warping* maximum. La complexité algorithmique de la distance DTW est $O(M * N)$ dans le cas général, M étant la longueur de la série et N celle de la fenêtre de *warping* autorisée. Cependant, en imposant une contrainte sur la fenêtre de warping autorisée, cette complexité peut ainsi être réduite à $O(M)$ [KR05]

- l'algorithme Optimal Subsequence Bijection (OSB) est fondé sur le concept d'appariement élastique des séries temporelles [LJKTM07]. Il permet de trouver la bijection optimale entre deux séries temporelles grâce à des facteurs de *warping* et *skip* (éléments sautés) entraînant une *pénalité* à préciser dans ces cas là. Cet algorithme nous fournit la distance d calculée pour la correspondence choisie mais aussi une valeur de *pathcost* représentant le coût du chemin choisi. Dans le cas où nous ne limitons pas la recherche, l'algoritme est de l'ordre $O(M^2 * N^2)$.

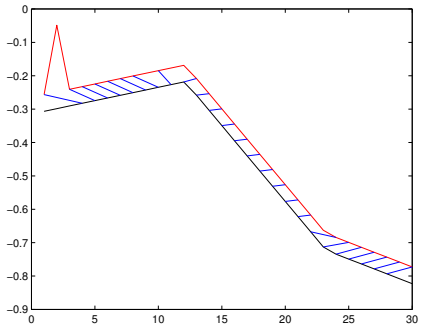

Figure 2.3 – La correspondance obtenue en utilisant l'algorithme OSB entre ces deux séries temporelles (d=0.0042 & pathcost = 0.0017. Dans ce cas on considère comme mesure de decision une pondération entre ces deux paramètres.). Il est intéressant de noter que contrairement à la DTW, la distance OSB permet de sauter des éléments extrêmes (outliers).

Méthodologie d'évaluation

Le but de la classification est de trouver à quelle classe appartient un son donné en entrée. On utilise la méthode *Leave-One-Out* pour laquelle chaque son du jeu de données est consideré comme une entrée que l'on cherche à classifier.

L'exactitude de la classification, pourcentage des sons bien classifiés, est calculée pour chaque descripteur puis pour toutes les combinaisons possibles, suivant ainsi la méthodologie d'évaluation proposée dans [EA12a]. Etant donné que le nombre de combinaisons croît exponentiellement, à partir du niveau 3 on ne garde qu'une partie du nombre total des descripteurs, un compromis entre temps de calcul et exhaustivité d'espace de calcul. Cette méthode de selection itérative nous permet d'obtenir des combinaisons allant jusqu'à 11 descripteurs.

Comme critère de décision on utilise quatre algorithmes différents : 1NN (1-Nearest-Neighbor), 5NN (5-Nearest-Neighbors), MOTS (MultiObjective Time Series matching) et HV-MOTS (HyperVolume MOTS).

Règle des k Plus Proches Voisins (The k Nearest Neighbors rule)

Définition : Étant donnés une série temporelle comme entrée (query) $Q = (q_1, ..., q_n)$, une base de donnés qui contient des séries temporelles DB, une mesure de similarité $\mathcal{D}(Q, T)$ et un nombre entier k, on cherche à trouver les k séries temporelles qui sont les plus similaires à Q. [EA12b]

$$\boxed{\begin{array}{c}\text{Chercher } \mathcal{L} = \{T_i | T_i \in DB\} \text{ tel que } |\mathcal{L}| = k \\ \text{et } \forall T_i \notin \mathcal{L},\ \mathcal{D}(Q, T) \leq \mathcal{D}(Q, T_i)\end{array}}$$

L'approche Multiobjectif

Les algorithmes d'optimisation multiobjective ont été proposés pour la résolution de problèmes impliquant l'optimisation simultanée de nombreux objectifs ([EA12a]) pouvant être parfois contradictoire.

Cette optimisation multiobjective est fondée sur le concept d'optimalité introduit par Vilfredo Pareto. Le front de Pareto $\mathcal{P}$ est defini comme l'ensemble des éléments qui sont les plus efficaces, qui ne sont pas *dominé* dans toutes les dimensions. Plus précisément :

Etant donnés deux points x et y de l'espace de decision S,
une solution y est *Pareto dominé* d'une solution x (noté $x \preceq y$)
ssi elle est dominée dans toutes les dimensions :
$\forall n \in \{1, ..., N\}, f_n(x) \leq f_n(y)$

A partir de cette notion d'optimalité et étant donnés l' *espace de decision* S et un ensemble de fonctions $F = \{f_1, ..., f_N\}$ à minimiser,
on définit l' *espace de critères* $C = \{f_1(x), ..., f_N(x) | x \in S\}$.

La solution optimale est définie comme la recherche des éléments efficaces d'une base de données qui minimisent conjointement un ensemble de mesures de similarité entre des séries temporelles (figure 2.5, points verts) - l'ensemble de ces éléments forment le *front de Pareto* $\mathcal{P}$ (figure 2.5, ligne bleue). La solution à un problème multiobjectif se résume donc à la recherche du front de Pareto $\mathcal{P}$.

Comme critère de selection nous utilisons deux mesures : le nombre d'occurrences de chaque classe dans $\mathcal{P}$ (MOTS). Cependant, le front de Pareto devenant trop inclusif pour un nombre de dimensions élevées. On introduit ainsi l'union des hypervolumes définis par les éléments de $\mathcal{P}$ (HV-MOTS).

L'hypervolume de dominance

En s'appuyant sur la mesure de Lebesgue, une mesure qui étend le concept intuitif de volume à des espaces ayant un nombre de dimensions quelconque , l'hypervolume est alors défini comme :

$\mathcal{H}(B) = \prod_{i=1}^{n}(b_i - a_i)$
définit l'hypervolume de la boîte produite par deux points
$a = (a_1, ..., a_n)$ et $b = (b_1, ..., b_n)$ dans $\mathbb{R}^n$, n $\in \mathbb{N}$

On cherche alors la portion de l'espace dominée par les éléments du front de Pareto $\mathcal{P}$ défini par $p_i \in \mathcal{P}, i \in \mathbb{R}_{\geq 1}$ (figure 2.4) :

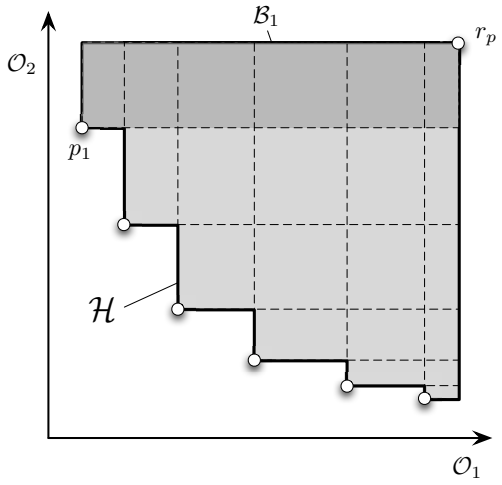

Figure 2.4 – L'hypervolume dominé par $\mathcal{P}$ étant donné un point de référence r_p

Etant donné un point de référence $r_p = (r_p^1, ..., r_p^n)$
$\mathcal{H}(\mathcal{P}) = \bigcup_i \mathcal{H}(B_i) = \bigcup_{p_i \in \mathcal{P}} \left[\prod_{j=1}^{n} (r_p^j - p_i^j) \right]$
définit l'union des hypervolumes dominés par le front de Pareto $\mathcal{P}$

MultiObjective Time Series Matching(MOTS)

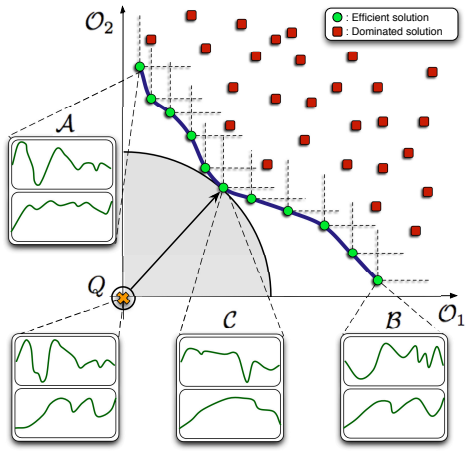

Figure 2.5 – MultiObjective Time Series matching

Le problème MOTS[1] est défini comme

$$\begin{cases} minD_Q^k(S) & k \in \{1 \dots K\} \\ s.t. & S \in DB \end{cases}$$

où Q est la *query* représenté par un ensemble de K séries temporelles, étant les objectifs de la recherche, S est un élément de la base de donnés DB représenté avec

[1] `http://repmus.ircam.fr/esling/mots.html`

un ensemble des séries temporelles correspondant à l'ensemble des objectifs de *query*. $D_Q^k(S)$ correspond à la similarité entre les séries temporelles du k^e objectif, ie. la similarité entre les séries Qk et Sk. Donc, le but est de trouver les éléments les plus similaires qui minimisent conjointement la distance entre les séries des objectifs comparées. La figure 2.5 présente un problème de recherche bi-objectif (O_1 et O_2), de façon à permettre une representation en 2 dimensions. La *query* est à l'origine de l'espace de décision étant donné que ses distances avec elle-même ont des valeurs nulles pour tous les objectifs. La solution A est l'élément le plus similaire pour l'objectif O_1 et B est respectivement l'élément le plus similaire pour O_2. L'élément C serait la solution optimum d'un problème monoobjectif à pondération égale. Nous pouvons observer que aucun des deux objectifs de C n'est très similaire avec les séries de Q, ce qui illustre l'intérêt d'une approche multiobjectif. En effet, cette approche permet des *queries* conjointes sur plusieurs dimensions, sans favoriser l'une d'entre elles. En outre, l'approche multiobjectif est un paradigme approprié lorsque les poids relatifs de chaque objectif recherché ne sont pas connus à l'avance, ce qui est le cas dans les problèmes de perception du timbre.

2.3 Evaluation : Paramètres et Résultats

Nous cherchons à trouver un modèle de prédiction pour la classification des sons environnementaux. Ce modèle nous fournira l'ensemble des descripteurs qui permettent de maximiser le score de classification perceptive 1 mais sur une base purement mathématique. Nous calculons des descripteurs fondés sur la transformée de Stockwell, et les IrcamDescriptors ainsi que différentes modélisations appliquées à ceux-ci pour représenter les sons du corpus. Afin d'expliquer la classification, nous examinons la similarité entre toutes ces représentations, en utilisant des algorithmes de comparaisons des séries temporelles différents (Euclidienne, DTW et OSB). Nous procédons à 6 étapes de calcul qui seront détaillées dans cette section. Dans un premier temps, nous calculons tous les descripteurs présentés, pour des valeurs de ré-échantillonage différents et nous examinons l'influence du paramètre *warping* d'algorithme DTW. Ensuite, nous comparons ces résultats en utilisant la distance donnée par OSB au lieu de DTW. Une version plus récente et riche de IrcamDescriptors est testée par la suite, pour les deux algorithmes de mesure de similarité. En plus, nous procédons à une validation du paramétrage suggéré pendant cette démarche. Enfin, nous présentons les résultats de la recherche de l'ensemble de descripteurs optimal spécifique à chaque classe séparément ainsi qu'aux séries temporelles seules.

Test 1 : Calcul exhaustif

Afin d'expliquer la classification subjective sur une base purement computationnelle, nous commençons par étudier les diverses propriétés acoustiques et perceptives des sons du corpus en question. Nous avons ainsi calculé des descripteurs de formes et de natures différentes. Les sons sont représentés par des descripteurs scalaires, des descripteurs par bande et des descripteurs continus, c'est-à-dire des séries temporelles. Ces formes temporelles ont été modélisées de 4 façons différentes. Les descripteurs fondés sur la transformée de Stockwell, nous fournissent 273 mesures représentant des caractéristiques des sons. Les IrcamDescriptors comprennent une gamme des descripteurs encore plus riche qui s'avère être suffisante pour décrire seuls les dimensions importantes du timbre de ce type de sons. IrcamDescriptors2_0 nous fournissent un ensemble de 408 descripteurs de nature différente (toutes représentations comprises).

Chaque son parmi les 55 est alors caractérisé par un grand nombre de descripteur. Pour chaque combinaison de descripteurs, nous regardons où chaque son se place par rapport aux autres. Le but étant de mettre cet ensemble de similarités en relation avec les données issues de la classification perceptive. Pour mesurer cette similarité entre séries temporelles, dans un premier temps, nous utilisons l'algorithme *DTW*.

Nous considérons alors :

- Descripteurs : Tous (STransform + IrcamDescriptors2_0) et toutes les modélisations (moyenne, écart type, forme originale, normalisée et SAX).
- Measure de similarité : DTW
- Nombre maximum de dimensions : 11 (Niveau 11)
- Nombre de descripteurs pris en compte par rapport au Level :
  si (nombre_de_descripteurs / 2) > (Maximum_dimension + 3) on divise par 4 et met à jour le nombre des descripteurs.
- Dans ce test exhaustif, et donc coûteux en termes de temps de calcul, on elimine le nombre de descripteurs même à partir du niveau 2.

Et nous examinons toutes les combinaisons possibles pour les valeurs suivantes :

- *resampling* : 8, 16, 32, 64, 128, 256
- *sax–alphabet* : 4, 8, 16, 20, 64
- *warping* : 0.01, 0.1, 0.15, 0.2 , 0.25, 0.33

Modèle de prediction / Classificateur

Parmi les méthodologies d'évaluation testées, les résultas obtenus vérifient bien la proposition que la performance de l'algorithme de classification HV-MOTS est supérieure à celle des autres classificateurs. On cherche alors la combinaison de descripteurs qui donne le meilleur score de classification pour HV-MOTS.

Une analyse statistique sur les resultats des 210 configurations de paramètres a été effectuée (*multiway analysis of variance (n-way ANOVA)*), après avoir vérifié que les données suivent une loi normale (figure 2.6). Les facteurs du paramétrage sont les variables indépendantes (internes). Dans cette même étude, on s'est également intéressé à l'influence éventuelle de la nature des features, selon la nomenclature proposée par Peeters dans [Pee04] (variable indépendante notée comme *Param* dans la figure en question). Il apparaît cohérent d'effectuer cette analyse sur les resultats du premier niveau, c'est-à-dire pour chacun des descripteurs indépendamment et ainsi éviter les interférences qui pourraient être introduites par la combinaison avec d'autres descripteurs.

Les résultats de l'analyse sont présentés dans le tableau2.7, où nous voyons qu'il apparaît une différence significative entre les différentes valeurs des facteurs de ré-échantillonnage et warping (figure 2.8).

Il est alors des plus intéressant de concentrer notre attention sur les résultats proposés pour la valeur du ré-échantillonage (figure 2.8). En effet, contrairement aux

résultats intuitifs auxquels on pourrait s'attendre, il semblerait qu'un ré-échantillonage très destructif (dans notre cas 8 points temporels pris sur des séries de plusieurs centaines) et donc réduisant l'information, nous permette une classification bien meilleure. L'algorithme ne se perd pas dans les fluctuations à basse granularité mais se concentre au contraire sur la forme globale. Cela pourrait à première vue sembler aller à l'encontre de notre intuition qu'une meilleure résolution temporelle devait permettre l'accès à une information plus exacte, et donc permettre une meilleure classification. Cependant, ce résultat s'accorde avec le codage de l'information dans notre mémoire. Il renforce également l'intérêt de la classification symbolique et encourage la poursuite d'un formalisation des prototypes décrivants une forme globale. De plus, ce résultat peut également s'interpréter comme une simplification extrême du principe de la DTW. En effet, cette réduction de donnée permet de limiter les distortions temporelles en marginalisant ainsi les fluctuations à court terme.

La valeur du ré-échantillonage suggérée est celle qui donne également le maximum global d'accuracy parmi toutes les configurations pour tous les niveaux.

Concernant la fenêtre de warping qu'on autorise, nous observons qu'une valeur plus petite que 0.25 est bien satisfaisante. Comme la valeur du ré-échantillonage suggérée est 8, la valeur choisie est égale à 0.15, ce qui implique l'utilisation d'un unique point de dilatation autorisé (celui d'avant et d'après). Comme expliqué précédemment, une valeur de ré-échantillonage aussi petite induit naturellement un tel 'warping'.

Parmi les trois modélisations de séries temporelles, la version originale (non normalisée) apparaît être la plus appropriée pour expliquer la classification. Les résultats proposent d'ailleurs d'éliminer la représentation SAX. En outre, la valeur de quantification de la représentation SAX n'influence pas la performance du classificateur (2.7).

Concernant la nature des descripteurs, les résultats font apparaître une forte pertinence des descripteurs du type *Energy*. Parmi ceux sélectionnés automatiquement par le classificateur, les descripteurs *Perceptual* sont les plus importants. Ce résultat reste valide pour le cas du ré-échantillonage à 8 points et ces observations confirment la consistance de la notion de *dynamique* étant donné le contexte de notre recherche.

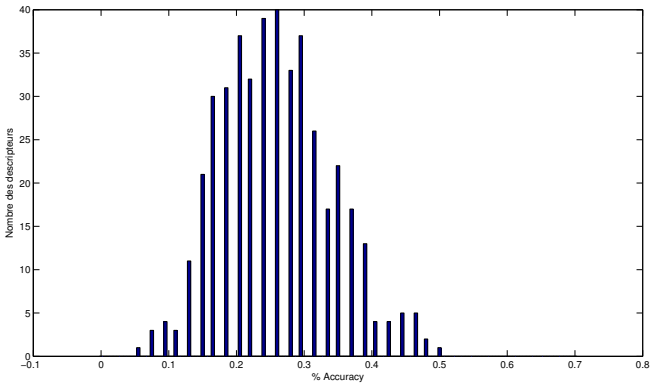

Figure 2.6 – Les donnés de l'une des 210 configurations. Elles suivent bien une loi normale. Les résultats sont similaires pour toutes les autres configurations.

Type III Sums of Squares

Source	df	Sum of Squares	Mean Square	F-Value	P-Value
Resampling	5	1,320	,264	11,211	,0001
DyTimeWarp	6	,375	,062	2,653	,0142
String	4	,072	,018	,760	,5510
Resampling * DyTimeWarp	30	,443	,015	,627	,9437
Resampling * String	20	,165	,008	,351	,9966
DyTimeWarp * String	24	,010	4,021E-4	,017	1,0000
Param	5	31,066	6,213	263,816	,0001
Param * Resampling	25	3,507	,140	5,956	,0001
Param * DyTimeWarp	30	,609	,020	,862	,6828
Param * String	20	,079	,004	,168	1,0000
Residual	26274	618,789	,024		

Dependent: Mesures

Figure 2.7 – Multiway Analysis of Variance de type 'sums of squares'. 3 variables indépendantes (intra) : Resampling (*resampling*), DyTimeWarp (*warping*) et String (*sax*) & 1 variable indépendante (inter) : Param (les descripteurs sont regroupés dans 6 catégories *Energy, Harmonic, Noise, Perceptual, Spectral, Temporal*)

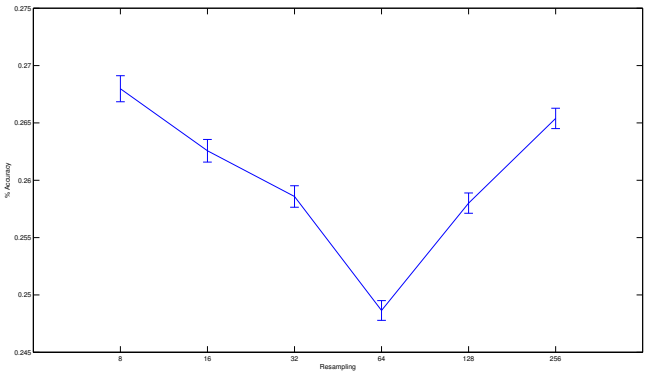

Figure 2.8 – Moyennes et écart-types pour la variable indépendante de*resampling*

Le meilleur pourcentage d'accuracy est obtenu grâce à la combinaison suivante. Cette combinaison de 7 descripteurs vérifie la classification perceptive avec un pourcentage de 76.4%.

Harmonic Spectral Spread Original
Perceptual Tristimulus Original
Relative Specific Loudness Standard Deviation
DMFCC Bands Mean
DDMFCC Bands Standard Deviation
Stransform DMFCC Standard Deviation
Stransform DDMFCC Alt Standard Deviation

Table 2.1 – Meilleur combinaison pour IrcamDescriptors2_0 avec DTW.

où

- Harmonic Spectral Spread : représente l'étalement du spectre autour sa valeur moyenne pour la partie harmonique du signal.

- Perceptual Tristimulus : Trois types de ratio d'energie qui décrivent bien la répartition d'énergie relative au premier partiel du spectre, qui est perceptivement important. Les trois ratios sont définis par :
$T1 = \alpha(1)/\sum_h(\alpha(h)), T2 = (\alpha(2)+\alpha(3)+\alpha(4))/\sum_h(\alpha(h))$ et $T3 = \sum_{h=5:H}(\alpha(h))/\sum_h(\alpha(h))$.

- Relative Specific Loudness : Une courbe d'égalisation du son définie comme la Loudness associée à chaque bande d'une échelle de Bark.

- DMFCC (Delta Mel Frequency Cepstral Coefficients) : La dérivée du premier ordre des MFCC par rapport au temps. MFCC consiste à une description globale du spectre avec peu de coefficients. Le Cepstrum correspond à la Transformée de Fourier (ou la Transformée en Cosinus Discrète) du logarithme du spectre. L'échelle de Mel (une échelle de fréquences construite sur la perception humaine) permet de mieux prendre en compte le filtrage du signal effectué par notre système auditif. Le premier coefficient représente l'energie totale tandis que les 12 coefficients suivants représentent les modulations d'énergie dans différentes bandes.

De plus, les matrices de confusion pour chaque descripteur et chaque combinaison ont été calculées pour expliquer la cohérence des classes. Nous pouvons ainsi observer que malgré une cohérence de classe relativement forte sur l'ensemble des descripteurs, on observe de grandes variations de la pertinence des classes (figure 2.9) en fonction de la combinaison étudiée.

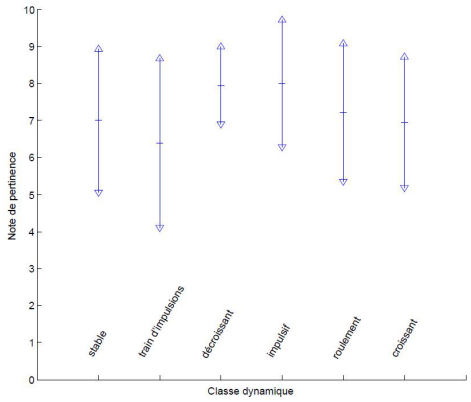

Figure 2.9 – Résultats de 2ème phase de l'expérience : évaluation de la pertinence des classes[MMHS10]

Nous pouvons observer que les classes *impulsif* et *stable* donnent des résultats satisfaisants (figure 2.10) de manière très consistante.

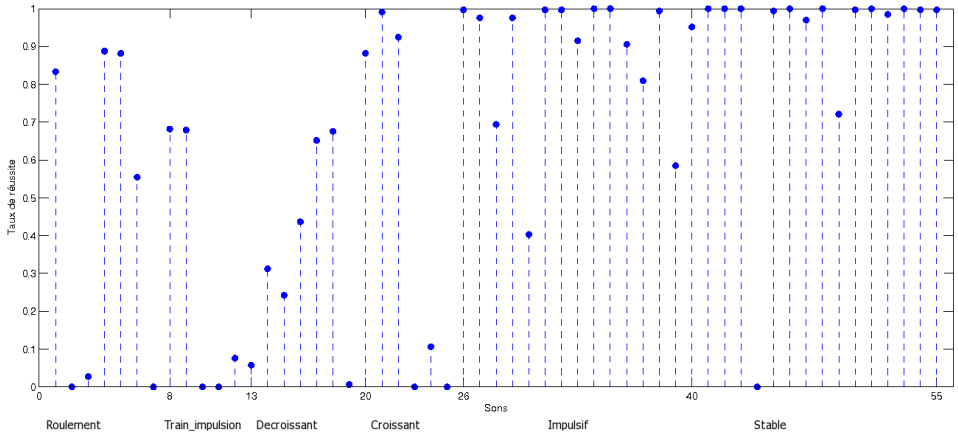

Figure 2.10 – Performance des sons par classe. Testé pour les 11 descripteurs du Level 7 qui combinés, nous donnent le meilleur pourcentage d'accuracy. Les sons apparaissent dans la même ordre que dans l'analyse de cluster 1.1

Test 2 : DTW Vs OSB

Pour comparer les deux mesures de similarité entre séries, nous avons procedé à un test de la même nature avec la valeur de ré-échantillonage à 8 points.
Nous ne restreindrons pas l'algorithme OSB et ainsi pour prendre en compte à la fois la distance et le coût du chemin choisi on définit une pondération :
$OSB = 0.75 * d + 0.25 * pathcost$.
Par contre, on garde tous les descripteurs au niveau 2 et on divise par 4 à partir du Level 3, pour respecter l'effet de la multidimensionalité, comme le temps de calcul n'est pas gênant cette fois-ci.

- Measure de Similarité : $OSB = 0.75 * d + 0.25 * pathcost$

- Nombre de descripteurs pris en compte par rapport au niveau :
  Nous éliminons le nombre de descripteurs qu'à partir du Level 3

Le pourcentage de réussite est exactement le même (76.4%). On arrive à obtenir un ensemble des descripteurs pertinents au Level 6 (8 descripteurs combinés par 6), au lieu du Level 7, du fait qu'on a gardé un nombre de descripteurs plus grand au Level 2.

Harmonic Spectral Spread Original
Perceptual Tristimulus Standard Deviation
Perceptual Spectral Kurtosis Original
Relative Specific Loudness Standard Deviation
DDMFCC Bands Std Dev
Stransform DMFCC Standard Deviation
Stransform DDMFCC Bands Mean
Stransform DDMFCC Alt Standard Deviation

Table 2.2 – Meilleure combinaison pour IrcamDescriptors2_0 avec OSB

où

- Perceptual Spectral Kurtosis : Une mesure permettant de quantifier la présence de pics au sein de la distribution du spectre (un spectre plat donnant une kurtosis nulle)

Test 3 : IrcamDescriptors2_7

Une version plus récente et plus riche de IrcamDescripteurs a été utilisée dans un deuxieme temps. Pour la même configuration que dans le cas précédent,
$\{resampling = 8, warping = 0.15, sax = 4\}$ pour DTW et
$\{resampling = 8$, osb sans restrictions & , mesure de décision $OSB = 0.75 * d + 0.25 * pathcost\}$,
nous calculons toute la gamme de IrcamDescriptors2_7. Cette version nous fournit une description des sons plus riche, car chaque descripteur temporel calculé est accompagné de sa première et deuxième dérivée. Comme l'analyse statistique du premier test a proposé que la représentation SAX n'est pas si significative, on ne la prend pas en compte. Par ailleurs, comme le nombre total de cette version de IrcamDescripteurs est assez grand (702, pour tous les representations) nous procédons à l'omission du calcul des descripteurs fondés sur la transformation de Stockwell,.

La Transformée de Stockwell, consiste à une transformée coûteuse en termes de temps de calcul et qui permet, comparativement aux autres algorithmes, d'obtenir une très bonne résolution fréquentielle aux très basses fréquences, ce qui n'est pas indispensable pour les sons étudiés. Cependant, nous observons que certains descripteurs issus de cette transformée appartiennent à la meilleure combinaison du classificateur de haute performance (87.3% au lieu de 80%), ce qui peut s'expliquer par son excellente résolution fréquentielle. Cependant, il s'avère qu'elle peut ainsi engendrer un effet de sur-entraînement (overtraining). Les descripteurs issus de la transformation de Stockwell, qui font partie de la meilleure combinaison sont sur-adaptés au jeu de

données utilisé ce qui fait que la nature des descripteurs semble peu en rapport avec le contexte de l'experience.

La classification perceptive est ainsi validée avec un pourcentage de réussite de 80% en utilisant n'importe quel algorithme entre les deux.

DTW *Level 4 - 80%*	**OSB** *Level 4 - 80%*
Energy Envelope Effective Duration	Energy Envelope Temporal Centroid
Harmonic Energy OriginalDTW	Loudness DeltaDelta StandardDeviation
Harmonic Spectral Slope	Noisiness OriginalOSB
Perceptual Spectral Variation DeltaDelta Bands StandardDeviation	Perceptual Spectral Roll-Off OriginalOSB
OSB avec STransfrom *Level 7 - 87.3%*	
Energy Envelope Effective Duration	
Energy Envelope Temporal Centroid	
Harmonic Spectral Deviation DeltaDelta	
Loudness DeltaDelta StandardDeviation	
Noisiness OriginalDTW	
STransform Spectral Roughness Mean	
STransform Spectral Skewness Original	

Table 2.3 – Meilleures combinaisons pour IrcamDescriptors2_7 (DTW & OSB)

où

- Energy Envelope Effective Duration : Une mesure qui représente la durée pour laquelle le signal est perceptivement significatif.
- Harmonic Energy : Représente l'energie de la partie harmonique du signal.
- Harmonic Spectral Slope : Mesure de décroissance de l'amplitude du spectre (partie harmonique du signal).
- Perceptual Spectral Variation : Représente le taux de variation du spectre au cours du temps.
- Harmonic Spectral Deviation : Variation de l'amplitude des harmoniques par rapport à l'enveloppe spectrale globale.
- Spectral Roughness : Une mesure de la rugosité du spectre.
- Spectral Skewness : Une mesure de l'asymétrie de la distribution du spectre autour de sa valeur moyenne.

Nous considérons comme meilleure combinaison celle donnée en utilisant l'algorithme OSB car la nature des descripteurs est plus en accord avec la classe des sons testés et avec l'hypothèse de l'experience sur laquelle notre étude a été fondée. Cette

hypothèse induisait les sujets à ce concentrer uniquement sur l'évolution temporelle de la dynamique comme critère de la classification. De plus, le descripteur *Energy Envelope Temporal Centroid* est celui qui donne la meilleure performance au niveau 1, permettant 54,5% d'accuracy sur la classification si on l'utilise seul. D'ailleurs, il s'agit d'un descripteur audio perceptivement important [Pee04]. Il correspond à la moyenne temporelle de l'enveloppe d'energie et permet la distinction entre sons percussifs et soutenus. D'autre part, dans ce cas il apparaît deux descripteurs instantanés qui pourraient nous servir dans la suite (corrélation avec les profils morphologiques §3). Une mesure basée sur le *Loudness* et l'evolution de la *Noisiness* font preuve du critère dynamique proposé pour la classification. Le *Perceptual Spectral Roll-Off* est un descripteur corrélé en quelque sorte avec le « harmonic/noise cutting frequency ».

Pour cette combinaison de 4, dans le même esprit que dans le figure 2.12 nous regardons quels sont les sons qui ne sont pas bien classés.

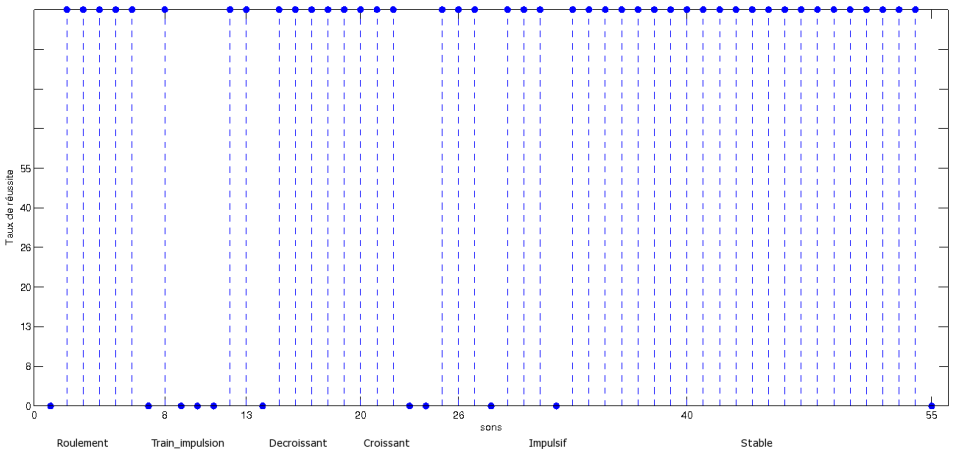

Figure 2.11 – Performance des sons par classe. Testé pour la combinaison de 4 descripteurs pertinents pour le cas de DTW. Les sons sont alors soit bien classés (1) soit mal classés (0). Ils apparaissent dans la même ordre que dans l'analyse de cluster 1.1

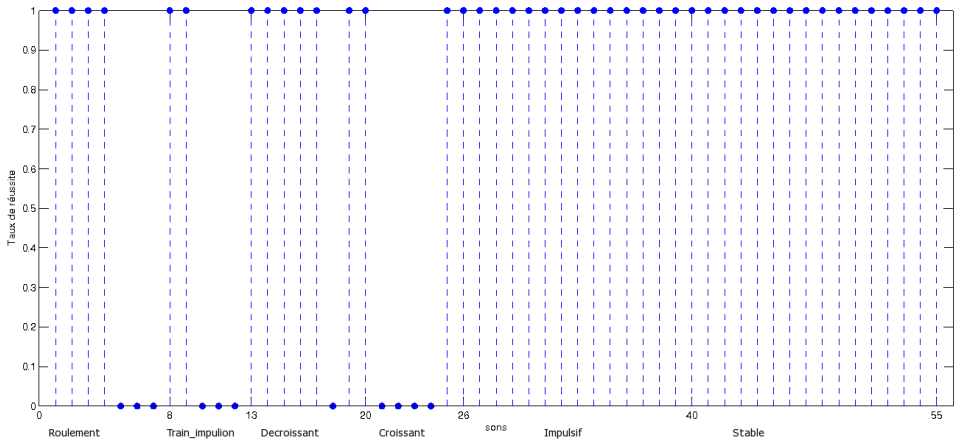

Figure 2.12 – Performance des sons par classe. Testé pour la la combinaison de 4 descripteurs pertinents pour le cas de OSB. Les sons sont alors soit bien classés (1) soit mal classés (0). Ils apparaissent dans la même ordre que dans l'analyse de cluster 1.1

Dans les trois figures 2.10, 2.11 et 2.12 il n'y a pas une forte relation qui pourrait proposer des sons qui sont toujours mal classés, même si le pourcentage d'accuracy est le même pour les deux derniers cas. La performance de sons dépend des descripteurs utilisés pour la classification et il n'est pas facile de tirer une conclusion globale, qui pourrait être mis en relation avec l'analyse de cluster qui a permis d'extraire les 6 classes 1.1. Il n'y a qu'un son de la classe *roulement* qui est mal classé dans l'intégralité des cas testés et en même temps il est le dernier élément lié à sa classe sur le dendrogramme. C'est le son 23 (loop_spin, le 7ième sur les figures), le dernier de la classe *roulement*. Nous pourrions donc considérer que cet élément faisait déjà figure d'outlier lors des analyses de l'étude précédente. Dans le figure 2.13 nous présentons les 8 points qui décrivent l'évolution temporelle des deux descripteurs instantanés de la meilleur combinaison pour cette classe. Selon l'algorithme, le son 23 appartient à la classe stable, ce qui semble cohérent si on regarde les formes globales des deux descripteurs de ce son.

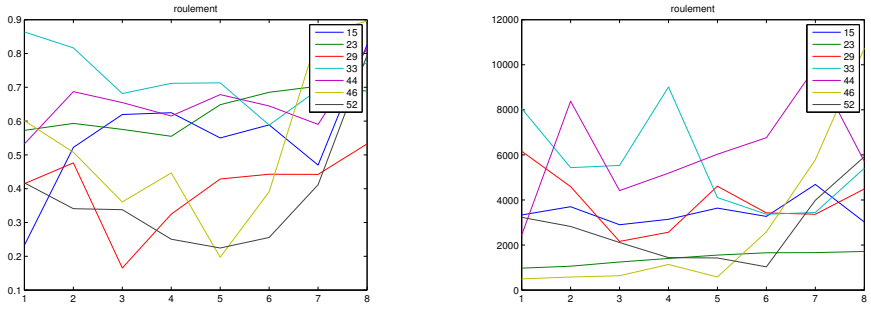

Figure 2.13 – *Noisiness Original* (gauche) et *Perceptual Spectral Roll-Off Original* (droit) pour les sons de la classe *roulement*.

Pour les autres sons toujours mal classés - 12 (10ième) et 19 (11ième) de la classe *train_impulsion* et 40 (23ième) de la classe *croissant* - ce sont plutôt leurs valeurs moyennes qui les différencient et non leur évolution temporelle. Cet observation peut suggérer que ces deux classes sont les moins cohérentes au niveau de la classification perceptive.

Concernant le paramétrage de la distance OSB, nous n'avons pas posé de restrictions aux points qui peuvent être choisis ou sautés. Cette liberté nous donne des valeurs de distances (d) assez petites et par conséquent des valeurs de coût de chemin assez importantes. Cette réflexion nous a mené à examiner la possibilité de re-définir la distance que nous avons considérée comme mesure de decision. On définit alors $OSB = 0.25 * d + 0.75 * pathcost$.

Cette modification n'impacte que faiblement les résultats qui se caractérise par une richesse et une diversité plus importante lors du choix des descripteurs. Au niveau 4 on obtient 2 descripteurs en plus (on a plus d'une combinaison cette fois-ci) et au niveau d'après on obtient un pourcentage d'accuracy légèrement amélioré :

Level 4 - 80%	*Level 5 - 81.81%*
Energy Envelope Temporal Centroid	Energy Envelope Temporal Centroid
Loudness Delta StandardDeviation	Loudness Delta StandardDeviation
Loudness DeltaDelta StandardDeviation	Loudness Delta Original
Noisiness OriginalOSB	MFCC
Perceptual Spectral Roll-Off OriginalOSB	Noisiness OriginalOSB
Sharpness Mean	

Table 2.4 – Les combinaisons de Level 4 et 5 pour IrcamDescriptors2_7 avec OSB ($OSB = 0.25 * d + 0.75 * pathcost$)

où
La Sharpness correspond à l'équivalent perceptif du centroid spectral, calculé en utilisant le Specific Loudness des bandes de Bark.

La performance des sons et des classes est identique à celle présentée précédemment 2.3.

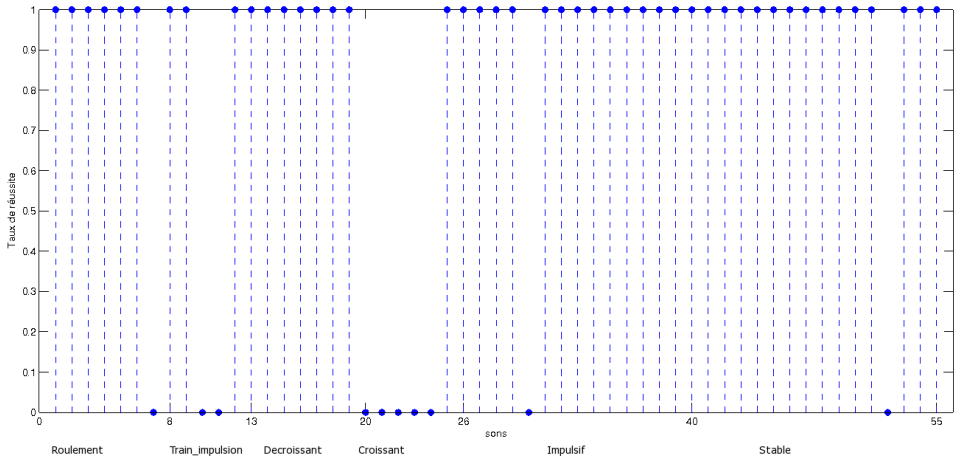

Figure 2.14 – Performance des sons par classe. Testée pour la combinaison de 5 descripteurs pertinents pour le cas de $OSB = 0.25 * d + 0.75 * pathcost$. Les sons sont alors soit bien classés (1) soit mal classés (0). Ils apparaissent dans la même ordre que dans l'analyse de cluster 1.1

Test 4 : Observation de variabilité du paramétrage

Nous présentons dans cette section l'impact du choix de la valeur des différents paramètres sur les résultats de classification pour IrcamDescriptors2_7. Nous procédons à un calcul en faisant varier la valeur de *resampling* comme dans le Test 1 pour deux valeurs de *warping* qui ont donné des resultats significativement différents dans cette étude-là.

- *resampling* : 8, 16, 32, 64, 128, 256
- *warping* : 0.15, 0.33

Les deux premières figures (2.15 et 2.16) ont été générés en prenant en compte tous les descripteurs qui ne sont pas de forme scalaire, du même manière que dans l'analyse des resultat du Test 1. Cependant, la performance des dérivées des descripteurs qui sont calculés en plus dans le cas de IrcamDescriptors2_7 fait que les resultats ne sont pas cohérents avec ceux du Test 1. C'est plutôt leur performance que nous observons dans ces deux premières figures. La dernière figure (2.17) a été générée en excluant les dérivées, mais pour des combinaisons de descripteurs deux par deux, où nous voyons bien la cohérence avec les resultats de la première analyse.

Test 5 : Meilleure Combinaison par Classe

Dans les trois figures 2.10, 2.12 et 2.11 présentés précédemment, nous voyons bien que le classificateur ne nous donne pas une performance du même niveau pour toutes les classes, comme proposé dans l'étude de départ 2.9. Ainsi, nous avons procédé également à la recherche des combinaisons de descripteurs qui peuvent représenter la cohérence dans une même classe. Nous avons effectué deux calculs en utilisant les

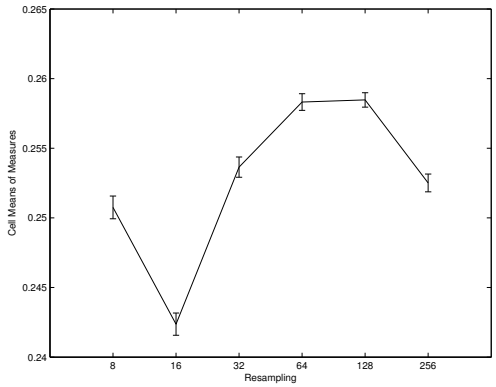

Figure 2.15 – Moyennes et écart-types pour la variable indépendante *resampling* sur les resultats du niveau 1.

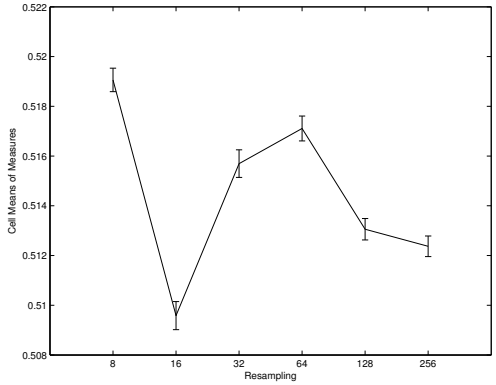

Figure 2.16 – Moyennes et écart-types pour la variable indépendante *resampling* sur les resultats du niveau 2.

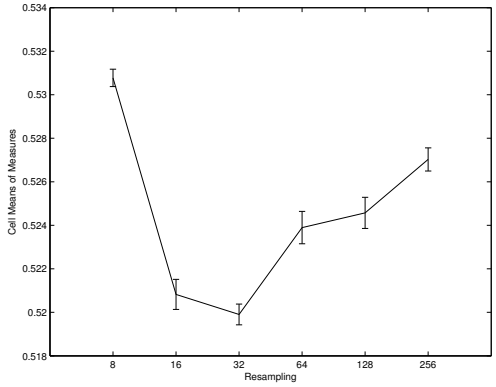

Figure 2.17 – Moyennes et écart-types pour la variable indépendante *resampling* sur les resultats du niveau 2 sans prendre en compte les dérivées de descripteurs.

deux algorithmes étudiés, DTW[2] et OSB[3]. Les resultats obtenus suggèrent qu'il existe des combinaisons qui font bien preuve de la cohérence intra classe (pourcentage de réussite 100%), mais ces descripteurs sont tous différents (Les résultats sont donnés en Annexe A). Parmi ces descripteurs, la *Loudness* est le descripteur le plus fréquemment sélectionné. Ses modélisations différentes apparaissent dans toutes les combinaisons, sauf pour les classes *impulsif* et *roulement*).

D'un point de vue plus général, les descripteurs regroupés par type sont :

[2] *resampling* = 8 & *warping*=0.15
[3] *resampling*=8 & *OSB=.75*pathcost+.25*d*

Descripteur	Nature	Classes
Harmonic Energy	Energy	croissant
Noise Energy	Energy	croissant, train impulsion
Total Energy	Energy	train impulsion
MFCC	Global Spectral	croissant, roulement, train impulsion
Energy Envelope Effective Duration	Global Temporal	croissant, impulsif
Energy Envelope Temporal Centroid	Global Temporal	décroissant, impulsif
Energy Envelope Temporal Increase	Global Temporal	croissant
Chroma	Harmonic	croissant
Harmonic Odd To Even Ratio	Harmonic	impulsif
Harmonic Spectral Centroid	Harmonic	croissant, impulsif
Harmonic Spectral Kurtosis	Harmonic	croissant, train impulsion
Harmonic Spectral Rolloff	Harmonic	décroissant, roulement
Harmonic Spectral Skewness	Harmonic	train impulsion
Harmonic Spectral Slope	Harmonic	croissant
Harmonic Spectral Spread	Harmonic	croissant, roulement
Harmonic Spectral Variation	Harmonic	croissant
Harmonic Tristimulus	Harmonic	croissant, décroissant
Inharmonicity	Harmonic	décroissant
Noisiness	Harmonic	croissant, impulsif
AutoCorrelation	Instantaneous Temporal	croissant
Signal Zero Crossing Rate	Instantaneous Temporal	croissant, train impulsion
Loudness	Perceptual	croissant, décroissant, stable, train impulsion
Perceptual Odd To Even Ratio	Perceptual	décroissant
Perceptual Spectral Centroid	Perceptual	train impulsion
Perceptual Spectral Deviation	Perceptual	décroissant
Perceptual Spectral Rolloff	Perceptual	stable
Perceptual Spectral Skewness	Perceptual	impulsif
Perceptual Spectral Slope	Perceptual	train impulsion
Perceptual Spectral Spread	Perceptual	décroissant
Perceptual Spectral Variation	Perceptual	décroissant, train impulsion
Perceptual Tristimulus	Perceptual	décroissant, train impulsion
Relative Specific Loudness	Perceptual	décroissant
Sharpness	Perceptual	roulement
Spread	Perceptual	impulsif
Spectral Centroid	Spectral	train impulsion
Spectral Decrease	Spectral	impulsif, stable
Spectral Rolloff	Spectral	impulsif, stable
Spectral Slope	Spectral	train impulsion
Spectral Variation	Spectral	train impulsion

Table 2.5 – Analyse synthétique des résultats du calcul par classe

Test 6 : Meilleure combinaison de séries temporelles

Dans le chapitre qui suit, on s'intéressera à étudier les profils perceptifs et à les comparer avec les descripteurs significatifs sélectionnés par analyse automatique. Dans cet esprit, nous avons éliminé les descripteurs scalaires et nous avons enfin procédé au calcul de la combinaison des descripteurs de forme des séries temporelles et ainsi permettre une comparaison directe avec les profils 3.1.

Level 3 DTW - 60%	*Level 5 OSB - 69%*
Harmonic Energy OriginalDTW	Energy Envelope Original
Fundamental Frequency DeltaDeltaOriginal	Fundamental Frequency OriginalOSB
Loudness	LoudnessDelta
Noise Energy OriginalDTW	Loudness DeltaDelta Original
Signal Zero Crossing Rate DeltaDeltaOriginal	Signal Zero Crossing Rate DeltaDelta
Total Energy Orignal	
Total Energy OrignalDTW	
Total Energy DeltaDelta OrignalDTW	

Table 2.6 – Les meilleures combinaisons pour des séries temporelles seules pour les deux méthodes : DTW (*resampling* = 8 & *warping*=0.15) et OSB (*resampling*=8 & OSB=.75*pathcost+.25*d).

où

- L'Envelope d'energie se trouve à la base du calcul des descripteurs globaux (scalaires).
- La Fréquence Fondamentale est un descripteur très significatif pour des signaux harmoniques.
- Le Signal Zero Crossing Rate correspond au nombre des fois que le signal traverse l'axe du zero. Sa valeur augmente quand le signal est bruité.

2.4 En résumé

Aux premièrs resultats du Test 1, notre attention a été attirée par le fait qu'une valeur de ré-échantillonage très faible augmente la performance du classifieur. Cette forme de moyennage temporel s'accorde avec une abstraction qui permettrait de se focaliser sur l'évolution globale qu'on observe plutôt que d'apporter une signification aux fluctuations à court terme. Ce résultat encourage une association entre les résultats computationnels et perceptifs, ce qui constitue le chapitre qui suit.

Concernant les descripteurs pertinents, la meilleure combinaison obtenue est celle présentée dans le Test 3 pour IrcamDescriptors2_7 et pour l'algorithme OSB. L'étape suivante d'analyse serait la recherche du paramétrage optimum de la distance OSB et de la meilleure métrique définie (pondération parmi les valeurs données par l'algorithme.)

Avant de passer au chapitre suivant, nous ré-examinons les resultats en ne s'intéressant qu'aux séries temporelles. Cette analyse nous permet d'extraire les descripteurs les plus pertinents qui nous serviront de référence pour une corrélation perceptive. Les *Noisiness Original* et *Perceptual Spectral Roll-Off Original* sont les deux séries temporelles qui font partie de la meilleure combinaison. Parmi les descripteurs proposés dans le dernier calcul pour les séries temporelles seules, le *Loudness Delta* (non-normalisé) est également celui avec le meilleur pourcentage de réussite (50%) pour le niveau 1. Parmi les autres, le *Energy Envelope Original* a aussi une bonne performance (44%)

au niveau 1. Dans ce même cas et avec la même performance apparaît aussi le *Inharmonicity Delta Delta*. Cet ensemble de descripteurs est donc celui qui est retenu pour étudier la corrélation perceptive avec les profils utilisateurs.

3. Profils morphologiques

3.1 Objectif

Un des trois objectifs du projet SOR 1 était la conception d'un formalisme pour la description de profils morphologiques proposés pour les sons environnementaux, et ce sous forme symbolique. Pendant la deuxième phase de l'expérience effectuée, il a ainsi été demandé aux 19 sujets de *dessiner les classes* suggérées selon l'analyse de clusters. Des profils prototypes ont été proposés de manière purement subjective. 1.2

Dans cette section, nous nous intéressons à definir un symbolisme qui soit moins arbitraire et qui pourrait être associé en quelque sorte avec les descripteurs temporels. Par exemple, cette étude pourrait proposer des stratégies différents, c'est à dire une tendance d'un groupe des sujets à 'suivre' un descripteur quelconque, et donc de présenter une forme de corrélation dans son appréciation de la morphologie du son.

3.2 Prétraitement des données

La tâche de 'dessiner les sons' est effectuée au moyen d'une interface temps-réel (Max/MSP) reliée à une tablette WACOM permettant de jouer les sons et recueillir les données morphologiques. Ces données comprennent alors les coordonnés (p, a) où p correspond au pas effectué dans un écart temporel donné et a à l'amplitude correspondante à cette position p (l'échantillonnage effectué étant non-lineaire). Seule une partie de la tablette a été prise en compte pour définir les *profils perceptifs*.

À cause d'un changement d'échelle effectué dans cette même étude, nous observons des multiples mesures du même p. Dans ce cas nous stockons que la dernière valeur de a. Puis, nous procédons à une interpolation (linéaire) pour tous les p de la partie autorisée, ce qui nous donne 992 points.

Ensuite, nous observons qu'un sous-échantillonnage de 128 points est satisfaisant pour conserver la forme globale des profils pour tous les sujets. Ce analyse a été fait pour faciliter la comparaison avec les descripteurs instantanés - 8 points puis interpolés. Nous comparons ainsi la régularité parmi les séries originales et ceux après le sous-échantillonnage pour des valeurs différentes. Nous nous appuyons sur l' *entropie approximée*, un outil de statistiques de régularité proposé pour l'analyse des données médicales, telles que la fréquence cardiaque, et qui s'est ensuite étendue aux autres applications (finance, psychologie).

Les résultats de la comparaison ont été moyennés par classe, puis par valeur testée :

Sous-échantillonage	32	64	128	256	512
Variation d' entropie approximée	10.8	6.1	3.4	2.3	2.7

Table 3.1 – Variation d' *entropie approximée* pour des valeurs de sous-échantillonnage données

3.3 Formalisme et Corrélation

Une fois que nous avons terminé cette étape de pré-traitement des séries temporelles sur les profils perceptifs dessinés, nous nous intéressons d'une part à proposer un formalisme définissant chaque classe et d'autre part à associer ces formes aves les descripteurs temporels issus de l'étude présentée précédemment.

Dendrogramme

Dans cette partie nous avons procédé à une analyse par clustering hiérarchique d'une part des 6 profils perceptifs des 19 participants décrivant les classes et d'autre part des descripteurs instantanés pertinents proposés pour les 55 sons répartis en ces classes. L'intérêt d'une telle analyse réside dans le fait qu'elle nous permettrait de vérifier la cohérence entre les données perceptives obtenues par classification libre (Figure 1.1) et le modèle mathématique de classification que nous avons employé. En effet, l'obtention d'un clustering similaire à celui obtenu lors de l'expérience précédente (ou tout du moins présentant de fortes corrélations), nous permettrait ainsi de valider le modèle obtenu.

Pour les profils dessinés, les résultats ne se sont pas avérés très concluants. Cependant (comme nous le détaillerons dans le paragraphe suivant), les profils perceptifs ne font pas preuve d'une forte corrélation intra-classe qui peut être observée par la présence d'un grand nombre d'outliers. Notamment, une variabilité très importante dans les catégories de formes influencent ainsi fortement les résultats.

En ce qui concerne les descripteurs, l'analyse des résultats se révèle des plus intéressante. Parmi les six descripteurs que nous considérons comme significatifs 2.4 obtenus par classification automatique lors de l'étape précédente, nous présentons le dendrogramme généré pour les formes définies par *Loudness Delta* dans la Figure 3.1

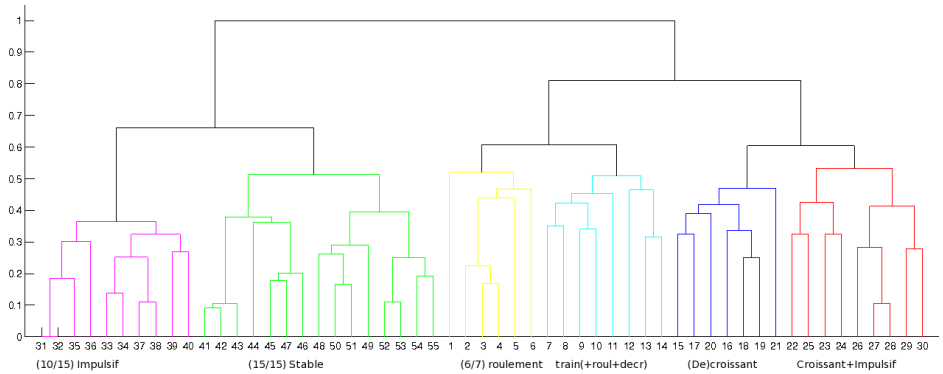

Figure 3.1 – L'analyse en cluster pour la dérivée de Loudness pour tous les sons du corpus. Nous voyons une répartition qui est en accord avec les résultats sur la performance des sons présenté au deuxième chapitre (le son 7 qui appartient à la classe *roulement* - et les 10, 11 - sont toujours mal classés).

Pour la génération de ce clustering hiérarchique, la comparaisons des formes du descripteur ont été calculé par la distance Procustéenne (présentée dans le paragraphe suivant). La matrice de distance est alors regroupée suivant une moyenne entre éléments. Nous utilisons ainsi la même méthode d'analyse que l'étude précédente.

Nous pouvons observer sur cette figure, une très forte corrélation avec les résultats perceptifs obtenus précédemment. En outre, les sons de la classe stable sont tous regroupés correctement. Les classes roulement et impulsif suivent également un regroupement proche de celui effectué de manière perceptive. Enfin, malgré une plus grande confusion dans les autres classes, nous observons tout de même une relative homogénéité dans la création des clusters.

L'analyse Procustéenne et sa généralisation

En statistiques, l'analyse procustéenne est une technique pour comparer des formes[1].

Pour comparer les formes de deux ou plusieurs objets de façon optimale, les objets doivent être d'abord superposés. La superposition Procustéenne est effectuée par transformation, rotation et mis à l'échelle optimale des objets. L'objectif est d'obtenir des placements et tailles semblables, en minimisant une mesure de différence de forme entre les objets, la distance Procustéenne.

La Generalized Procrustes Analysis (GPA) s'appuie sur la méthode d'analyse Procustéenne afin de superposer un ensemble de formes au lieu d'une paire de formes.

L'algorithme est le suivant :
1. Choisir une référence (parmi l'ensemble d'étude, ou définie arbitrairement)

[1] Cette technique a été nommée à partir de Procuste (Προκρούστης), un bandit de la mythologie grecque qui forçait ses victimes à s'allonger sur un lit et modifiait violemment leur taille pour que celle-ci correspondent à la taille du lit.

2. Utiliser l'analyse procustéenne pour superposer chaque forme à la référence
3. Calculer la moyenne de cet ensemble après les superpositions
4. Si l'écart entre la moyenne et la référence est plus grand qu'une certaine valeur, définir comme référence la moyenne et retourner à l'étape 2.

Nous avons ainsi appliqué cet algorithme à l'ensemble des profils dessinés en considérant chaque classe indépendamment. Notre objectif dans cette analyse est d'obtenir des formes prototypiques (semblable à celles choisies précédemment de manière arbitraire) qui pourraient représenter ces classes, mais suivant cette fois-ci une analyse purement pragmatique des profils dessinés. Nous définissons ensuite deux manières de choisir la forme de référence de l'analyse. La première utilise les profils du même participant pour toutes les classes. La deuxième méthode calcule la "forme centroid" (en considérant les séries temporelles comme une distribution de points bi-dimensionnels). Nous comparons ces resultats avec ceux donnés après avoir adapté à notre contexte l'algorithme GPA proposé par Simon Preston[2] pour l'analyse des formes en 3 dimensions.

[2] `http://www.maths.nottingham.ac.uk/personal/pmzspp`

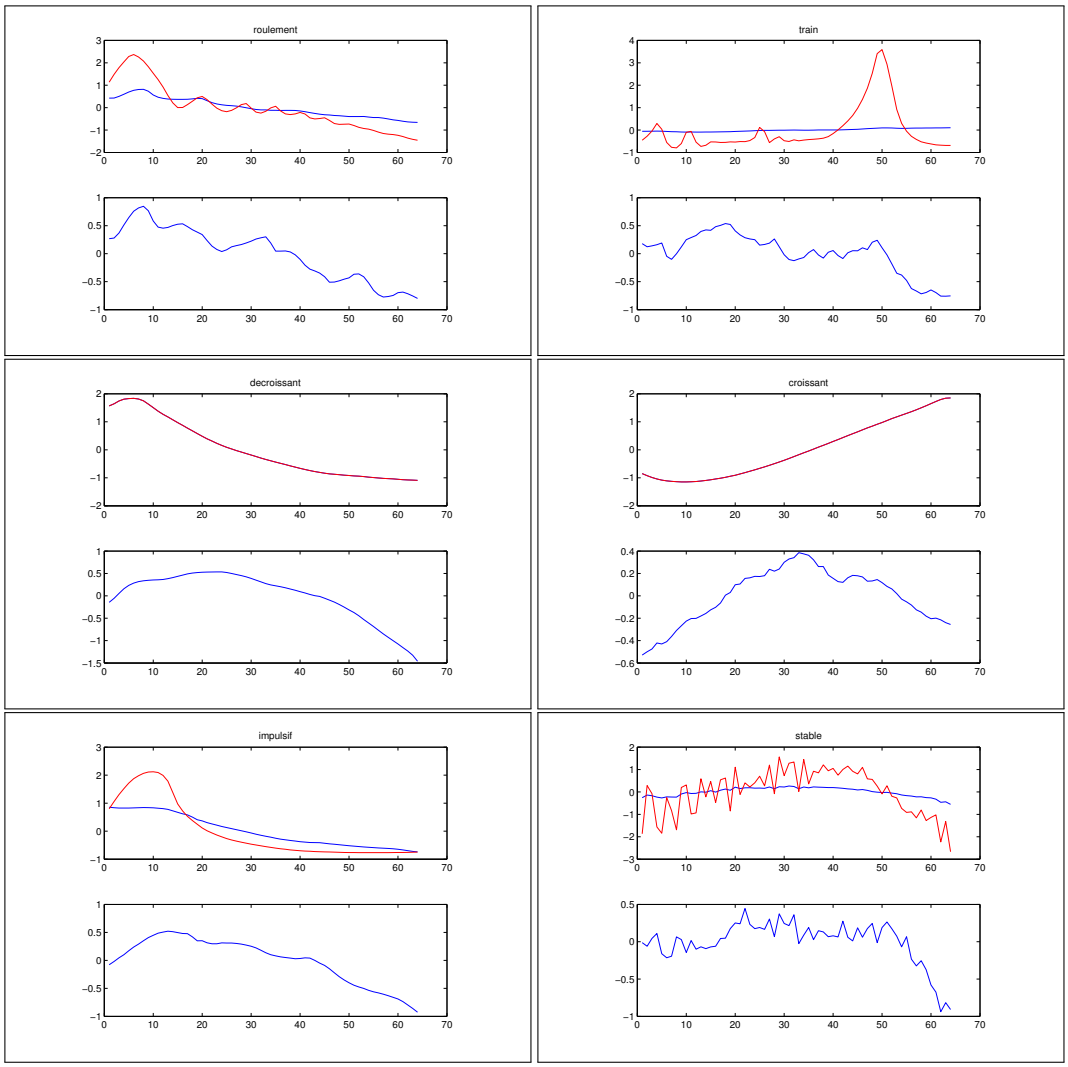

Figure 3.2 – La formalisation proposée par l'analyse Procustéenne généralisée (blue), étant donnée comme référence la forme en rouge, suivie par celle obtenue en utilisant l'algorithme de Preston.

Nous observons que la formalisation proposée est fortement dépendante de la série sélectionnée comme référence. Il est possible d'analyser ce résultat de plusieurs manières. Tout d'abord, la corrélation intra-classe au niveau des profils dessinés présente de grande disparités, comme l'a suggéré le resultat de l'analyse en cluster effectuée précédemment. D'autre part, il semblerait que les profils de certains sujets soient radicalement opposés à la nature intuitive de la classe. Ainsi ces outliers observés (Figure 1.2), provoque une incohérence dans la distribution des données qui modifient fortement les profils obtenus. Cette observation, est des plus frappantes en ce qui concerne la classe *impulsif* pour l'analyse par référence et la classe *croissant* pour l'analyse par centroid. Nous observons cependant, des résultats très cohérents pour les classes de *roulement* et *décroissant*. Les resultats montrent que dans le cas d'algorithme présenté, il faudra définir une référence moins arbitrairement. Dans le

cas de celui de S. Preston les resultats seront assurément meilleurs si nous ne prenons pas en compte les formes atypiques (outliers).

Corrélation

Afin d'explorer la correlation parmi les descripteurs instantanés pertinents et les profils perceptifs nous procédons à une analyse multivariée (ordination). Pour ce faire, nous utilisons une méthode de calcul en composantes principales permettant également d'expliquer un ensemble des variables dépendantes par la variance d'un ensemble des variables indépendantes. Nous effectuons ainsi une Analyse de Redondance (RDA), en utilisant le Fathom_Toolbox pour Matlab. Ce Toolbox comprend un ensemble des fonctions statistiques pour l'analyse multivariée des données écologiques[3].

Nous avons effectué cette analyse en utilisant les 3 distances différentes : DTW, OSB et Procrustes. Dans la figure qui suit nous présentons le meilleur analyse obtenue pour le cas de Procrustes.

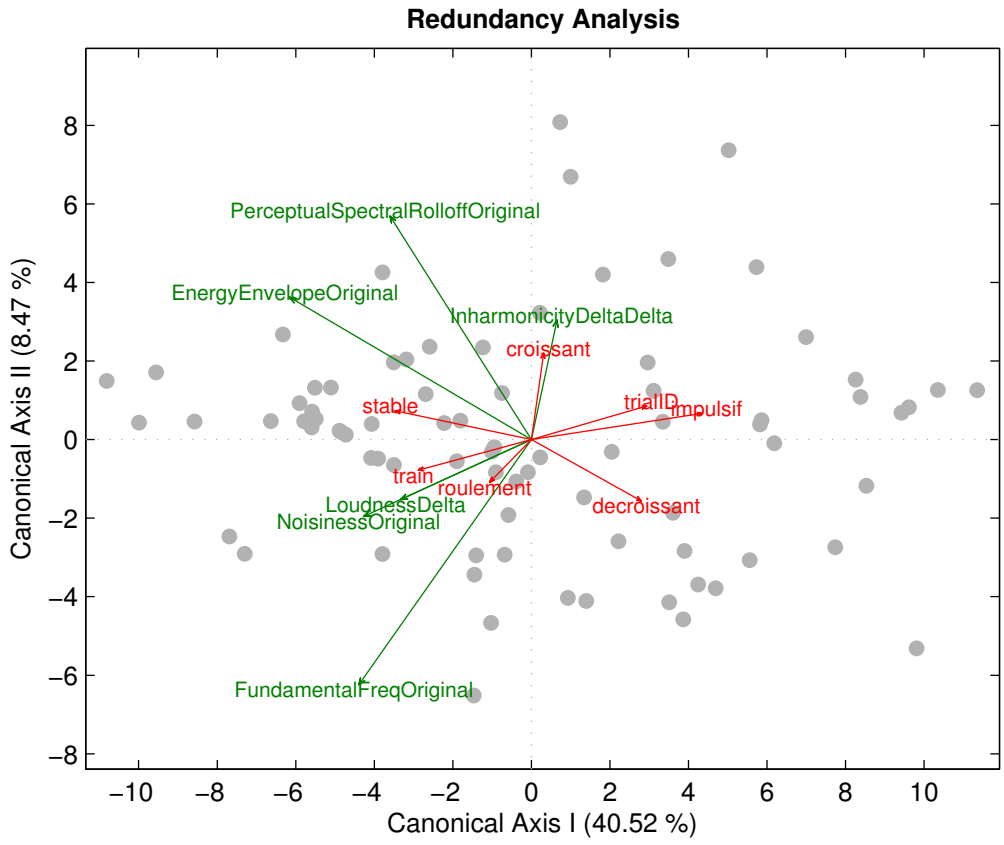

Figure 3.3 – Analyse de Redondance des descripteurs instantanés pertinents et des profils perceptifs regroupé par classe.

[3] `http://seas.marine.usf.edu/~djones/matlab/matlab.html`

L'axe principal est expliqué par des descripteurs de type Energétique/Perceptif et associé aux profils des classes *stable* et *impulsif*. Cela semble logique si nous considérons le fait que la mémoire favorise les événements bien structurés. Ce sont aussi les deux classes qui se sont avérés les plus cohérentes dans l'étude du chapitre précédent.

Les profils dessinés pour les sons de la classe *croissant* sont associés aux descripteurs de type plutôt Harmonique. Cette observation pourrait expliquer certaines incohérences dans les analyses précédentes. En effet, malgré les instructions données aux sujets de se focaliser sur l'évolution de la dynamique des sons, il est possible qu'un biais harmonique ai pu les influencer.

Les profils des classes *train impulsion* et *roulement* sont très proches et on remarque une forte corrélation avec la première dérivée de la Loudness, ce qui est cohérent avec nos analyses précédentes.

Les formes atypiques (outliers) sont détectées en amont de cette analyse et ne sont ainsi pas prises en compte. Ce pré-traitement amène deux sujets à être intégralement exclus et certain profils d'autres participants ont également été rejetés.

3.4 En résumé

L'Analyse Procustéenne semble appropriée pour expliquer les profils perceptifs. Une fois que nous avons détecté les formes atypiques dans l'analyse multivariée présentée, nous pouvons revenir en arrière et aller plus loin dans l'analyse de la formalisation des profils perceptifs.

4. Perspectives

Une deuxième experience du même type permettra une validation perceptive des résultats présentés. L'affinement des formes obtenues par le GPA pourrait constituer des identificateurs des 6 classes. L'idée serait alors de demander aux sujets d'associer les sons d'un nouveau corpus à ces 6 catégories des formes. Pour mettre en valeur les resultats expérimentaux, il sera alors primordial que les sons aient la même durée et que le pré-traitement s'attache à une synchronisation temporelle parfaite (les événements sonores arrivent en même temps).

D'autre part, ce serait intéressant d'explorer l'association parmi les ensembles multivariés des fonctions représentant chaque son ([HJTW12]). Ainsi relier la corrélation entre un grand nombre de descripteurs ayant la même évolution pour un son donné, avec le renforcement induit sur la perception morphologique (classification libre). L'hypothèse étant qu'une évolution conjointe pourrait affecter les caractéristiques morphologiques. Cette corrélation mutli-ensemble pourrait nous servir également à l'association des profils perspectifs aux descripteurs instantanés. Si ces descripteurs font partie des descripteurs pertinents proposés, cela serait une validation computationnelle des résultats présentés dans cette étude.

Dans un deuxième temps, l'étude exactement opposée de sons ayant une décorrélation maximum de leur descripteurs permettrait d'évaluer l'influence relative et la possible existence d'une pondération subjective dans la perception de la morphologie.

Ce serait aussi intéressant de comparer les résultats avec ceux d'une experience effectuée en ayant comme sujets des enfants. Leur intelligence intuitive pourrait nous fournir des résultats plus éclairés sour l'acquisition d'information sur l'environnement.

A. Meilleure Combinaison par Classe

croissant *Level 3 -100%*	**decroissant** *Level 2 -100%*	**impulsif** *Level 2 - 100%*
DTW		
Autocorrelation Delta Bands StandardDeviation	Energy Envelope Temporal Centroid	Energy Envelope Effective Duration
Chroma Delta Bands StandardDeviation	Fundamental Frequency OriginalDTW	
Energy Envelope Effective Duration	Fundamental Frequency Delta OriginalDTW	Harmonic Energy Delta Mean
Energy Envelope Temporal Increase	Harmonic Spectral Roll Off Original	Noisiness DeltaDelta StandardDeviation
Fundamental Frequency Delta Original	Harmonic Spectral Roll Off OriginalDTW	Spectral Decrease Delta Delta StandardDeviation
Harmonic Spectral Centroid Delta	Harmonic Tristimulus Bands Mean	Spectral RollOff Delta StandardDeviation
Harmonic Spectral Kurtosis Delta Bands Mean	Inharmonicity Delta Delta Original	Spread Delta Original
Harmonic Spectral Skewness Bands Mean	Loudness Delta StandardDeviation	
Harmonic Spectral Slope Bands NormDTW	Loudness Delta Original	
Harmonic Spectral Slope DeltaDelta Bands NormDTW	Loudness Delta OriginalDTW	
Harmonic Spectral Variation Delta	MFCC DeltaDelta Bands Mean	
Harmonic Tristimulus Delta Bands StandardDeviation	Noisiness Delta OriginalDTW	
Loudness DeltaDelta Original	Perceptual OddToEvenRatio Delta Bands Mean	
Noise Energy Delta StandardDeviation	Perceptual Spectral Deviation Delta	
Noisiness Delta StandardDeviation	Perceptual Spectral Deviation DeltaDelta Bands Mean	
Noisiness Delta StdDev	Perceptual Spectral Spread	
	Perceptual Spectral Variation Delta Bands StandardDeviation	
	Perceptual Tristimulus Delta Bands Mean	
	Relative Specific Loudness DeltaDelta Bands Mean	
	Sharpness Delta StandardDeviation	
	Spectral Crest Delta Bands Mean	
OSB		
AutoCorrelation Delta Bands StandardDeviation	Energy Envelope Temporal Centroid	Energy Envelope Effective Duration
AutoCorrelation DeltaDelta Bands StandardDeviation	Harmonic Spectral Rolloff Original	Energy Envelope Temporal Centroid
Chroma Delta Bands StandardDeviation	Harmonic Tristimulus Bands Mean	HarmonicEnergy Delta Mean
Energy Envelope Effective Duration	Inharmonicity Delta OriginalOSB	Harmonic OddToEvenRatio DeltaDelta Bands NormOSB
Energy Envelope Temporal Increase	Inharmonicity DeltaDelta Original	Noisiness DeltaDelta StandardDeviation
Harmonic Spectral Centroid Delta	Loudness Delta StandardDeviation	Perceptual Spectral Skewness Delta Bands NormOSB
Harmonic Spectral Kurtosis Delta Bands Mean	Loudness Delta Original	Spectral Decrease DeltaDelta StandardDeviation
Harmonic Spectral Slope DeltaDelta Bands NormOSB	MFCC DeltaDelta Bands Mean	Spectral Rolloff Delta StandardDeviation
Harmonic Spectral Spread Bands NormOSB	Perceptual OddToEvenRatio Delta Bands Mean	Spread Delta Original
Harmonic Spectral Spread Delta	Perceptual Spectral Deviation Delta	
Harmonic Spectral Variation Delta	PerceptualSpectral Deviation DeltaDelta Bands Mean	
Harmonic Tristimulus Delta Bands StandardDeviation	Perceptual Spectral Spread	
Loudness DeltaDelta Original	Perceptual Spectral Variation Delta Bands StandardDeviation	
Noise Energy Delta StandardDeviation	Perceptual Tristimulus Delta Bands Mean	
Noisiness Mean	Relative Specific Loudness DeltaDelta Bands Mean	
Noisiness Delta StandardDeviation		
Noisiness DeltaDelta StdDev		
Signal Zero Crossing Rate DeltaDelta Original		

roulement *Level 3 - 100%*	**stable** *Level 1 - 100%*	**train** *Level 2 - 100%*
DTW		
Harmonic Spectral Centroid Delta Harmonic Spectral Kurtosis Bands NormDTW Harmonic Spectral Roll Off Delta OriginalDTW Harmonic Spectral Spread Delta Harmonic Tristimulus Delta Delta Bands StandardDeviation MFCC Delta Bands StandardDeviation Relative Specific Loudness DeltaDelta Bands StandardDeviation Sharpness StandardDeviation	Perceptual Spectral Roll Off DeltaDelta OriginalDTW	Harmonic Spectral Skewness Delta Bands Mean Loudness DeltaDelta StandardDeviation MFCC Bands StandardDeviation MFCC Delta Bands Mean Noise Energy Delta Original Perceptual Spectral Centroid Delta Bands Mean Perceptual Spectral Slope Bands StandardDeviation Perceptual Spectral Variation DeltaDelta Bands StandardDeviation Perceptual Tristimulus Delta Signal Zero Crossing Rate Delta StandardDeviation Spectral Centroid DeltaDelta Bands StandardDeviation Spectral Slope Spectral Slope Delta Bands StandardDeviation Spectral Slope DeltaDelta Bands StandardDeviation Spectral Variation Bands Norm Spectral Variation Delta Bands Mean Total Energy Delta Original Total Energy Delta OriginalDTW
OSB		
Harmonic Spectral Centroid Delta Harmonic Spectral Rolloff Delta OriginalOSB Harmonic Spectral Spread Delta MFCC Delta Bands StandardDeviation Sharpness StandardDeviation	Loudness Delta OriginaOSB Perceptual Spectral Rolloff DeltaDelta OriginalOSB Sharpness Delta OriginalOSB Spectral Decrease Delta OriginalOSB Spectral Decrease DeltaDelta OriginalOSB Spectral Rolloff Delta OriginalOSB Spectral Rolloff DeltaDelta OriginalOSB	Harmonic Spectral Kurtosis Bands NormDTW Harmonic Spectral Skewness Delta Bands Mean Loudness DeltaDelta StandardDeviation MFCC Bands StandardDeviation MFCC Delta Bands Mean MFCC DeltaDelta Bands StandardDeviation Noise Energy Delta Original Noise Energy Delta OriginalOSB Perceptual Spectral Centroid Delta Bands Mean Perceptual Spectral Slope Bands StandardDeviation Perceptual Spectral Variation DeltaDelta Bands StandardDeviation Perceptual Tristimulus Delta Signal Zero Crossing Rate Delta StandardDeviation Spectral Centroid DeltaDelta Bands StandardDeviation Spectral Slope Spectral Slope Delta Bands StandardDeviation Spectral Slope DeltaDelta Bands StandardDeviation Spectral Variation Bands Norm Spectral Variation Delta Bands Mean Total Energy Delta Original Total Energy Delta Original OSB

Table des figures

Liste des tableaux

Bibliographie

[BC94] D. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *AAAI-94 workshop on knowledge discovery in databases*, volume 2, 1994.

[Der01] E. Deruty. « les descripteurs morphologiques des sons : descripteurs2.doc », rapport interne ecrins. mai 2001.

[EA12a] P. Esling and C. Agon. Multiobjective time series matching for audio classification and retrieval. *IEEE Transactions on Speech Audio and Language Processing 2012 (in review)*, 2012.

[EA12b] P. Esling and C. Agon. Time series data mining and analysis. *ACM Computing Surveys 2012 (Accepted proof)*, 2012.

[God06] R. I. Godøy. Gestural-sonorous objects : embodied extensions of schaeffer's conceptual apparatus. *Organised Sound 11(2) : 149–157 2006 Cambridge University Press*, 2006.

[HJTW12] H. Hwang, K. Jung, Y. Takane, and T. Woodward. Functional multiple-set canonical correlation analysis. *Psychometrika*, 2012.

[KR05] E. Keogh and C.A. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and information systems*, 7(3) :358–386, 2005.

[LJKTM07] Q. Wang L. J.Latecki, S. Koknar-Tezel, and V. Megalooikonomou. Optimal subsequence bijection. 2007.

[LKLC03] J. Lin, E. Keogh, S. Lonardi, and B. Chiu. A symbolic representation of time series, with implications for streaming algorithms. *Proceeding of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, ACM New York, NY, USA, 2003, pp. 2-11, 2003.

[MMHS08] A. Minard, N. Misdariis, O. Houix, and P. Susini. « descripteurs audio de type morphologique pour les sons environnementaux ». projet sampleorchestrator, deliverable sp2-4-2. Technical report, sept. 2008.

[MMHS10] A. Minard, N. Misdariis, O. Houix, and P. Susini. Categorisation de sons environnementaux sur la base de profils morphologiques. In *10eme Congres Francais d'Acoustique*, 2010.

[MMS+10] N. Misdariis, A. Minard, P. Susini, G.Lemaitre, S. McAdams, and Etienne Parizet. Environmental sound perception :metadescription and modeling based on independent primary studies. *Journal on Audio, Speech, and Music Processing*, 2010.

[PD08] G. Peeters and E. Deruty. Automatic morphological description of sounds. *Proceed. 8ème Congres Français d'Acoustique*, Paris, juin 2008.

[PD09] G. Peeters and E. Deruty. Sound indexing using morphological descriptions. *IEEE Transactions on Audio, Speech and Language Processing*, 2009.

[Pee04] G. Peeters. A large set of audio features for sound description (similarity and classification) in the cuidado project. Technical report, 2004.

[Rio01] V. Rioux. « projet ecrins / validation expérimentale phase i : descripteurs morphologiques », rapport interne ecrins. Technical report, novembre 2001.

[Sch66] P. Schaeffer. Traité des objets musicaux. *Paris, France : Seuil*, 1966.

[Sma97] D. Smalley. Spectromorphology : explaining sound-shapes. *Organised Sound 2(2) : 107–26 1997 Cambridge University Press*, 1997.

[SML96] R. G. Stockwell, L. Mansinha, and R. P. Lowe. Localization of the complex spectrum : the s transform. *IEEE Transactions on Signal Processing 44 (4), p 998-1001*, 1996.

[Van79] N. J. VanDerveer. *Human Perception of Environmental Sounds.* PhD thesis, Cornell University, 1979.

[WDR99] M.M. Wanderley, P. Depalle, and X. Rodet. *Contrôle Gestuel de la Synthèse Sonore.* Interfaces Homme-Machine et Creation Musicale. Paris : Hermès Science Publishing, 1999.