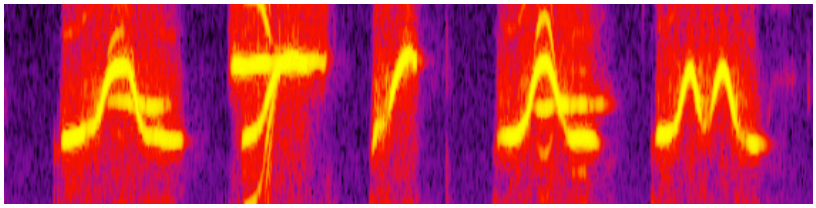
Carlo Baugé
Stage de fin d'études Master ATIAM



Mémoire de Master 2

Étude des similarités acoustiques entre sons environnementaux

Institutions	IRCAM, UPMC Telecom ParisTech
Encadrants	Mathieu Lagrange, Nicolas Misdariis

20 août 2012

Table des matières

Table des figures

Liste des tableaux

Résumé

Dans la plupart des applications de traitement audio, il est important de comprendre les caractéristiques des sons qui rendent compte le mieux de la perception humaine.

Nous nous intéressons particulièrement à la représentation des sons et les mesures de similarité (ou distance) entre sons. Nous montrons l'intérêt de la représentation « scattering » ([AM11]), basée sur une transformée en ondelettes, ainsi que deux variantes. Nous montrons également l'apport de ces représentations par rapport aux représentations et algorithmes de l'état de l'art.

Dans une volonté d'ancrer notre étude sur une meilleure compréhension de la notion de similarité entre sons au sens perceptif du terme, nous proposons également de nouvelles possibilités d'évaluation de nos algorithmes relativement à certaines données perceptives.

Abstract

For most of audio signal processing applications, it is essential to formalize mathematically characteristics of sounds that are important for human perception.

We focus in this report on sound representations and sound similarity measures that can be derived from those representations. In this context, we show the potential of the scattering representation ([AM11]), based on a wavelet transform, as well as two variants specifically tailored for the purpose of modeling environmental and musical sounds. We additionally show the advantage of these representations compared to state of the art representations and algorithms.

In order to achieve a better comprehension of sound similarity in general, we also propose novel methods for evaluating computational approaches with respect to human perception.

Introduction

Dans de nombreuses application de traitement du signal audio, on dispose de nombreuses données que l'on tente d'organiser entre elles afin d'en dégager du sens. Pour obtenir une représentation fine et pertinente, il est important de comprendre les caractéristiques qui rendent compte le mieux de la perception humaine des sons.

Si l'on connaît relativement bien les premières étapes bas niveau de la perception auditive, nous connaissons très peu les processus de représentation supérieurs.

La compréhension de la perception auditive est donc un sujet de recherche encore ouvert aux frontières des neurosciences, de la psycho-acoustique et des mathématiques de traitement du signal. Le travail de ce stage s'inscrit dans cette volonté de mieux comprendre la perception auditive de manière formelle en se concentrant sur l'étude du timbre des sons environnementaux. En effet ceux-ci couvrent un éventail de sons très large acoustiquement allant de textures sonores proches d'un bruit à des sons très harmoniques. Notre perception de ces derniers est probablement moins affectée par des informations de haut niveau sémantique que la perception de la parole ou de morceaux de musique par exemple et sera plus proche des informations purement acoustiques des sons. Bien que notre approche se concentre sur des aspects de traitement du signal, nos travaux se sont basés sur des travaux en neurosciences, en particulier les travaux de Shamma ([WS95],) et en psycho-acoustique, comme les travaux de Gaver ([Gav93]) ou de l'équipe Perception et Design Sonore de l'IRCAM ([HLM+12], [SML06a], [SML06b]).

Après une étude de l'état de l'art en la matière, nous nous intéresserons à un type particulier de représentation des sons, la représentation « scattering », créée par Stéphane Mallat et nous étudierons les éléments qu'une telle représentation peut apporter à notre compréhension de la perception auditive.

Chapitre 1

Motivations et Objectifs

1.1 Motivations

Le stage s'intéresse à la notion de distance (ou dissimilarité) perceptive avant tout entre sons environnementaux. Celle-ci correspond à la différence entre sons perçus par l'oreille humaine ; elle peut se représenter dans un espace dans lequel les sons sont placés de manière a faire correspondre les distances (au sens d'une mesure physique) avec les dissimilarités (perceptives). Nous avons par la suite étendu l'approche à d'autres types de sons (parole, instruments de musique).

Afin d'appréhender correctement les motivations du stage, il convient de s'interroger sur la façon dont on se représente et dont on compare les sons habituellement. En partant de la perception humaine telle que nous l'expérimentons tous les jours lorsqu'elle est orientée vers une tâche, il est naturel de penser notre perception ou représentation tout d'abord comme une transformation du signal perçu – la variation de pression au tympan – sous une forme plus adaptée à la tâche en question. Cette transformation globale, qui est l'aboutissement sans doute d'une série de traitements plus élémentaires, s'accompagne d'une perte d'information. En effet la complexité du signal de départ est réduite de sorte à éliminer la totalité ou une partie des informations inutiles pour la tâche et à ne garder que l'essentiel. Cette dernière représentation qui précède immédiatement le jugement ou la prise de décision, présente donc deux principaux enjeux pour nous. Le premier consiste à comprendre quelle est cette « bonne » information que nous devons garder et quelles sont au contraire les dimensions du signal qu'il faudra éliminer. Le deuxième consiste à trouver la meilleure forme ou structure pour figurer cette information. Contenir la bonne information indiquerait donc une potentialité de bien pouvoir accomplir la tâche (puisque l'information est là) alors que bien structurer l'information voudrait dire être capable de juger ou séparer correctement les éléments.

Une représentation du signal proposée par Stéphane Mallat ([AM11]), la représentation scattering, semble répondre à ces enjeux. Le stage se propose donc d'utiliser cette représentation afin de la comparer à d'autres types de représentations plus classiques du type du Mel-spectre ou des Mel-Frequency Cepstral Coefficients (MFCC).

L'étude de la dissimilarité se démarque d'une tâche de classification dans la mesure où la dissimilarité entre deux sons n'est pas directement liée à leur classe. Ainsi dans cette approche, les sons sont placés dans un continuum perceptif relativement aux autres sons, différemment de la classification qui apportera une caractérisation discrète des sons. L'objectif est donc davantage celui de reproduire un espace des distances perçues entre deux sons que d'associer une classe à chaque son. Cette approche, comparée à l'approche de classification, repose sur l'idée que la perception des sons n'est pas seulement catégorielle et donc discrète mais peut être vue comme un espace métrique continu dans lequel les sons acquièrent leur identité par les relations de similarité qui les lient. De plus, cette approche présente certains avantages liés notamment aux

bases de données dont nous disposons et qui seront présentées dans la prochaine section. En effet certaines de ces bases contiennent peu d'éléments par classe et restent globalement aussi assez petites (de 56 à 100 sons au total), ce qui dissuade d'utiliser l'approche classification et donc de découper les bases en des bases encore plus petites d'apprentissage et test. Enfin l'approche de la dissimilarité permet de ne pas dépendre des bases de données dans le calcul des distances entre sons. De cette façon la dimension de l'espace de représentation des sons ne pose pas de problème direct comme en classification. Cependant ceci n'empêche pas évidemment d'être dépendant des bases de données au moment de l'évaluation.

De la classification à la dissimilarité, l'évaluation peut être repensée. En effet celle-ci ne concerne plus autant un attribut binaire indiquant l'appartenance à une classe que la distance entre chaque élément et tous les autres. Comme nous le verrons, pour les bases de données de sons environnementaux nous disposons de matrices de distances entre les sons, issues d'expériences perceptives ([GKW07], [HLM+12]) qui peuvent être utilisées comme références pour une évaluation mieux adaptée.

Ainsi le travail durant le stage se caractérise à la fois par la tâche de dissimilarité choisie, la recherche d'une représentation plus adaptée à cette tâche à l'aide de la représentation scattering et la recherche d'une évaluation adaptée.

1.2 Protocole expérimental

De façon classique, notre problématique peut être découpée en trois phases. Une première phase de représentation des sons, puis l'application d'un algorithme de calcul (ici mesure de distance entre les représentations) et enfin une phase d'évaluation (voir figure 1.1).

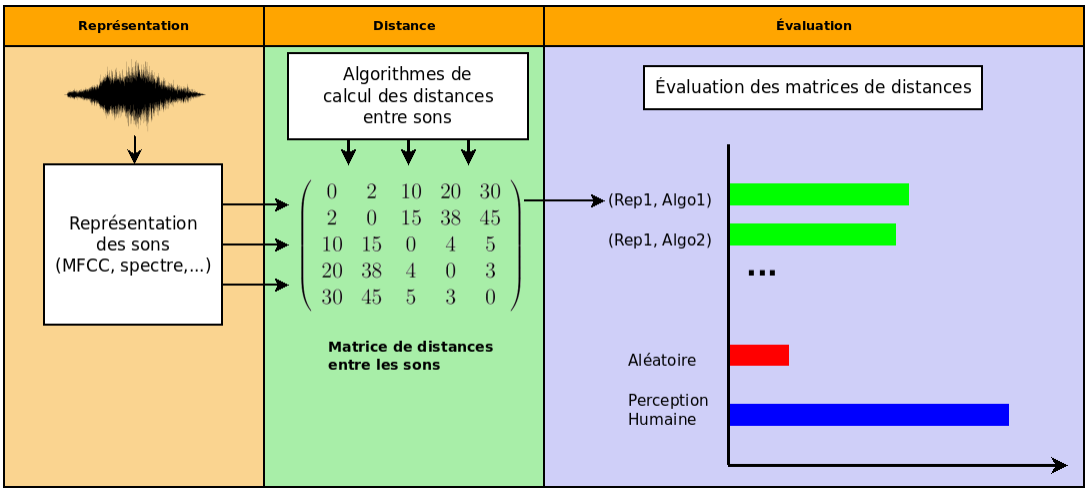


FIGURE 1.1 – Schéma général du protocole expérimental

La phase de représentation va consister typiquement à transformer les signaux sonores en une représentation mieux structurée et plus facilement manipulable comme le spectrogramme, le Mel-spectre ou les MFCC ou la représentation scattering. De la forme d'un vecteur ou d'une matrice, cette dernière servira d'entrée à une métrique de distance. Comme nous le verrons plus en détail par la suite, nous serons amenés à utiliser des métriques de distance aussi bien sur des matrices (comme il est habituel pour le spectrogramme par exemple) que sur des vecteurs (par exemple lorsqu'on ne considère qu'une seule fenêtre temporelle). En effet la taille de la fenêtre d'analyse sera un élément important de notre travail. Pour une fenêtre d'analyse petite

(inférieure à la taille de signal), la représentation sera sous forme matricielle temps-fréquence alors que pour une seule fenêtre d'analyse, la représentation sera un vecteur fréquence.

Ainsi les métriques de distance à utiliser seront différentes en fonction du type de représentation. Nous nous pencherons sur deux métriques de distance entre représentations matricielle et qui constitueront l'état l'art : l'algorithme Bag of Frame (BOF) et l'algorithme Dynamic Time Warping (DTW). Pour les représentations vectorielles, nous utiliserons les distances usuelles (euclidienne, cityblock, cosinus) ainsi qu'une mesure de distance issue du coefficient de corrélation de Spearman.

Nous obtenons de cette façon des matrices de distances entre chaque paire de sons que nous évaluerons à l'aide d'une métrique connue : le Mean Average Precision (MAP). Afin de situer nos résultats plus précisément nous comparerons systématiquement nos performances à des performances aléatoires ainsi qu'à des performances « humaines » dans la mesure du possible.

Nous étudierons également des façons plus fines d'évaluer ces matrices de distance, d'une part à l'aide d'un outil de visualisation que nous avons conçu et développé et d'autre part à l'aide d'une métrique très proche du Mean Average Precision (MAP) que nous proposerons dans la section 3 et qui tient compte de données issues d'expériences sur la perception.

Chapitre 2

Bases de données

Nous nous sommes appuyés principalement sur trois bases de sons environnementaux liées aux articles [GKW07] et [HLM+12]. Bien que notre travail se soit concentré avant tout sur ces bases, nous avons dans un deuxième temps élargi notre champ d'expérimentation à d'autres types de bases de données : trois bases de données d'instruments de musique, une base de donnée de parole pour la reconnaissance de chiffres lus et une base de données de sons d'insectes.

2.1 Égalisation des sons

Avant de commencer tout traitement nous avons égalisé les sons de chaque base de données. Ce même type d'égalisation a été appliqué à toutes les bases de données. Il s'agit du type d'égalisation utilisé dans [GKW07].

Pour une base de données de N sons $\mathcal{B}= (x_i)_{i\in[1,N]}$, on considère pour chaque son la moyenne quadratique d'une fenêtre de 100 ms autour du maximum (RMS_i). On égalise alors tous les sons par rapport à

$$m = min_{i\in[1,N]} RMS_i$$

Les sons égalisés $\hat{x}_i$ sont calculés de la façon suivante :

$$\hat{x}_i = \frac{m}{RMS_i} \cdot x_i$$

Cette égalisation est particulièrement utile dans notre cas car les sons sont de longueurs différentes et peuvent contenir des moments de silence au début et à la fin qui ne doivent pas influencer l'égalisation.

2.2 Bases de sons environnementaux

Les trois bases de sons environnementaux que nous présentons dans cette section ont servi pour des expériences sur la perception sonore que nous décrivons brièvement.

2.2.1 Base « Gygi » ([GKW07])

Cette base est composée de 100 sons environnementaux de courte durée (entre 0.5 et 4 secondes) organisés en 50 classes de 2 sons. Celles-ci couvrent un large éventail de sons environnementaux comme des sons d'animaux (miaulement, hennissement, meuglement), des éternuements, des cris de bébés ou des bruits de machines tels que des bruits de moteurs (avion,

scie électrique, essuies-glaces) en passant par des sons harmoniques (harpe), des sons d'objets (clavier d'ordinateur, glaçons tombant dans un verre) et des sons de la nature (tonnerre, vagues, pluie). Chaque paire de sons appartenant à la même classe correspond à deux représentants choisis « aussi éloignés que possible » à l'intérieur de la même classe. Les sons sont monophoniques, d'une résolution de 16 bits, échantillonnés à 44,1 kHz.

Cette base a été utilisée dans le cadre d'une étude visant à mesurer la similarité entre les sons environnementaux. Les enjeux étaient de trouver l'information acoustique qui caractérise un objet ou un événement ou de comprendre comment ceux-ci sont identifiés en l'absence de spécificité acoustique. 4 expériences ont été menées dont la première nous intéresse pour la suite. Pendant la première, 4 sujets ont noté la similarité de manière assez libre entre chacune des 10000 paires de sons – donc incluant les paires de sons identiques et en présentant les paires dans les deux sens – sur une échelle de 1 (complètement dissimilaires) à 7 (aussi similaires que possible). On leur a demandé d'utiliser possiblement toute l'échelle de 1 à 7 et d'essayer d'établir à 4 la réponse moyenne. À l'issue de l'expérience chacun des sujets a donc fourni une matrice de similarité 100 par 100.

En résumé, nous avons accès sur cette base à la fois aux classes des sons et aux matrices de distances établies par les 4 sujets lors de cette première expérience. Ceci nous permettra d'avoir plusieurs références pour l'évaluation sur cette base.

2.2.2 Bases « Houix » ([HLM+12])

Dans [HLM+12], Houix et al. s'intéressent davantage à la catégorisation des sons environnementaux.

Deux bases sont associées à [HLM+12], chacune correspondant à une expérience différente. La première base que l'on appellera « Houix1 » contient 60 sons qu'on entend habituellement dans une cuisine (robinet, casseroles, cafetière, micro-onde). La deuxième base (Houix2) est composée de 56 sons uniquement "d'intérieur" provoqués par des objets solides ou par l'action d'un agent sur un objet avec ou sans l'aide d'un outil.

D'un point de vue plus technique les sons sont monophoniques (durant entre 0,5 et 10 secondes pour Houix1 et entre 0,5 et 6,9 secondes pour Houix2) ajustés au niveau normal attendu dans une cuisine (au maximum 75 dB), avec une résolution de 16 bits et échantillonnés à 44,1 kHz).

Pour la première expérience, on a placé sur un écran les 60 sons de la base Houix1 représentés par des points et l'on a demandé à 15 sujets de regrouper en des catégories libres les sons qui leur semblaient partager les mêmes caractéristiques ou faire partie du même groupe. Le nombre de classes créé varie entre 6 et 24 selon le sujet et les résultats sont représentés sous forme d'une matrice de dissemblance qui sera utile pour la suite de nos travaux. Suite à cette expérience une analyse en dendrogramme ainsi qu'une analyse lexicale ont permis de mettre en évidence 5 classes, en accord avec les études de Gaver sur la perception des sons environnementaux ([**?**]) : les sons solides (35 sons), liquides (13 sons), gazeux (4 sons) et les sons de machines (8 sons).

Afin d'étudier plus en détail des catégorisations plus fines que celles mises en évidence par la première expérience, une deuxième expérience similaire à la première a été menée sur la base Houix2 contenant donc des sons uniquement de la classe « solides ». Ici des instructions plus précises ont été données aux participants qui devaient organiser les sons selon l'action associée à chaque son et non selon les objets impliqués. L'analyse des résultats de cette dernière expérience a permis d'organiser en classes hiérarchiques les sons de la base Houix2 comme le montre la figure 2.1.

Finalement, comme pour la base Gygi, nous avons accès à la fois à des informations de classes (simple pour la base Houix1 et hiérarchique pour la base Houix2) et à deux matrices de distances

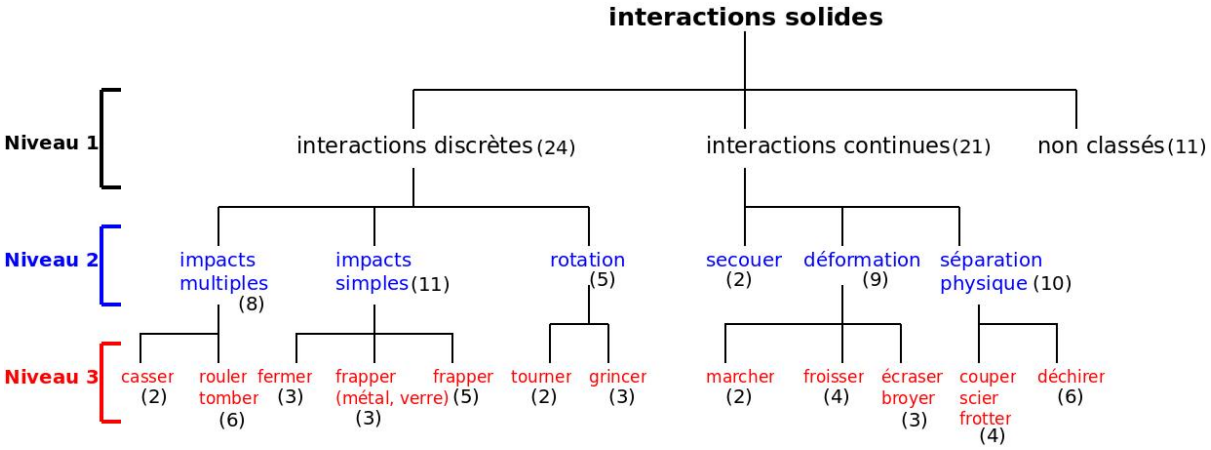


FIGURE 2.1 – Résumé des deux analyses lexicales sur les résultats de l'expérience 2. Les chiffres entre parenthèses indiquent le nombre de sons pour chaque classe de l'arbre

entre les sons pour les bases Houix1 et Houix2.

2.2.3 Extension de la base Gygi

Comme nous l'avons vu, les trois bases que nous avons utilisées sont relativement petites et contiennent peu d'éléments par classe, notamment la base Gygi qui ne contient que 2 éléments par classe. La base Gygi étant la base la plus variée, c'est aussi celle pour laquelle il est facile de trouver d'autres sons similaires pour l'agrandir. Nous avons donc décidé d'enrichir la base Gygi avec de nouveaux sons pour porter le nombre minimum d'éléments par classe à 5.

Pour cela, nous avons cherché des sons enregistrés avec des caractéristiques techniques minimum (formats sans perte, échantillonnés à 44,1 kHz et 16 bits de résolution), correspondant à chaque classe présente dans la base Gygi et se situant dans les limites définies par les deux représentants existants. Pour que cette nouvelle base soit réutilisable le plus facilement possible, nous n'avons considéré que des sons libres de droits ou sous des licences peu restrictives du type des licences Creative Commons.

Malheureusement, pour respecter les critères que nous nous sommes fixés, trois classes sur cinquante ne contiennent que 4 représentants : la classe « electric saw », « horse run » et « wipers ».

Afin de collecter le sons de la base Gygi Extended, nous nous sommes principalement appuyés sur une interface web à la base Freesound (`http://www.freesound.org/`) créée par Mathias Rossignol qui permet d'ajouter simplement un son de Freesound à une base de données en donnant la possibilité d'ajouter des informations supplémentaires (tags) et d'extraire une partie d'un son donné.

Ceci nous a permis de trouver la majorité des sons supplémentaires (182 sons sur 194) mais nous nous sommes également appuyés sur le site `http://www.soundbible.com` (4 sons) qui propose des sons sous des licences peu restrictives, sur le moteur de recherche Findsounds (`http://www.findsounds.com/`) et sur des recherches libres sur internet. Les détails des nouveaux sons sont précisés en annexe D.

2.3 Bases d'instruments

Après une étude sur les bases de sons environnementaux, nous nous sommes attelés à trois bases de sons d'instruments de musique. Souvent ces bases sont attachées à des études de classification obtenant parfois de très bons résultats. Notre objectif en utilisant ces bases n'est pas de se comparer aux études existantes puisque notre approche s'intéresse à la similarité et non à la classification.

2.3.1 Base « Iowa » ([Fri97])

La « Music Intruments Samples » (MIS) a été créée à l'université de l'Iowa en 1997. Il s'agit d'enregistrements monophoniques en chambre anéchoïque (d'une résolution de 16 bits et échantillonnés à 44,1 kHz) pouvant durer jusqu'à 10 secondes.

La base que nous appellerons dans la suite base Iowa utilise des sons tirés de cette base. Elle est constituée de 3637 notes isolées jouées par 15 instruments d'orchestre symphonique.

Pour des raisons pratiques de temps de calcul, nous n'avons considéré que les trois premières secondes de chaque son. Par cette opération 1183 sons ont été écourtés, ceux-ci sont principalement des sons tenus pour lesquels nous perdons notamment la phase d'atténuation.

2.3.2 Base « Real World Computing (RWC) » ([GN03])

La RWC music database a été construite par le RWC Music Database Sub-Working Group du Real World Computing Partnership (RWCP) du Japon. Elle est constituée de 6 bases : la Popular Music Database, la Royalty-Free Music Database, la Classical Music Database, la Jazz Music Database, la Music Genre Database, et la Musical Instrument Sound Database (voir [GN03]).

C'est de cette dernière, la Musical Instrument Sound Database qu'est tirée la base que nous avons utilisée et que nous nommerons RWC. La variabilité à l'intérieur d'une même classe est due à des modes de jeu différents, des musiciens et des instruments différents ainsi que des hauteurs de notes et des dynamiques différentes.

La base RWC contient 6138 notes isolées jouées par 14 instruments présents également dans la base Iowa (seule la classe clarinette basse n'y est pas représentée). Tous les sons ont également une résolution de 16 bits et sont échantillonnés à 44,1kHz.

De même que la base Iowa, nous avons écourtés les sons trop longs (1792 sons) à trois secondes.

2.3.3 Base « solosDb » ([ERD06a] et [ERD06b])

Une partie de cette base a été construite à partir d'enregistrements « live » et en studio (au format PCM ou mp3 à 64 kbps) alors qu'une autre partie vient de la RWC Jazz Music Database ([ERD06a], [ERD06b], voir section précédente).

Cette base contient 505 morceaux de musique classique, jazz ou classique contemporaine pouvant durer plusieurs minutes et joués par 19 instruments d'orchestre symphonique ou jazz (saxophone, piano, violon, violoncelle, batterie, clarinette, clarinette basse, etc...) dont les instruments utilisés dans les base Iowa et RWC.

Pour des raisons de temps de calcul, nous n'avons pas utilisé les sons tels quels mais nous n'avons gardé que trois secondes au début de chaque morceau lorsque leur longueur était supérieure à trois secondes. Le tableau 2.1, résume les caractéristiques de durée des sons après cette opération.

2.4 Autres bases

Enfin nous avons utilisé deux autres bases : une base de données de sons de parole énonçant des chiffres isolés et une base de données pour la reconnaissance d'insectes.

2.4.1 Base « Digit » ([LD91])

La base issue de [LD91] a été créée en 1982 et rassemble 25000 séries de chiffres énoncées par plus de 326 hommes, femmes et enfants des États-Unis. Les locuteurs sont répartis de la façon suivante :

- 111 hommes entre 21 et 70 ans
- 114 femmes entre 17 et 59 ans
- 50 garçons entre 6 et 14 ans
- 51 filles entre 8 et 15 ans

Le choix des locuteurs s'est fait également dans le but de représenter équitablement les différents accents régionaux. Pour chaque région, 5 hommes blancs, 5 hommes noirs, 5 femmes blanches, 5 femmes noires au moins ont été sélectionnés.

Pour notre travail et pour des raisons que nous verrons plus tard, nous nous sommes limités aux chiffres énoncés seuls, ce qui laisse 347 énoncés organisés en 11 classes (chiffres de 1 à 9 plus les classes "zero" - prononcé en toutes lettres - et "o").

2.4.2 Base « Insects » ([CKB12])

La base de données Insects est diffusée sur internet dans le cadre d'un concours (en novembre 2012) de classification d'insectes. Les données sont sous forme de sons au format WAVE (résolution de 16 bits et échantillonnage à 16kHz) mais proviennent de données optiques obtenues lors du passage d'un insecte à travers un faisceau laser.

Les sons sont très courts (de l'ordre de 1 s) et l'énergie du signal est concentrée dans un temps encore plus court (de l'ordre de 100 ms) ce qui nous a poussés à segmenter les sons de sorte à extraire au plus près de là où l'énergie du signal est concentrée. Ainsi les caractéristiques de la nouvelles base après segmentation sont précisées dans le tableau 2.1.

2.5 Résumé

Le tableau 2.1 résume les caractéristiques de chaque base :

	Durée des sons (s)				Nombre d'éléments par classe				Nombre total	
Bases	max	min	moy	σ	max	min	moy	σ	classes	éléments
Bases de sons environnementaux										
Gygi	4.00	0.58	2.32	0.86	2	2	2.00	0.00	50	100
GygiExt	4.80	0.21	2.40	0.92	12	4	5.88	1.77	50	294
Houix1	10.46	0.47	2.92	1.98	35	4	15.00	13.83	4	60
Houix2	6.95	0.58	2.28	1.49	24	11	18.67	6.81	(3,10,13)	56
Bases de sons instrumentaux										
SolosDb	3.00	2.48	3.00	0.03	68	6	26.58	19.05	19	505
Iowa	3.00	0.07	2.52	0.49	536	98	242.47	120.71	15	3637
RWC	3.00	0.23	2.44	0.60	888	198	438.43	218.11	14	6138
Base de parole										
Digit	1.54	0.63	0.98	0.16	38	26	31.55	3.27	11	347
Autre base										
Insects	0.61	0.14	0.21	0.07	100	100	100.00	0.00	5	500

TABLE 2.1 – Tableau résumant les statistiques des différentes bases de sons. Les statistiques concernent deux caractéristiques des bases : les classes et la durée des sons. Le maximum, minimum, la moyenne et l'écart-type sont données pour ces deux caractéristiques

Chapitre 3

Évaluation

Avant de présenter la partie algorithmique, il est utile de s'interroger sur l'évaluation. En effet comme nous l'avons suggéré au début de ce rapport, la tâche de dissimilarité que nous nous sommes proposée pose un problème différent d'une tâche de classification bien que toutes deux partagent certaines similitudes. Nous allons donc essayer dans cette partie de clarifier un peu mieux les différentes questions que pose notre tâche en termes d'évaluation et nous exposerons les outils qui nous ont permis d'y apporter une réponse.

3.1 Réflexion sur les qualités d'une métrique d'évaluation

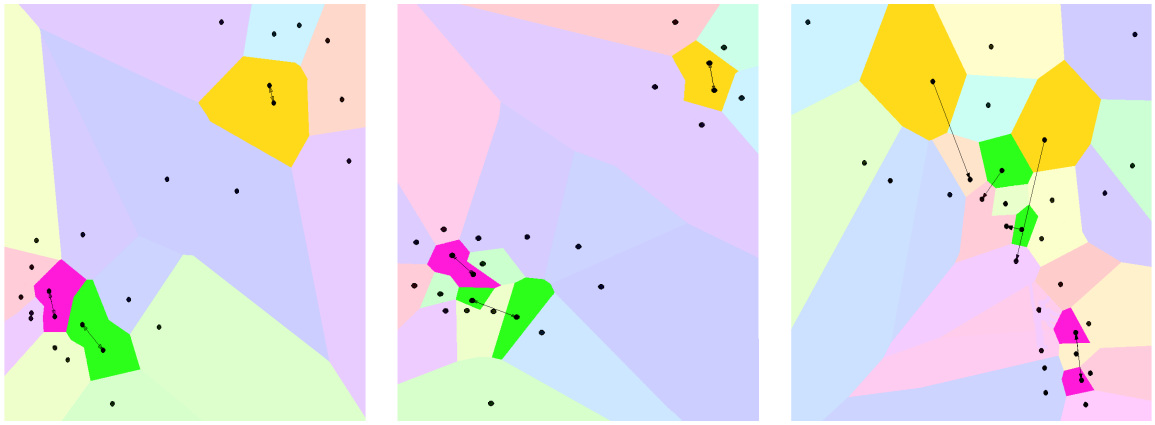
L'objet sur lequel nous devrons évaluer les performances de nos couples (représentation, distance) est une matrice de distances entre tous les sons d'une base de données. Nous connaissons pour cela la classe de chaque son et, dans le cas des bases Gygi et Houix, nous avons également des matrices de distances de référence.

En utilisant uniquement les classes comme référence, on s'aperçoit clairement que l'on ne peut pas analyser toute l'information présente dans la matrice de distances à évaluer car il y a plus d'information dans une matrice de distance que dans une liste d'appartenance à des classes. Plus précisément, on peut expliquer l'information supplémentaire contenue dans une matrice de distances en en dégageant deux aspects : le positionnement des sons entre eux à l'intérieur d'une même classe (distances intra-classes) et le positionnement des classes entre elles (distances inter-classes).

La figure 3.1 illustre ces aspects. Les deux figures les plus à gauche sont des représentations graphiques – effectuées à l'aide d'une analyse MDS et d'un diagramme de Voronoï (voir Annexe A) – de matrices de distances issues des expériences sur la perception de [GKW07]. La figure la plus à droite montre la même représentation d'une matrice de distances calculée à partir d'un algorithme d'état de l'art. Dans cette représentation chaque point correspond à un son, la couleur qui l'entoure correspond à sa classe et les flèches relient un son à son plus proche voisin [1].

Si l'on considère les trois classes mises en évidence, on remarque les deux aspects cités plus haut : une organisation inter-classes générale qui varie – dans les deux premières figures, la classe Klaxon (jaune) est éloignée des classes Chasse d'eau et Tonnerre qui sont très proches – et une organisation intra-classes : à l'intérieur de chaque classe, les sons sont proches les uns des autres, suivant une certaine organisation. Sous ce dernier point de vue, on voit clairement que le plus proche voisin reste dans la même classe sur les deux premières figures mais sur la dernière, le plus proche voisin de certains sons n'appartient pas à la même classe. Ces deux aspects sont

1. Sur la figure 3.1 il peut sembler étrange que les flèches ne pointent pas toujours vers le plus proche voisin "visible" sur la figure, cela est du à la visualisation qui ne respecte pas parfaitement les distances de la matrice de distance. Ainsi le plus proche voisin réel est indiqué par la flèche



Humain moyenne | Humain sujet 1 | État de l'art

FIGURE 3.1 – Illustration des différences inter- et intra-classes pour différentes matrices de distances. Les données sont issues de la base Gygi. Les classes de couleur plus vive sont les classes Klaxon (jaune), Chasse d'eau (vert) et Tonnerre (violet).

complémentaires et l'on imagine aisément que si l'on n'évalue que l'un des deux, l'autre peut varier sans répercussion sur la mesure d'évaluation. En particulier si l'on s'appuie uniquement sur les classes des sons, l'organisation inter-classes ne peut être évaluée. On peut également noter que même l'organisation intra-classe n'est évaluée que de manière partielle.

En fonction des informations disponibles pour l'évaluation (classes des sons ou matrice de distances de référence), nous adapterons notre mesure d'évaluation. Lorsque l'on prendra pour référence la classe de chaque son, nous nous appuierons sur une métrique d'évaluation beaucoup utilisée en recherche d'information, le Mean Average Precision (MAP). Afin de tenir compte des matrices de distances perceptives dont on dispose pour certaines bases (voir section 2.2), nous proposerons une extension du MAP qui calculera les performances par rapport à une matrice de distances de référence.

3.2 Mean Average Precision (MAP)

La mesure principale que nous avons utilisée pour évaluer nos résultats est issue du domaine de la recherche d'information. Elle est utilisée typiquement pour évaluer les résultats d'une requête à une base de données. Ici nous l'utilisons dans un contexte légèrement différent, nous allons donc la présenter dans le cadre de l'évaluation d'une matrice de distances en prenant comme référence les classes des sons.

Le MAP est calculé à partir de deux mesures plus simples : la « precision » et le « recall ». Nous allons donc tout d'abord présenter ces deux mesures avant d'expliquer le calcul du MAP. Pour cela nous considérons un son n, l'ensemble de ses k plus proches voisins $V_n(k)$ (k-voisinage) et l'ensemble des sons de sa classe C_n.

3.2.1 Précision

La précision est la proportion de sons de la même classe que n dans son k-voisinage. Elle s'écrit :

$$precision_n(k) = \frac{|V_n(k) \cap C_n|}{|V_n(k)|}$$

La figure 3.2 illustre cette formule pour une matrice de distances entre 100 points organisés en 10 classes de 10 sons. Deux cas sont tracés : le cas idéal et le cas aléatoire. Dans le cas idéal où les plus proches voisins sont tous de la bonne classe, la précision vaut 1 lorsque k est inférieur à 9 (soit $k \leqslant$ nombre d'éléments dans la classe-1) puis décroît en $1/x$ jusqu'à 9.09% (soit nombre d'éléments dans la classe-1 / nombre de points-1). Les performances des algorithmes se situeront donc entre les deux courbes idéale et aléatoire.

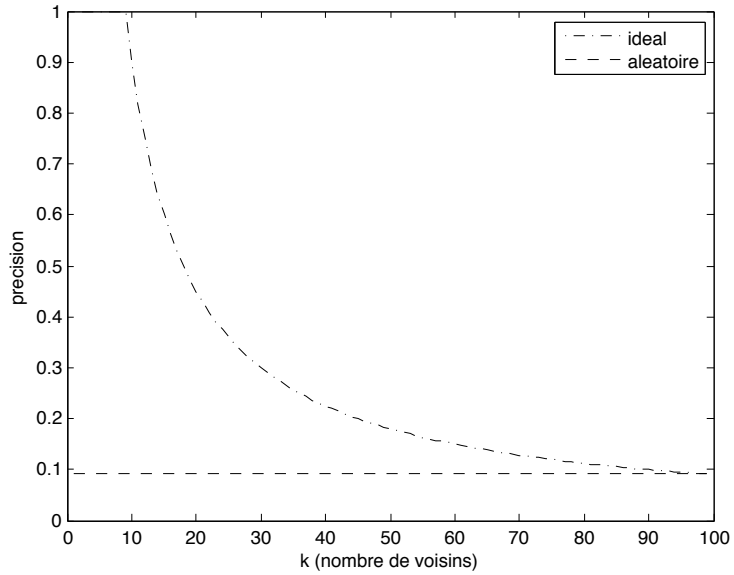


FIGURE 3.2 – Illustration du comportement de la précision en fonction du nombre de voisins considéré.

3.2.2 Recall

Dans un k-voisinage du point n, le recall est le nombre de voisins de la même classe que n sur le nombre d'éléments dans la classe. Il s'écrit :

$$recall_n(k) = \frac{|V_n(k) \cap C_n|}{|C_n|}$$

La figure 3.3 montre les cas idéal et aléatoire pour la mesure de recall. On s'aperçoit bien que lorsque les plus proches voisins sont tous de la même classe que n, le recall augmentera très rapidement jusqu'à atteindre 1 (avec une pente de 1/nombre d'éléments par classe-1).

3.2.3 MAP

On remarque que les deux mesures que nous venons de présenter fournissent des informations complémentaires. En effet, à partir du nombre de voisins de la bonne classe, la précision tient compte du rang auquel ils se trouvent par rapport à n alors que le recall s'intéresse au nombre de voisins de la bonne classe par rapport au nombre total d'éléments dans la classe. L'Average Precision permet d'avoir un compromis entre ces deux métriques et de ne plus dépendre du rayon k de voisinage.

Si l'on a N sons au total, l'Average Precision résume dans une mesure ces deux aspects de la façon suivante :

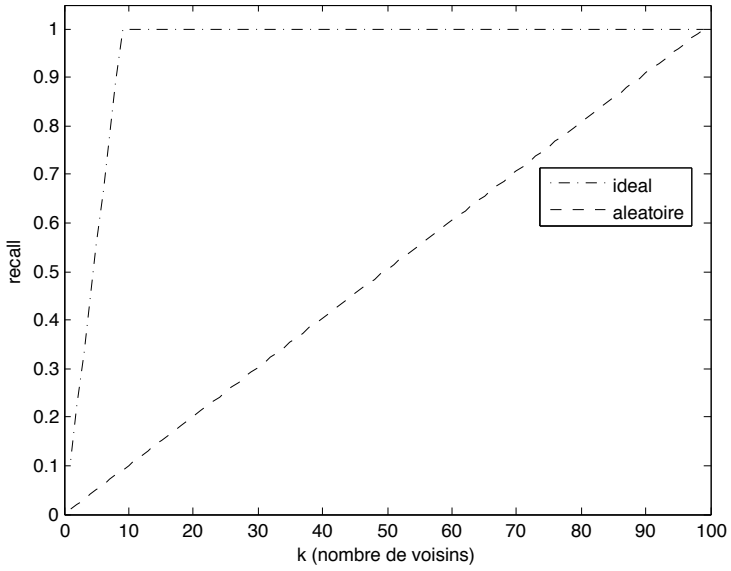


FIGURE 3.3 – Illustration du comportement du recall en fonction du nombre de voisins considéré.

$$AveP_n = \sum_{k=1}^{N} precision_n(k) \cdot (recall_n(k) - recall_n(k-1))$$

À partir de là, le MAP est la moyenne de l'Average Precision pour n variant de 1 à N.

La figure 3.4 montre l'arbitrage que fait le MAP entre précision et recall. Les exemples sont tirés des résultats de la base de données Gygi. À gauche, il s'agit du résultat pour une matrice de distances issues de l'expérience sur la perception et l'exemple de droite montre les performances pour un algorithme état de l'art.

À ce stade, un point supplémentaire est à noter. En effet le MAP s'attache comme nous l'avons vu au rang des voisins d'un point n, qui est déterminé à partir d'une ligne ou une colonne de la matrice de distances. Donc si certains points sont à la même distance de n, il y a une indétermination sur le rang de ces points, et la valeur du MAP peut varier et perd de son sens. Pour y remédier on peut par exemple moyenner les valeurs du MAP sur toutes les permutations possibles des valeurs doublées.

3.3 Extension du MAP (SMAP)

3.3.1 Présentation générale

Le MAP tel que présenté ci-dessus prend comme référence les classes des points et ne peut par conséquent pas donner d'information sur l'organisation inter et intra-classe abordée précédemment. Nous présentons ici une extension du MAP que nous appellerons Similarity Mean Average Precision (SMAP), qui prend comme référence une matrice de distances et donne une évaluation de l'organisation inter-classes et intra-classes d'une matrice de distances.

Considérons une matrice de référence et une matrice à évaluer, des distances entre N points. L'idée du SMAP est de calculer le MAP avec des classes construites à partir de la matrice de référence. Pour construire ces classes, on se fixe un nombre d'éléments par classe c et l'on construit N classes $(C_i)_{i\in[1,N]}$ en les générant à partir de chacun des N points. Pour $i \in [1, N]$, la classe C_i sera l'ensemble des c plus proches voisins (selon la matrice de distances de référence) du point i

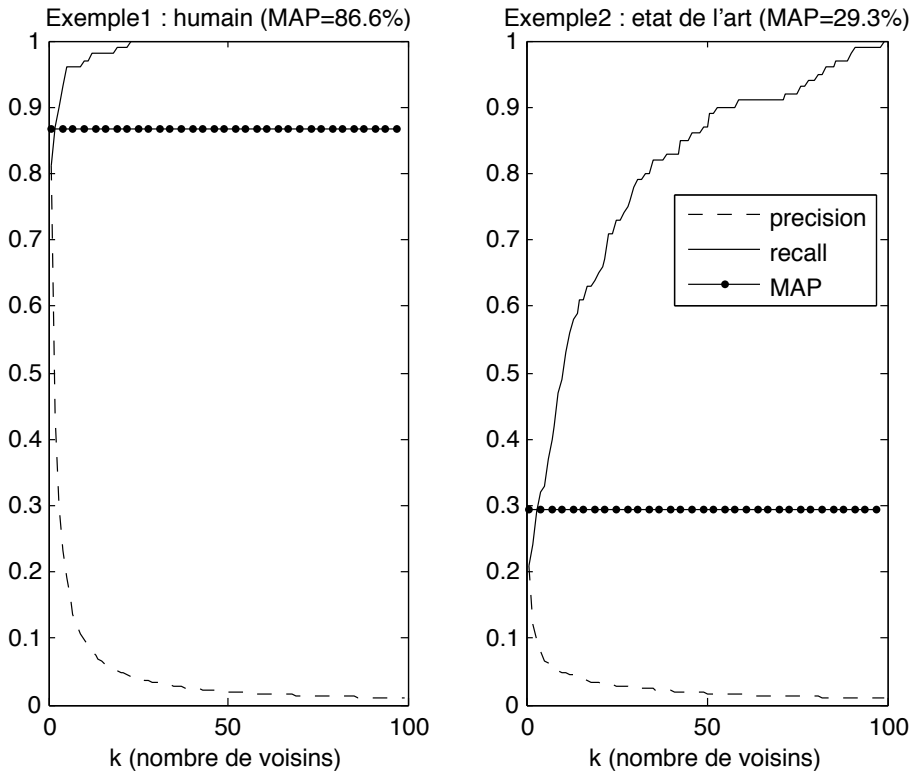


FIGURE 3.4 – Exemples résumant la précision, le recall et le MAP

et pour le calcul du MAP, l'$AveP_i$ sera donc calculée en prenant la classe C_i pour référence. On remarque que les classes C_i peuvent potentiellement se recouvrir voire être identique mais il n'y a pas d'ambiguïté dans la classe à considérer lors du calcul d'$AveP_i$.

Lorsqu'on dispose à la fois d'une référence sous forme de classes et sous forme de matrice de distance, sous certaines conditions, on peut avec le SMAP retrouver exactement le MAP usuel calculé à partir des classes des sons. Pour cela il suffit que le nombre d'éléments par classes soit constant et que la matrice de référence ait un MAP égal à 100%.

Nous avons vu comment calculer le SMAP à un horizon c, c'est-à-dire en créant N classes de c éléments. Pour un c petit, cela donne une idée de la façon dont la matrice à évaluer respecte l'organisation locale des points de la matrice de référence. En augmentant c, la mesure devient plus tolérante sur les différences d'organisations locales et se concentre davantage sur des structures plus grandes.

Ainsi on peut tracer une courbe d'évolution du SMAP en fonction de c, comme le montre la figure 3.5. Les courbes correspondent aux SMAP de trois matrices de distances entre les sons de la base Gygi. La matrice de référence est la moyenne des matrices de distances fournies par les 4 sujets de l'expérience décrite dans [GKW07] (voir 2.2.1). Trois matrices de distances ont été évaluées à l'aide du SMAP. La première courbe correspond la matrice de distances du sujet 1. La deuxième est la matrice de distances donnée par un algorithme d'état de l'art (BOF) que nous verrons dans la suite et enfin la droite correspond aux valeurs du SMAP aléatoire.

Plusieurs points valent la peine d'être remarqués. Tout d'abord la valeur du SMAP pour $c = 2$ est très proche du MAP. En effet en considérant les 50 classes (2 éléments par classe) de la base Gygi le SMAP vaut 86.6% pour le sujet 1 (contre un MAP de 86.4%) et 30.2% pour l'algorithme BOF (contre un MAP de 29.4%). Les valeurs ne sont pas identiques car la matrice de référence a un MAP de 95.1%. D'autre part, on remarque que le SMAP est compris dans

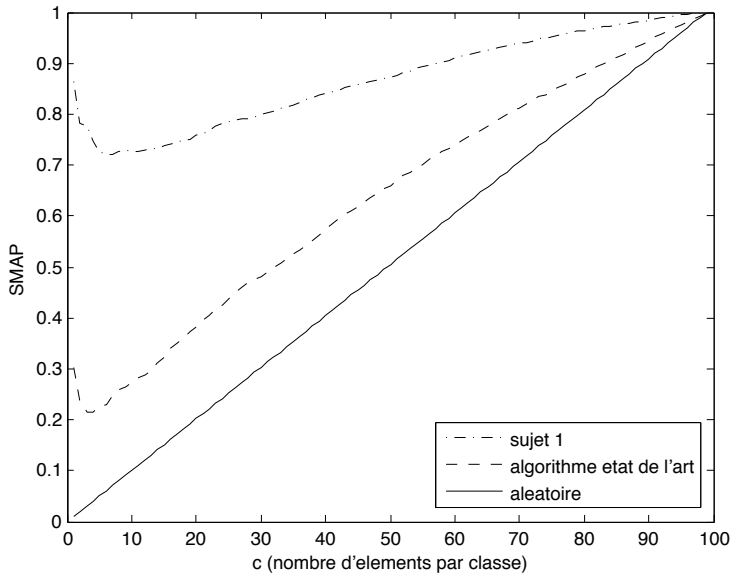


FIGURE 3.5 – Évolution du SMAP selon le nombre d'éléments par classe. Trois cas sont tracés : le résultat pour une matrice issue d'un sujet humain, pour une matrice issue d'un algorithme état de l'art (BOF) et le résultat théorique aléatoire.

un triangle et se rapproche inévitablement de 1 lorsque c est assez grand. En effet, à l'extrême, lorsqu'il y a 100 éléments par classe, tous les éléments appartiennent à la même classe et le MAP vaut 100%.

Chapitre 4

État de l'art

Dans ce chapitre, nous présenterons à la fois les types de représentation et les mesures de distances sur ces représentations qui constituent l'état de l'art. Nous allons donc commencer par introduire les 3 types de représentation spectrogramme, Mel-spectre et Mel-Frequency Cepstral Coefficients (MFCC) en nous attardant sur cette dernière et nous continuerons en expliquant le fonctionnement des algorithmes BOF et DTW permettant de mesurer la distance entre représentations.

4.1 Représentations des sons

Nous avons choisis d'utiliser trois types de représentations usuelles : le spectrogramme, le Mel-spectre et les MFCC.

Le spectrogramme est la représentation plus simple et s'obtient en calculant le module d'une transformée de Fourier discrète à fenêtres du signal.

Le Mel-spectre est un spectrogramme où l'échelle linéaire des fréquence est convertie en échelle de Mel, logarithmique. En effet celle-ci est basée sur la perception humaine des sons et a été construite de façon à ce qu'une variation de la fréquence perçue comme constante corresponde à une variation constante dans l'échelle de Mel. L'échelle des fréquences est linéaire jusqu'à environ 1000Hz puis logarithmique au-delà (voir [Log00]).

Le Mel-spectre sert de base au calcul des Mel-Frequency Cepstral Coefficients (MFCC). Ces derniers sont calculés en appliquant une Discrete Cosine Transform (DCT) au logarithme des coefficients du Mel-spectre ([Log00]). Les MFCC sont utilisés dans de nombreux domaines de traitement audio, en particulier ceux traitant des sons vocaux ou harmoniques. En effet en s'appuyant sur le modèle d'un son x considéré comme une excitation e filtrée par un filtre h ($x(t) = e \star h(t)$), on arrive facilement à séparer ces deux composantes e et h grâce aux MFCC en considérant respectivement les coefficients DCT de basses fréquences ou de hautes fréquences. De plus ces derniers ont tendance à décorreler les coefficients du Mel-spectre et à concentrer la variance des coefficients du spectre dans les coefficients MFCC de basses fréquences DCT.

Dans notre contexte est en nous basant sur [ADP07], nous avons utilisé des fenêtres d'analyse de 2048 échantillons et un recouvrement des fenêtres de 50%.

4.2 Mesures de distance entre sons

Les algorithmes correspondant à l'état de l'art sont l'algortihme « Bag of Frame (BOF » de J-J. Aucouturier ([ADP07]) et l'algorithme de « Dynamic Time Warping (DTW) » de Dan Ellis ([TE03]).

4.2.1 Bag of Frame (BOF)

L'algorithme BOF ([ADP07]) s'appuie sur des Gaussian Mixture Model (GMM) ([Rey09]) pour calculer une représentation intermédiaire des sons en utilisant les représentations que nous venons de voir, c'est-à-dire des listes de vecteurs fréquence ordonnés dans le temps. L'algorithme BOF considère ces vecteurs fréquence sans tenir compte de l'ordre temporel (ce qui a priori est une lacune du procédé), comme des réalisations indépendantes d'une variable aléatoire. L'algorithme estime donc dans un premier temps, pour chaque son les paramètres d'un modèle GMM à $\mathcal{M}$ gaussiennes :

$$p(x_t) = \sum_{m=1}^{\mathcal{M}} \pi_m \mathcal{N}(x_t, \mu_m, \Sigma_m) \tag{4.1}$$

où π_m est le coefficient de contribution de chaque gaussienne au modèle, et $\mathcal{N}(x_t, \mu_m, \Sigma_m)$ une gaussienne de moyenne μ_m et variance Σ_m appliquée à l'instant x_t. L'estimation des paramètres se fait à travers un algorithme E-M classique.

Chaque son est donc représenté par un modèle GMM. La distance entre modèles est calculée grâce à une approximation de la divergence de Kullback-Leibler (KL) à l'aide d'un algorithme de Monte-Carlo. Si l'on note p_A et p_B les densités de probabilité associées à deux modèles GMM A et B respectivement, on tire n points aléatoires $(x_i)_{i\in[1,n]}$ et l'approximation de la divergence KL entre A et B est donnée par la formule :

$$\tilde{d}_{KL}(A, B) = \frac{1}{n} \sum_{i=1}^{n} \frac{p_B(x_i)}{p_A(x_i)} \tag{4.2}$$

Afin d'obtenir une distance, on moyenne les divergences $d_{KL}(A, B)$ et $d_{KL}(B, A)$.

Dans notre contexte, nous trouverons certaines difficultés à l'application de cet algorithme. En effet nos sons sont très courts et l'estimation des paramètres des GMM doit se baser par conséquent sur peu d'échantillons. C'est pour cela que nous avons choisi d'utiliser 1, 2 ou 3 gaussiennes par modèle. De plus la dimension de l'espace de représentation varie beaucoup entre la représentation en spectrogramme, Mel-spectre et MFCC. En effet plus la dimension est grande, plus l'algorithme a besoin d'échantillons pour estimer les paramètres de manière cohérente. Le spectrogramme étant la mesure produisant un espace de dimension la plus importante, nous avons appliqué l'algorithme BOF uniquement sur les représentations Mel-spectre et MFCC (comme dans [ADP07]).

4.2.2 Dynamic Time Warping (DTW)

L'algorithme DTW ([TE03]), contrairement à BOF, permet de tenir compte de l'ordre temporel des vecteurs fréquences ainsi que d'éventuelles déformations dans le temps. Celui-ci calcule dans un premier temps une matrice de similarité $S = (s_{ij})$ entre deux sons A et B où s_{ij} est la similarité entre le vecteur fréquence de A à l'instant i (e_i) et le vecteur fréquence de B à l'instant j (e_j), calculée selon la formule :

$$s_{ij} = \frac{< e_i | e'_j >}{\|e_i\| \cdot \|e'_j\|} \tag{4.3}$$

où $< x|y >$ désigne le produit scalaire usuel.

Une fois la matrice de similarité calculée entre chaque paire de sons, la distance entre deux sons de longueurs n et m est donnée par le coût du chemin optimal dans la matrice de similarité correspondante entre le premier coefficient s_{11} et le dernier coefficient s_{nm}.

Le calcul du chemin optimal se fait de manière récursive. Appelons C_{ij} le coût du chemin minimal allant de de s_{11} à s_{ij}. Alors en prenant comme condition initiale $C_{11} = s_{11}$ on a :

$$C_{ij} = s_{ij} + min(C_{i-1,j}, C_{i-1,j-1}, C_{i,j-1}) \tag{4.4}$$

Le coût du chemin optimal est donc donné par le coefficient C_{nm}.

Chapitre 5

Algorithme proposé

Dans ce chapitre nous présenterons dans un premier temps la représentation scattering et dans un deuxième temps les mesures de distance que nous avons utilisées sur cette représentation. Nous commencerons par exposer les raisons qui nous ont poussés à utiliser la représentation scattering et nous présenterons ensuite la représentation scattering d'un point de vue formel ainsi que deux extensions de cette représentation : CLS et scattering combined. Enfin dans la section sur les mesures de distance, nous mettrons l'accent sur une mesure de distance inhabituelle : la distance de Spearman.

5.1 Motivations

L'intérêt pour nous de la représentation scattering vient de ses propriétés d'invariance ou stabilité par déformation. Comme nous l'avons évoqué au début du rapport, une « bonne » représentation d'un son est une représentation qui entre autre contient l'information utile pour une tâche donnée et qui a éliminé le reste de l'information. Dans la recherche d'une mesure de similarité entre deux sons, un premier type d'information que l'on veut éliminer du signal est la variance par translation temporelle. En effet on aimerait que deux sons translatés dans le temps l'un par rapport à l'autre mais parfaitement identiques par ailleurs aient la même représentation. Cette première propriété est apportée par la représentation de Fourier. En appliquant la transformée de Fourier on peut obtenir une représentation pareille à celle de la figure 5.1.

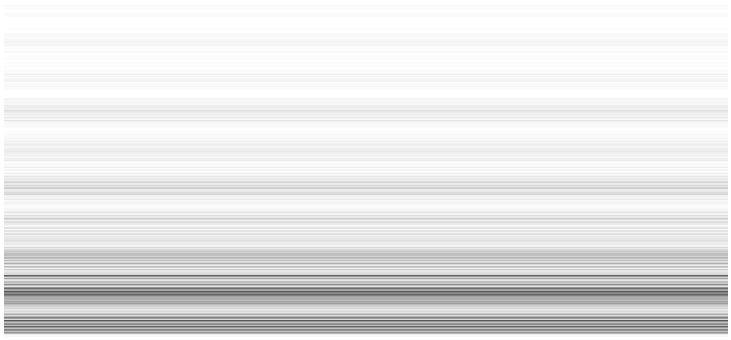


FIGURE 5.1 – Représentation temps-fréquence d'une transformée de Fourier appliquée sur la totalité d'un signal (ici, un bruit de sirène issu de la base Gygi)

On voit clairement que pour assurer l'invariance par translation temporelle, on a éliminé la notion de temps dans la représentation et l'on a perdu de cette façon toutes les variations et

modulations fréquentielles qu'il ya pu y avoir au cours du temps. Pour les récupérer, on peut utiliser une transformée de Fourier avec des fenêtres d'analyse plus petites, on obtient alors sur le même signal que précédemment quelque chose ressemblant à la figure 5.2. Mais si la propriété d'invariance par translation temporelle est respectée à l'intérieur de chaque fenêtre d'analyse, nous l'avons perdue à plus grande échelle sur le signal et le même signal translaté dans le temps aura une représentation temps-fréquence également translatée.

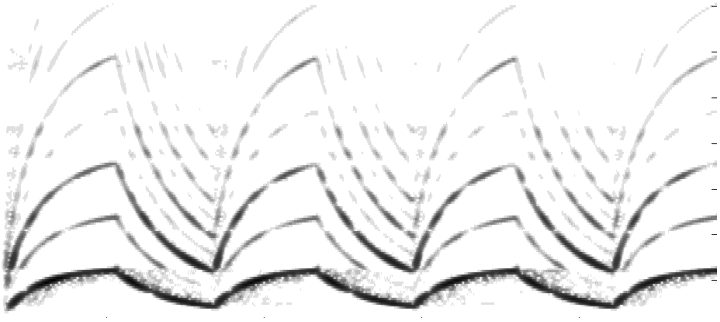


FIGURE 5.2 – Représentation temps-fréquence d'une transformée de Fourier à fenêtres sur un bruit de sirène issu de la base Gygi

Comme nous le verrons, la représentation scattering permet de garder la propriété d'invariance par translation temporelle tout en gardant l'information des modulations fréquentielles.

Une deuxième propriété que l'on aimerait avoir dans nos représentation est une propriété de stabilité par translation fréquentielle. En effet, dans le cas de signaux harmoniques par exemple, lorsque deux sons harmoniques de fréquence fondamentale proche sont jouées (comme sur la figure 5.3(a)), les spectres résultant des deux signaux sont très différents en hautes fréquences alors que perceptivement, les deux sons sont très proches.

(a) Représentation sur en spectre à largeur de bandes constantes

(b) Représentation sur un spectre en bande de Mel

FIGURE 5.3 – Exemple de spectre de deux sons harmoniques de fréquences fondamentales légèrement différentes.

Une solution peut être par exemple d'utiliser le Mel-spectre au lieu d'un spectre usuel. Les bandes plus larges en hautes fréquences permettent d'obtenir des représentations plus proches, comme le montre la figure 5.3(b). Nous verrons que la représentation scattering permet d'obtenir le même type de stabilité par translation fréquentielle.

5.2 Représentation scattering ([AM11])

D'un point de vue formel la représentation scattering est une représentation hiérarchique d'un signal (on parle de scattering à l'ordre 1, 2, etc...) calculée à l'aide d'un opérateur U_J utilisant un banc de filtres en ondelettes. Cet opérateur est appliqué en cascade n fois sur le signal pour obtenir la représentation à l'ordre $n-1$. D'un point de vue pratique, la représentation à l'ordre 1 ressemble à une représentation en bandes de Mel du signal et l'objectif des ordres suivants et de récupérer l'information qui n'a pas été captée par l'ordre 1.

5.2.1 Présentation formelle de l'opérateur

On construit tout d'abord un banc de filtre en ondelettes $(\psi_j)_{j<J+P}$ à partir d'une ondelette mère ψ, d'une largeur de bande en octaves de $1/Q$, que l'on va dilater. En fréquence, cette ondelette mère couvre environ l'intervalle $[2Q\pi - \pi, 2Q\pi + \pi]$.

Les J filtres de plus haute fréquence sont construits de la façon suivante :

$$\psi_j(t) = a^{-j}\psi(a^{-j}t) \text{ où } a = 2^{1/Q} \text{ et } j \leqslant J$$

On obtient ainsi des ondelettes dont le support temporel est croissant avec j. Elles sont de plus en plus basses fréquences avec j croissant, couvrant l'intervalle $[2Q\pi a^{-j} - \pi a^{-j}, 2Q\pi a^{-j} + \pi a^{-j}]$ (donc de plus en plus petit avec j croissant. Le facteur a^{-j} devant ψ sert à avoir des ondelettes de même amplitude en transformée de Fourier. L'ondelette de plus basse fréquence est alors centrée en fréquences sur $2\pi a^{-J}$.

Pour couvrir les plus basses fréquences, on construit P filtres (pour $j \in [J, J+P[$ où $P = \lfloor \frac{Q}{2} - 1 \rfloor$) espacés linéairement, de même support temporel ($2Qa^J$) et de même largeur de bande ($2\pi a^{-J}$) que l'ondelette ψ_J.

Enfin les basses fréquences qui ne sont pas couvertes par ces P derniers filtres sont couvertes par un filtre passe-bas Φ_J couvrant l'intervalle fréquentiel $[-\pi a^{-J} \pi a^{-J}]$.

On peut alors définir l'opérateur U_J applicable à un signal $x(t)$:

$$U_J x(t) = \begin{pmatrix} x \star \Phi_J(t) \\ |x \star \psi_j(t)| \end{pmatrix}_{j<J+P} \tag{5.1}$$

L'opérateur va donner une sortie : le signal moyenné par le filtre Φ_J et va également calculer un spectrogramme multi-échelle correspondant à l'application du banc de filtre $(\psi_j)_{j<J+P}$ sur le signal et ensuite l'application du module sur ce résultat. Ce dernier calcul n'est pas sorti et ne sert qu'au calcul des ordres suivants. Ainsi les ordres suivants bénéficient d'un échantillonnage fin.

5.2.2 Calcul de la représentation

Ainsi l'ordre 0 est simplement $x \star \Phi_J$. La figure 5.4 schématise le calcul de l'ordre 1 : on applique le banc de filtre en ondelettes au signal x et on calcule le module. Ceci nous donne une représentation temps-fréquence multi-échelle à laquelle on applique le filtre passe-bas Φ_J. On obtient ainsi la représentation scattering à l'ordre 1.

On peut remarquer à ce stade que l'ordre 1 peut ressembler fortement à un spectre de Mel si l'on remplace le banc de filtre en ondelettes par un banc de filtre de Mel. Ce que l'on a perdu à l'ordre 1, soit ce qui a été brouillé par le filtre Φ_J est l'information que l'on va récupérer aux ordres suivants.

Pour calculer l'ordre 2, on considère chaque bande de fréquence du spectrogramme multi-échelle (c'est-à-dire les $|x \star \psi_j|$) comme un nouveau signal auquel on peut appliquer le même

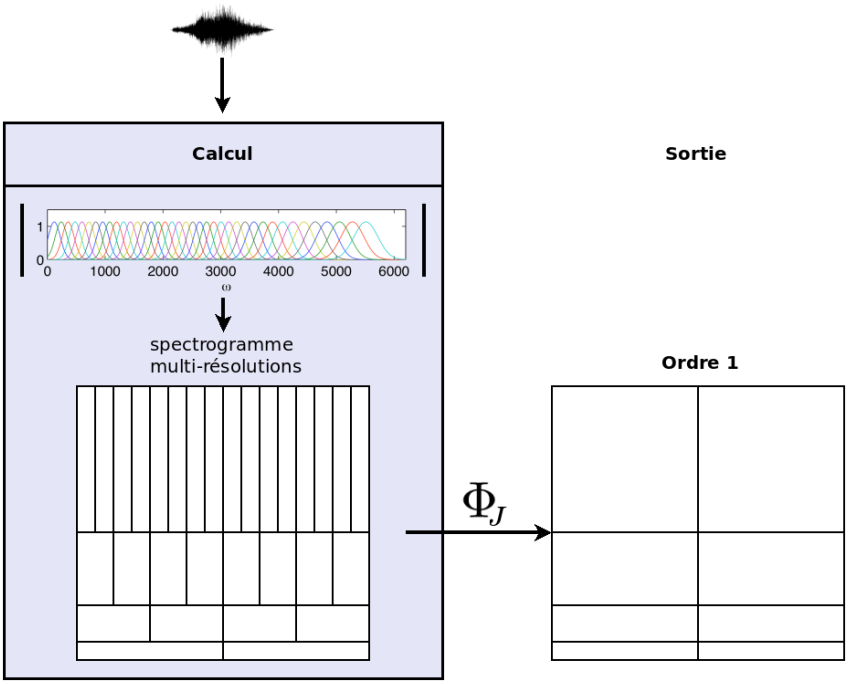


FIGURE 5.4 – Schéma du calcul de la représentation scattering à l'ordre 1. Le calcul interne est représenté à gauche en bleu clair et la sortie est située à droite.

type d'opération. Ainsi la figure 5.5 schématise le calcul de l'ordre 2. À partir du spectrogramme multi-échelle, on applique à chaque bande le banc de filtre en ondelettes suivi du module. Cela nous donne pour chaque bande de fréquence du spectrogramme d'origine (soit chaque $|x \star \psi_j|$) un nouveau spectrogramme multi-échelle que l'on « moyenne » en temps grâce au filtre Φ_J. L'ordre 2 est donc constitué de $J + P$ spectrogrammes multi-échelle.

À l'ordre 2 on peut résumer la représentation scattering d'un signal x par la formule suivante :

$$S_J x(t) = \begin{pmatrix} x \star \Phi_J(t) \\ |x \star \psi_{j_1} \star \Phi_J(t) \\ ||x \star \psi_{j_1}| \star \psi_{j_2}| \star \Phi_J(t) \end{pmatrix}_{j_1, j_2 < J+P} \tag{5.2}$$

On remarque que le nombre de coefficients augmente très rapidement, cependant on peut vérifier (voir [**?**]) que si $j_2 < j_1 + log_a(Q/2)$ alors $||x \star \psi_{j_1}| \star \psi_{j_2}| \star \Phi_J(t) \approx 0$. Ainsi il est inutile de considérer les coefficients qui vérifient cette condition. De cette façon, pour un signal de N échantillons, on obtient environ $Qlog_2(N/Q)$ coefficients d'ordre 1 et $Q^2/2log_2^2(N/Q^2)$ coefficients d'ordre 2.

Comme le montre Anden et alii dans [**?**], le fait de considérer les ordres supérieurs d'un signal, permet de capturer de mieux en mieux toute l'énergie du signal. Sur la base de sons (GTZAN [ToCS02]) utilisée par Anden et alii, l'énergie du signal capturée par l'ordre 1 est de 73.0% et celle capturée par l'ordre 2 de 98.1% lorsqu'on utilise une fenêtre d'analyse de 5.9s. Ceci nous dit qu'il ne sert pas a priori d'aller au delà de l'ordre 2 (mais ce n'est pas sûr, car l'information utile pourrait être contenue dans les 1.9% d'énergie du signal restants) et nous permet d'utiliser des fenêtres d'analyse longues sans pour autant perdre trop d'information. Nous nous sommes donc réduits à la représentation scattering à l'ordre 2 et nous avons utilisé des fenêtres d'analyse longues (quelques secondes) couvrant la totalité du signal. Dans la suite nous ne considérons donc plus que les deux premiers ordres scattering.

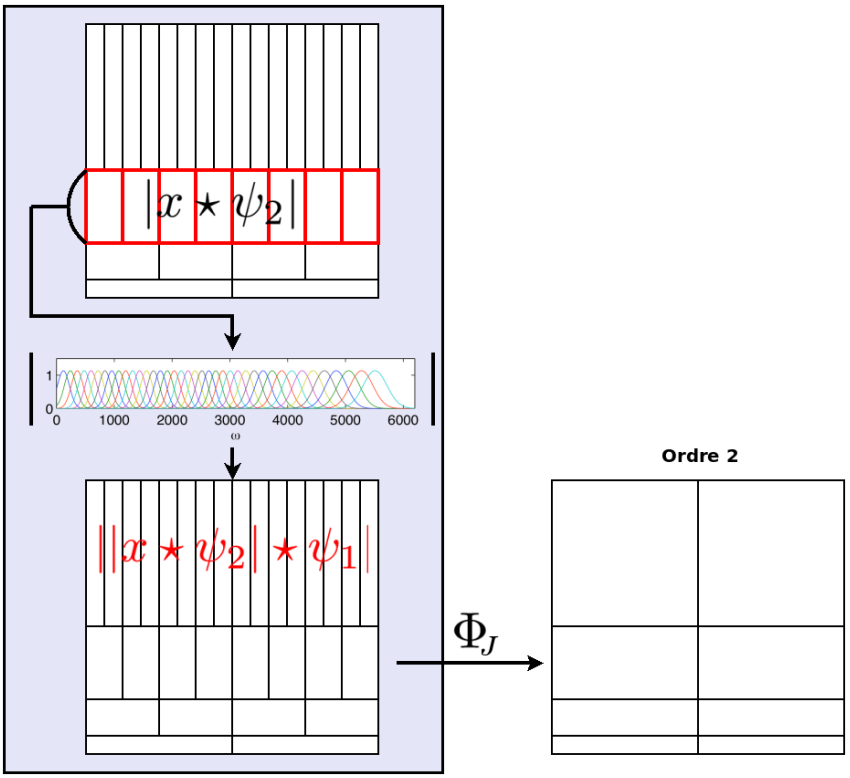


FIGURE 5.5 – Schéma du calcul de la représentation scattering à l'ordre 2 de $|x \star \psi_2|$.

5.3 Cosine Log Scattering (CLS)

Comme nous l'avons vu au chapitre précédent, la représentation en MFCC est utilisée dans de nombreuses applications audio, en particulier celles traitant des sons musicaux ou vocaux. En effet ce type de son peut être modélisé par une source $e(t)$ filtrée par $h(t)$. Les MFCC permettent de séparer facilement ces deux composantes e et h. La représentation CLS est à la représentation scattering ce que les MFCC sont au Mel-spectre.

Nous avons vu que l'ordre 1 de scattering ressemble au Mel-spectre qui sert de base au calcul des MFCC. Ainsi on peut retrouver une représentation similaire aux MFCC à l'aide de la représentation scattering à l'ordre 1. Pour cela il suffit d'appliquer une DCT au logarithme des coefficients scattering d'ordre 1.

Un processus similaire permet de séparer partiellement les composantes e et h dans les coefficients scattering d'ordre 2. Pour expliquer ce processus, plaçons nous à une fenêtre temporelle donnée. Les coefficients scattering à l'ordre 1 et 2 peuvent être représentés alors comme décrit sur la figure 5.6.

Les coefficients d'ordre 1 peuvent être représentés comme un vecteur fréquence indexé par i. À l'ordre 2 chaque coefficient est indexé à la fois par i (index du premier filtre) et par j (correspondant à l'index du second filtre). Pour obtenir la représentation CLS à l'ordre 2 on calcule le logarithme des coefficients d'ordre 2 et l'on applique une DCT d'abord sur les colonnes (c'est-à-dire selon j) puis sur les lignes (selon i). On obtient alors une matrice paramétrée par deux indices de fréquences DCT.

Comme l'indique [AM11], la représentation CLS décorrèle bien les coefficients scattering et permet de concentrer l'énergie du signal sur un plus petit nombre de coefficients, les coefficients basses-fréquences DCT. Ainsi, comme c'est souvent le cas pour les MFCC, seuls les premiers

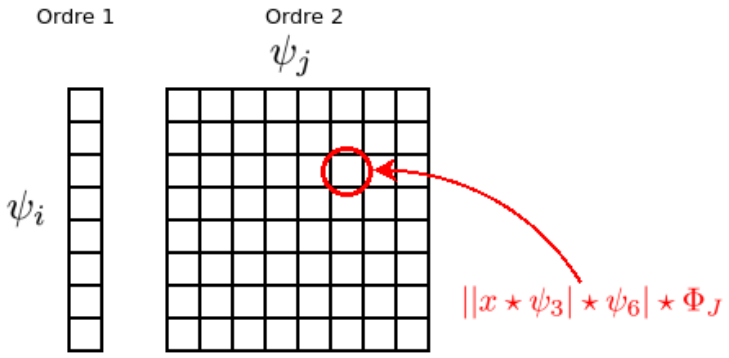


FIGURE 5.6 – Schéma représentant les coefficients scattering à l'ordre 1 et 2 pour une fenêtre temporelle donnée.

coefficients (basses fréquences DCT) sont porteurs d'information utile. Nous avons donc introduit un paramètre k exprimé en pourcentage du nombre total de coefficients, définissant un masque retenant les k premiers pourcentages des coefficients CLS.

5.4 Scattering « Combined »

Suite aux résultats de premières expérimentations de la représentation CLS, nous avons ressenti le besoin de chercher une représentation ayant une forme plus forte de stabilité par translation fréquentielle. Celle-ci s'est traduite par la représentation scattering « combined » qui propose une forme de stabilité par translation fréquentielle à la fois plus forte et paramétrable.

Comme pour CLS, nous partons de la représentation scattering à l'ordre 2 et nous nous replaçons à instant t_1. Soit y_{t_1} le vecteur constitué des coefficients scattering d'ordre 1. On peut alors le considérer comme un signal s'écrivant $y(n) = |x \star \psi_n| \star \Phi_J(t_1)$ et lui appliquer l'algorithme scattering à l'ordre 1. Nous choisissons pour paramétrer cette deuxième application de l'algorithme d'utiliser $Q = 1$ comme facteur de qualité. On obtient ainsi à partir des coefficients d'ordre 1 un spectrogramme (voir figure 5.7) où l'axe « temporel » représente en réalité les fréquences du signal de départ et l'axe fréquentiel représente les « fréquences des fréquences ».

On remarque que de cette façon, on a construit une représentation invariante par translation fréquentielle, par bandes de fréquences, ces dernières étant déterminées par la famille de filtres en ondelette du deuxième algorithme scattering. Sur la figure 5.7, le spectrogramme est constitué de deux fenêtres à l'intérieur desquelles l'invariance par translation fréquentielle est assurée. On comprend donc que le choix de la taille des fenêtres est important pour paramétrer la sévérité avec laquelle notre représentation doit être invariante par translation fréquentielle. À un extrême, soit une seule fenêtre, l'invariance est totale, la représentation est totalement délocalisée en fréquence. À l'autre extrême, si les fenêtres sont extrêmement courtes, il n'y aura presque aucune invariance par translation fréquentielle.

Nous avons appliqué un procédé analogue sur les coefficients d'ordre 2 en considérant les vecteurs d'ordre 2 paramétrés par i et j (indices des filtres à l'ordre 1 et 2 respectivement) à j fixé. L'application du deuxième algorithme de scattering ce fait donc sur les signaux $y_{j,t_1}(n) = ||x \star \psi_n| \star \psi_j| \star \Phi_J(t_1)$. Comme pour l'ordre 1, cette deuxième application de l'algorithme scattering est caractérisée par un facteur de qualité $Q = 1$.

D'un point de vue qualitatif, on autorise de cette façon une plus grande souplesse dans la façon dont les fréquences (ou formants ou harmoniques) varient au cours du temps.

Enfin, pour nos expériences, la représentation scattering combined sera paramétrée par le facteur de qualité Q de la représentation scattering et par le nombre de fenêtres fréquentielles

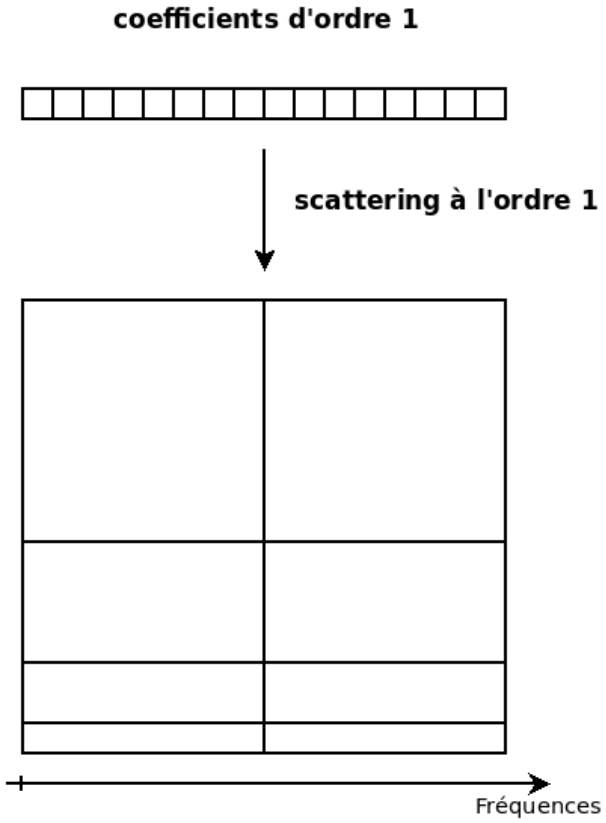


FIGURE 5.7 – Schéma de l'application de l'algorithme scattering à l'ordre 1 sur les coefficients d'ordre 1. Les fréquences

T_{co} déterminant la sévérité de l'invariance par translation fréquentielle que l'on veut atteindre. Plus T_{co} sera grand, moins l'invariance sera sévère.

5.4.1 Rappel sur les particularités d'implémentation

Dans notre contexte, nous rappelons que nous avons choisi d'appliquer l'algorithme de scattering sur la totalité du signal en gardant Q comme paramètre à faire varier. La représentation est « éclatée » sous forme de vecteur contenant à la fois les coefficients scattering d'ordre 1 et 2. Afin de s'assurer que tous les sons d'une même base auront un vecteur de représentation de même taille nous avons ajouté des zéros à la fin des sons trop courts pour que tous aient la même taille.

Concernant la représentation CLS, nous avons introduit le paramètre k définissant le nombre de coefficients plus basses fréquences DCT à considérer.

Concernant la représentation combined, nous avons introduit le paramètre T_{CO} donnant une façon d'établir le degré d'invariance par translation fréquentielle souhaité.

5.5 Mesures de distance

Après avoir introduit les représentations scattering, nous exposons maintenant les métriques de distances entre ces représentations. La représentation sous forme de vecteurs de même taille pour tous les sons d'une même base nous permet d'utiliser des distances entre vecteurs usuelles en commençant par les plus simples : distance euclidienne, distance « valeur absolue de la diffé-

rence » – que l'on appellera cityblock – et distance en cosinus. Dont les formules sont rappelées ci-dessous pour deux vecteurs u et v :

$$\text{euclidienne} : d_e(u,v) = \sqrt{\sum_i (u_i - v_i)^2} \tag{5.3}$$

$$\text{cityblock} : d_{cy}(u,v) = \sum_i |u_i - v_i| \tag{5.4}$$

$$\text{cosinus} : d_{co}(u,v) = 1 - \frac{\sum_i u_i v_i}{\sqrt{\sum_i u_i^2}\sqrt{\sum_i v_i^2}} \tag{5.5}$$

Nous avons également utilisé une quatrième distance, la distance de Spearman que nous présentons ci-après.

5.5.1 Distance de Spearman

La distance de Spearman est issue du coefficient de corrélation de Spearman entre deux variable aléatoires X et Y. Si l'on considère deux vecteurs $(X_i)_{i\in[1,n]}$ et $(Y_i)_{i\in[1,n]}$ de n tirages de ces variables aléatoires et l'on appelle $x = (x_i)_{i\in[1,n]}$ et $y = (y_i)_{i\in[1,n]}$ les vecteurs des rangs des coefficients X_i et Y_i respectivement alors la corrélation de Spearman s'écrit :

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{5.6}$$

D'un point de vue qualitatif, le fait d'utiliser le rang plutôt que les coefficients en eux-mêmes permet de calculer la corrélation de deux variables aléatoires reliées de façon monotone, et non uniquement de façon linéaire comme une corrélation de Pearson plus classique.

Étant donné que la corrélation prend des valeurs entre -1 et 1, la distance de Spearman s'écrit $1 - \rho$ et prend donc des valeurs entre 0 et 2.

En rapport avec nos représentations de signaux, calculer la corrélation entre deux vecteurs fréquence permet d'être invariant aux homothéties comme pour la distance cosinus et d'être plus stable aux déformations du vecteur qui conservent le rang.

Dans la mesure où la distance de Spearman utilise le rang des vecteurs, le nombre de vecteurs possible est beaucoup plus réduit et l'on peut se retrouver facilement avec des vecteurs ayant le même rang, en particulier si le nombre de coefficients est petit. Ainsi, si une matrice de distances entre les sons d'une base peut contenir des doublons et nous avons vu que dans ce cas l'utilisation du MAP comme métrique d'évaluation des matrices de distance doit être faite avec précaution (voir section 3.2).

Chapitre 6

Résultats et analyse

Nous présentons dans ce dernier chapitre les résultats de nos expérimentations sur les bases de sons (rappelées sur le tableau 6.1), ordonnés par algorithme de distance. Nous commencerons par présenter les résultats aléatoires comme repère puis nous présenterons les résultats pour les algorithmes état de l'art BOF et DTW ainsi que les résultats de l'application de distances usuelles sur des représentations MFCC. Enfin nous présenterons les résultats des représentations CLS et scattering combined couplées aux distances usuelles. Le tableau 6.14 résume ces différentes sections.

Nous rappelons que les résultats sont calculés à l'aide du MAP (voir le chapitre 3) et sont écrits sous forme de pourcentage. Pour la base Houix2 qui peut être étiquetée par trois niveaux de classes, nous écrirons les résultats pour ces trois niveaux de classes en nommant Houix23 par exemple les résultats de la base Houix2 pour le 3ème niveau de classes.

Enfin, comme nous l'avons vu dans le chapitre 3, lorsqu'on utilise le MAP il faut faire attention aux éventuels doublons dans les matrices de distance. Si cela n'est pas un problème pour la plupart des couples distance-représentation que nous utilisons, nous avons remarqué que des doublons pouvaient être présents lorsqu'on utilise la distance de Spearman. Afin de contourner ce problème nous avons appliqué la solution décrite au chapitre 5.5.1. Malgré cela son interprétation reste délicate, c'est pourquoi, afin de garder une certaine clarté dans l'analyse des résultats, nous ne tiendrons pas compte de cette distance dans ce qui suit.

	nombre de classes	nombre d'éléments	MAP aléatoire
Bases de sons environnementaux			
Gygi	50	100	5.1
GygiExt	50	294	3.6
Houix1	4	60	43.6
Houix2	(3,10,13)	56	(39.4,20.6,14.7)
Bases de sons instrumentaux			
Iowa	15	3637	8.4
RWC	14	6138	8.9
SolosDb	19	505	8.7
Base de parole			
Digit	11	347	10.3
Autre base			
Insects	5	500	20.8

Table 6.1 – Rappel des caractéristiques des bases de sons

Nous avons calculé le MAP d'une matrice aléatoire, moyenné sur 100 tirages, pour chaque base. Les résultats sont présentés sur le tableau 6.1.

Les résultats varient beaucoup d'une base à l'autre à cause de leur morphologie (nombres d'éléments et nombre d'éléments par classe). D'une manière générale plus le nombre de classes est faible comme pour Houix1 et Houix21 (respectivement 4 et 3 classes) plus le MAP est élévé.

6.1 Résultats des distances usuelles sur les MFCC

Afin de mieux pouvoir interpréter les résultats de scattering uniquement du point de vue de la qualité de la représentation, nous avons calculé les résultats pour les mesures de distance usuelles – qui seront utilisées avec les représentations scattering CLS et combined) avec une représentation en MFCC.

Pour chaque base, nous avons donc calculé la représentation en MFCC en prenant les 13 premiers coefficients ou tous les coefficients (soit 32 coefficients), puis nous avons calculé les distances entre ces représentations à l'aide des distances euclidienne, cityblock, cosinus et spearman. Les MFCC sont calculés sur des fenêtres de 2048 échantillons et un recouvrement de 50%.

Dans ce contexte nous avons mené deux expériences. Pour la première nous avons calculé la moyenne et la variance des vecteurs de représentation de chaque son et nous avons calculé la distance (euclidienne) entre les moyennes d'une part et la moyenne des divergences de Kullback-Leibler (KL) entre les gaussiennes caractérisées par la moyenne et la variance de chaque représentation MFCC d'autre part. Pour deux gaussiennes en dimension k $G_1(\mu_1, \Sigma_1)$ et $G_2(\mu_2, \Sigma_2)$ la divergence KL s'écrit :

$$d_{KL}(G_1, G_2) = \frac{1}{2}(tr(\Sigma_1^{-1}\Sigma_2) + (\mu_1 - \mu_2)^T\Sigma_1^{-1}(\mu_1 - \mu_2) - ln(\frac{det\Sigma_2}{det\Sigma_2}) - k)) \tag{6.1}$$

La distance entre G_1 et G_2 est donc calculée en moyennant $d_{KL}(G_1, G_2)$ et $d_{KL}(G_2, G_1)$. Nous avons utilisé la variance uniquement des coefficients MFCC, les matrices Σ_i sont donc diagonales. Sur les 4 bases de sons environnementaux nous avons également effectué la même expérience avec la covariance des coefficiens MFCC au lieu de la variance. Cependant les résultats sont similaires et légèrement moins élevés que lorsque la variance simple est utilisée. Cela peut être du à l'imprécision du calcul de l'inversion des matrices de covariance. C'est pourquoi nous n'avons pas donné suite à ces calculs pour les autres bases et nous ne rapportons ici que les résultats pour la variance simple.

Les résultats sont présentés dans le tableau 6.2.

Les résultats sont la plupart du temps légèrement au-dessus des performances aléatoires, les deux extrêmes étant la base Houix2 qui reste aux performances aléatoires et les base d'instruments qui s'écartent le plus des performances aléatoires.

La deuxième expérience a consisté à calculer directement les distances entre les moyennes des vecteurs de représentation MFCC à 13 ou 32 coefficients. Les résultats sont résumés dans le tableau 6.3.

On remarque qu'ici les performances sont en général meilleures que pour la première expérience. Celles-ci donnent une meilleure base de comparaison pour les résultats scattering. En effet les mêmes mesures de distances sont utilisées, seule la représentation change et l'on devrait retrouver des résultats similaires pour la représentation d'ordre 1 car l'ordre 1 ressemble aux MFCC.

	MFCC 13		MFCC 32	
	moy	moy+var	moy	moy+var
gygi	22.7	28.0	23.5	**31.6**
gygiExt	21	27	20.9	**27.9**
houix1	47.5	**50.0**	47.5	48.8
houix21	39.5	38.8	39.6	39.2
houix22	19.7	20.2	19.6	**20.5**
houix23	12.8	14.5	12.7	13.1
iowa	23.4	28.2	28.4	21.1
rwc	30.2	28.4	28.7	22.9
solosDb	35.4	**36.8**	33.8	31.0
digit	45.4	41.0	45.1	28.3
insects	29.0	34.7	30.5	**35.5**

TABLE 6.2 – Résultats de la première expérience en représentation MFCC (les meilleurs résultats par base, sur les deux expériences sont notés en gras).

	MFCC 13				MFCC 32			
distances	eucl	city	cos	spear	eucl	city	cos	spear
gygi	22.7	26.8	26.9	17.7	23.5	24.0	27.0	17.7
gygiExt	20.5	20.9	19.5	15.8	20.9	21.4	20.2	0.0
houix1	47.5	48.3	43.6	44.6	47.5	48.4	43.5	43.8
houix21	39.5	39.2	**40.2**	39.9	39.6	39.6	40.1	37.8
houix22	19.7	20.2	20.2	21.3	19.6	20.3	20.2	19.6
houix23	12.8	13.3	13.7	**15.9**	12.7	12.9	13.5	16.1
iowa	**32.6**	31.5	29.0	27.8	28.4	25.2	25.6	17.7
rwc	30.2	**30.5**	28.5	27.8	28.7	27.3	26.9	19.7
solosDb	35.4	36.4	34.5	32.9	33.8	32.7	33.0	19.4
digit	45.4	45.2	**47.4**	41.9	45.1	43.3	47.2	21.3
insects	29.0	29.2	31.4	29.0	30.5	30.9	34.1	32.3

TABLE 6.3 – Résultats pour chaque couple représentation MFCC à 13 ou 32 coefficients et distance euclidienne (eucl), cityblock (city), cosinus (cos) ou spearman (spear)

6.2 Résultats pour l'algorithme BOF

Nous avons appliqué l'algorithme BOF à 1, 2 ou 3 gaussiennes par modèle, couplé aux représentations en Mel-spectre et en MFCC[1].

En effet, les sons de nos bases étant courts (quelques secondes) soit naturellement soit à cause d'une segmentation préalable, les modèles GMM ont peu d'échantillons pour estimer les paramètres des modèles et utiliser plus de 1 ou 2 gaussiennes mène à de moins bons résultats. Seule la base Houix2 possède des sons assez longs pour que l'utilisation de 3 gaussiennes donne de meilleurs résultats. Enfin nous n'utilisons pas le spectrogramme comme représentation car le nombre de dimensions est trop important et il faudrait utiliser des fenêtres d'analyse trop longues et donc perdre trop d'information.

Le tableau 6.4 présente les résultats par base et en fonction du nombre de gaussiennes et de la représentation utilisée. Il faut noter que cet algorithme n'a pas été utilisé sur la base Insects

1. L'implémentation de BOF est détaillée au chapitre 4

Rep		Log			MFCC		
NbGauss		1	2	3	1	2	3
Bases	gygi	29.4	20.8	21.8	30.8	30.1	**31.8**
	gygiExt	23.2	19.1	20.9	**27.0**	25.3	25.4
	houix1	**54.6**	53.7	51.1	50.9	50.8	50.2
	houix21	39.4	42.2	43.7	39.2	38.9	**46.3**
	houix22	19.8	21.8	**27.6**	20.6	21.0	23.3
	houix23	13.0	15.2	**25.2**	14.1	15.1	20.9
	iowa	28.3	23.2	21.1	28.9	**29.8**	
	rwc	**30.0**			28.8	28.1	
	solosDb	26.5	27.1	27.0	37.6	38.7	**39.7**
	digit	31.3	33.4	31.9	**45.9**	41.0	33.7
	insects						

TABLE 6.4 – Résultats de l'algorithme BOF à 1, 2 ou 3 gaussiennes, couplé à la représentation Mel-spectre (noté log) ou MFCC (noté mfcc). Les cases vides correspondent à des cas où l'algorithme n'a pas convergé.

car les sons ont une durée trop petite (<1s). Pour d'autres bases, certaines cases du tableau valent zéro. Il s'agit de cas où l'algorithme n'a pas donné de résultats pertinents à cause d'un trop grand nombre de dimensions par rapport à la taille des sons (soit le nombre d'échantillons pour estimer les paramètres du modèle).

6.3 Résultats pour l'algorithme DTW

Nous avons appliqué l'algorithme DTW aux bases de sons en utilisant les trois types de représentation usuels : spectrogramme, Mel-spectre et MFCC [2]. Pour des raisons de temps de calcul du coût du chemin optimal dans la matrice de similairté, les bases plus volumineuses n'ont pas été traitées avec la représentation spectrogramme.

	Spec	*Log*	*MFCC*
Gygi	**25.8**	20.7	20.4
GygiExt	15.2	**19.3**	13.9
*Houix*1	54.9	**55.5**	52.5
*Houix*21	**40.6**	39.8	39.4
*Houix*22	**20.9**	20.3	20.4
*Houix*23	13.2	**14.5**	14.3
Iowa	0	**32**	30.6
RWC	0	**30.2**	25.5
SolosDb	26.7	**33.8**	32.4
Digit	34.1	**70.1**	54.6
Insects	30.3	**42.1**	28.9

TABLE 6.5 – Résultats de l'algorithme DTW en fonction de la représentation utilisée

On remarque que ce dernier algorithme est moins performant pour les bases environnementales que l'algorithme BOF mais plus performant pour les bases instrumentales, en particulier

2. L'implémentation de DTW est détaillée au chapitre 4

les bases contenant des notes isolées. Enfin les performances sont également élevées pour la base de parole Digit.

6.4 Résultats pour CLS

6.4.1 Présentation

Comme nous l'avons vu à la section 5.3, l'apport de ces représentations concerne l'invariance par translation temporelle et la stabilité par translation fréquentielle que l'on obtient. Lorsqu'on calcule la représentation jusqu'au deuxième ordre, ces caractéristiques nous permettent de choisir des fenêtres d'analyse longues sans perdre d'information sur les structures temporelles à l'intérieur de la fenêtre. Nous distinguerons donc la représentation à l'ordre 1 et à l'ordre 2.

Dans ce contexte, le calcul de la représentation CLS se fait de la manière suivante. On considère une fenêtre (temporelle) d'analyse englobant la totalité du signal, de taille T, pour tous les sons d'une même base. Pour que T reste fixe alors que les sons sont de longueurs différentes, on ajoute autant de zéros que nécessaire à la fin de chaque son. On pose le facteur de qualité Q en paramètre et à partir de T et Q on détermine le nombre de filtres logarithmiques (J) et le nombre de filtres linéaires (P). Comme nous l'avons décrit à la section 5.3, nous utiliserons de plus un masque sur les coefficients scattering paramétré par k, écrit en pourcentage du nombre total de filtres (K). Ce paramètre caractérise le masque à la fois de l'ordre 1 et de l'ordre 2.

En résumé la représentation CLS est un vecteur fréquence, paramétré par l'ordre considéré (1 ou 2) et par deux paramètres : Q (variant entre 1 et 16[3]) et k (variant entre 10% et 100%).

Concernant les mesures de distances, nous avons utilisé les distances euclidienne, cityblock, cosinus et spearman. Cependant nous ne parlerons que des trois premières distances, principalement parce que leur comportement plus stable est plus facile à interpréter.

Afin de présenter les résultats de façon claire, nous n'écrivons ici que la partie plus significative des résultats, et nous laissons les résultats complets en annexe B.

6.4.2 Observations pour les bases de sons environnementaux

Considérons tout d'abord les bases de sons environnementaux en laissant de côté Houix2. Les résultats nous montrent qu'à travers les 3 bases (Gygi, Gygi Extended, Houix1), plus le facteur Q est élevé (plus la résolution fréquentielle est grande), meilleures sont les performances quels que soient les autres paramètres (soit k et le type de distance utilisé).

Par exemple la figure 6.1 montre pour la base Gygi Extended l'évolution du MAP en fonction du facteur Q pour $k = 10\%$ et pour chaque type de distance.

Si l'on ne considère que l'ordre 2, à travers les 3 bases et à paramètres égaux par ailleurs, la distance cityblock est celle qui donne les meilleurs résultats. En moyenne, quels que soient les paramètres, cityblock donne des résultats 2.8 points de pourcentage supérieurs aux distances euclidiennes et cosinus. À l'ordre 1 cette prédominance est moins évidente mais toujours visible, la distance cityblock ne donne des résultats en moyenne que 0.6% points supérieurs à ceux des autres distances.

Enfin pour ce qui est du paramètre k, on remarque un comportement particulier pour les bases Gygi et Gygi Extended. À l'ordre 2 et pour les grandes valeurs de Q ($Q = 8$ ou $Q = 16$), les performances baissent lorsque k augmente – c'est-à-dire lorsqu'on considère également les fréquences DCT plus hautes –, alors que l'inverse est constaté à l'ordre 1 et à l'ordre 2 pour des valeurs de Q petites. En particulier pour la base Gygi, la baisse en question est de près de 7 points de pourcentage. Pour la base Houix1 les performances sont croissantes avec k.

3. Pour des raisons de temps de calcul, Q ne varie que jusqu'à 8 pour certaines bases volumineuses

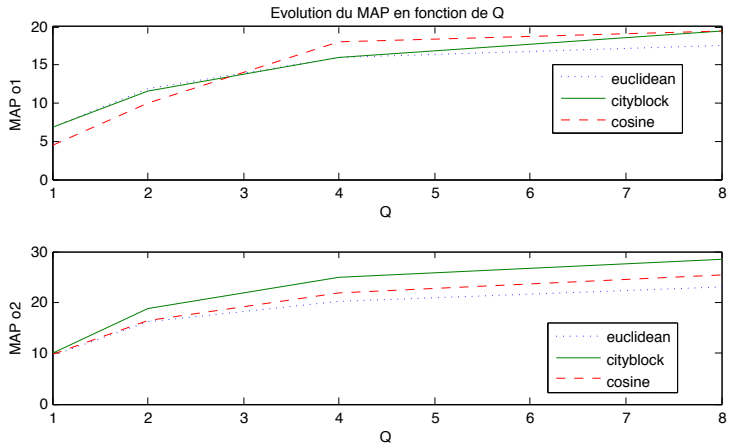


FIGURE 6.1 – Évolution du MAP en fonction de Q pour la base Gygi Extended et k=10% (en haut CLS ordre 1 - en bas CLS ordre 2)

Les résultats pour $Q = 16$ ($Q = 8$ pour Gygi Extended) et $k = 10\%$ sont écrits sur le tableau 6.6.

	BOF/DTW	CLS ordre 1	CLS ordre 2
Gygi	31.8	23.9	39.3
Gygi Extended	27.0	19.4	28.4
Houix1	55.5	54.8	53.4
Houix21	46.3	39.7	40.0
Houix22	27.6	22.5	22.3
Houix23	25.2	15.0	14.7

TABLE 6.6 – Résultats de la représentation CLS sur les bases de sons environnementaux - paramètres : $Q = 8$ pour Gygi Extended et $Q = 16$ pour Gygi et Houix1, $k = 10\%$, distance cityblock

Nous avons, dans notre analyse, laissé la base Houix2 de côté, car ses résultats sont très proches des résultats aléatoires et varient très peu selon les paramètres. Une explication de ce phénomène peut venir du fait que les classes demandées ne se basent pas assez sur des propriétés acoustiques des sons mais se basent davantage sur des propriétés sémantiques, basées sur l'action qui produit le son.

6.4.3 Observations pour les bases de sons instrumentaux

Pour les 3 bases d'instruments (Iowa, SolosDb et RWC), les résultats sont plus éloquants. Comme pour les bases de sons evironnementaux, plus Q est grand, meilleurs sont les résultats. Il en va de même pour k, sauf pour les bases Iowa et RWC qui présentent la même inversion de tendance que les bases Gygi et Gygi Extended pour $Q = 8$. En effet pour $Q = 8$, les performances décroissent avec k croissant (jusqu'à 10 points de pourcentage en moins).

La différence avec le type de bases précédentes vient de la meilleure métrique de distance qui est cosinus, mais la distance cityblock donne des résultats très proches de cosinus.

Nous présentons donc dans le tableau 6.7 les résultats pour les distances cityblock et cosinus, pour $k = 10\%$ et $Q = 8$ ou $Q = 16$.

		cityblock		cosinus	
	BOF/DTW	CLS ordre 1	CLS ordre 2	CLS ordre 1	CLS ordre 2
Iowa	32.0	44.8	47.3	47.0	50.4
RWC	30.2	38.6	44.8	37.1	44.6
SolosDb	39.7	30.5	35.2	32.3	35.8

TABLE 6.7 – Résultats de la représentation CLS sur les bases d'instruments - paramètres : $Q = 8$ pour Iowa et RWC et $Q = 16$ pour solosDb, $k = 10\%$

6.4.4 Autres bases

La base insects donne des résultats similaires aux précédentes. En effet les performances sont croissantes selon Q et k sauf pour $Q = 8$ et $Q = 16$ à l'ordre 2 où l'on retrouve un comportement particulier lorsqu'on utilise la distance cityblock. En effet dans ces conditions les performances atteignent un maximum pour $k = 50\%$.

Les performances de la base digits suivent un comportement similaire en fonction des paramètres Q et k. Les meilleures performances sont atteintes pour $Q = 16$ et $k = 10\%$. Le type de distance utilisée influe peu sur les résultats bien que le maximum soit atteint pour la distance cityblock.

	BOF/DTW	CLS $Ordre1$	CLS $Ordre2$
Insects	42.1	69.3	74.2
Digit	70.1	30.1	50.0

TABLE 6.8 – Résultats de la représentation CLS sur les bases Insects et Digit pour la distance cityblock - paramètres pour Insects : $Q = 8$ et $k = 50\%$ - paramètres pour Digit : $Q = 16$ et $k = 10\%$

6.4.5 Résumé des résultats

Le facteur le plus important pour les résultats est Q (il faut un Q élevé) mais lorsqu'il s'approche de grandes valeurs, des phénomènes inattendus peuvent se produire en relation avec k à l'ordre 2. Pour les comprendre on peut s'interroger sur ce qu'apporte à la fois un facteur Q élevé et un k grand, notamment par rapport à l'ordre 1 où ce phénomène n'apparaît pas.

On peut penser que l'on a besoin d'un facteur Q élevé d'une manière générale mais pour un Q élevé, avec k croissant, on inclue des fréquences DCT de plus en plus élevées à l'ordre 2 (que l'on peut voir comme des modulations très rapides en fréquence) avec une résolution fréquentielle trop grande (facteur Q élevé) donc la représentation devient trop sensible à de l'information inutile.

6.5 Résultats pour scattering Combined

6.5.1 Présentation

Les expérimentations sur la représentation scattering combined ont été menées d'une façon analogue. Les mesures de distance restent les mêmes et seuls certains paramètres de la représentation sont modifiés. Comme nous l'avons vu à la section 5.4 la représentation combined part d'une représentation scattering puis applique de nouveau l'algorithme selon les fréquences. Nous gardons donc le facteur de qualité Q comme paramètre de la représentation scattering de base

	BOF/DTW	CLS ordre 1	CLS ordre 2
Gygi	31.8	23.9	39.3
Gygi Extended	27.0	19.4	28.4
Houix1	55.5	54.8	53.4
Houix21	46.3	39.7	40.0
Houix22	27.6	22.5	22.3
Houix23	25.2	15.0	14.7
Iowa	32.0	47.0	50.4
RWC	30.2	38.6	44.8
SolosDb	39.7	30.5	35.2
Insects	42.1	69.3	74.2
Digit	70.1	30.1	50.0

TABLE 6.9 – Résultats pour la représentation CLS

et pour paramétrer l'application de l'algorithme selon les fréquences, nous utilisons le nombre de fenêtres d'analyse en fréquence (T_{co}). Q varie toujours entre 1 et 16 et T_{co} varie entre 1 et 32.

6.5.2 Observations pour les bases de sons environnementaux

À l'ordre 1, pour les bases de sons environnementaux Gygi, Gygi Extended, une grande délocalisation fréquentielle est requise (soit $T_{co} = 1$ ou 2) mais avec un facteur de qualité important (Q =8 ou 16), alors que pour la base Houix1, les paramètres Q et T_{co} = ont peu d'influence et les résultats restent très stables. Enfin les distances cityblock et cosinus sont celles qui donnent les meilleurs résultats à paramètres égaux.

À l'ordre 2 on remarque que si une grande délocalisation en fréquence est requise (T_{co} = 1 ou 2 également), le facteur Q perd de son importance et a peu d'influence sur les résultats pour Q =2,4 ou 8. Enfin la distance cosinus donne de manière claire les meilleurs résultats. Le fait que T_{co} petit donne les meilleurs résultats indique qu'une délocalisation fréquentielle totale est requise et laisse penser que cette représentation est plus pertinente dans ce cas que la représentation CLS. Ceci est confirmé par les résultats, plus élevés avec combined qu'avec CLS.

	BOF/DTW	combined ordre 1	combined ordre 2
Gygi	31.8	30.0	44.4
Gygi Extended	27.0	20.9	38.9
Houix1	55.5	52.0	59.0
Houix21	46.3	41.7	38.8
Houix22	27.6	23.3	20.6
Houix23	25.2	15.7	12.6

TABLE 6.10 – Résultats de la représentation combined sur les bases de sons environnementaux - paramètres : $Q = 4$, $T_{co} = 1$ et distance cosinus

6.5.3 Observations pour les bases de sons instrumentaux

Pour les bases de sons instrumentaux, on remarque que le facteur Q doit être important quelle que soit la base ou l'ordre considéré. Cependant, par rapport au paramètre T_{co}, les bases SolosDb d'une part et Iowa et RWC d'autre part ont un comportement différent. En effet si

SolosDb obtient les meilleurs résultats pour $T_{co} = 16$, les base Iowa et RWC requièrent un T_{co} petit ($= 1$).

	BOF/DTW	combined ordre 1	combined ordre 2
Iowa	32.0	35.7	39.9
RWC	30.2	39.0	40.4
SolosDb	39.7	29.4	28.0

TABLE 6.11 – Résultats des bases de sons intrumentaux pour la représentation combined - paramètres pour Iowa et RWC : $Q = 8$, $T_{co} = 1$, distance cityblock - paramètres pour SoloDb : $Q = 16$, $T_{co} = 16$, distance cosinus

6.5.4 Autres bases

La base Insects requiert à la fois un facteur Q et un paramètre T_{co} élevé. Ce qui indique que la représentation n'est pas adaptée à ce cas. Les résultats de CLS, plus élevés, le confirment.

Si la distance cosinus semblait être la plus adaptée jusqu'ici, la base Digit fait exception et obtient de meilleurs résultats avec la distance euclidienne. Comme pour la représentation CLS, $Q = 4$ ou 8 conduit aux meilleures performances et de manière particulière pour cette base $T_{co} = 4$ donne les meilleurs résultats de manière instable car si l'on s'en éloigne les performances baissent d'environ 5%.

	BOF/DTW	combined ordre 1	combined ordre 2
Digit	70.1	53.4	60.2
Insects	42.1	62.5	65.9

TABLE 6.12 – Résultats des bases Digit et Insects pour la représentation combined - paramètres pour Digit : $Q = 8$, $T_{co} = 4$ et distance euclidienne - paramètres pour Insects : $Q = 16$, $T_{co} = 16$ et distance cosinus

6.6 Récapitulation des résultats

Le tableau 6.14 récapitule les résultats par base. Pour les algorithmes d'état de l'art, les meilleurs résultats sont écrits et pour les représentations scattering, les résultats pour les paramètres présentés précédemment sont écrits (et ne correspondent donc pas aux meilleurs résultats). Ces paramètres sont rappelés dans le tableau 6.13.

On remarque à travers ce dernier tableau une différence notable entre les bases de sons instrumentaux et les bases de sons environnementaux. En effet la représentation qui donne les meilleures performances est scattering combined pour les sons environnementaux et CLS pour les bases de sons instrumentaux. Ceci peut être rapproché des études de Gaver ([Gav93],[HLM+12]), en particulier lorsqu'il introduit différents types d'écoute et fait la distinction entre « musical listening et « everyday listening ». Dans le premier type d'écoute, l'auditeur fait attention à certaines propriétés acoustiques des sons comme le timbre ou le « pitch » et dans le deuxième type d'écoute, l'auditeur se concentre sur les phénomènes qui ont causé le son. Ces deux types d'écoute ne dépendent pas du son écouté mais du contexte d'écoute. Ainsi dans notre cas les bases instrumentales peuvent nécessiter plus de précision en fréquence et donc les propriétés d'invariance de la représentation combined ne sont pas désirées. À l'inverse, les bases de sons environnementaux doivent être analysées davantage en termes de variations que ce soit en fré-

	CLS			combined		
	Q	k	distance	Q	T_{co}	distance
Gygi	16	10%	city	4	1	cos
Gygi Extended	8	10%	city	4	1	cos
Houix1	16	10%	city.	4	1	cos.
Houix21	16	10%	city.	4	1	cos.
Houix22	16	10%	city.	4	1	cos.
Houix23	16	10%	city.	4	1	cos.
Iowa	8	10%	cos.	8	1	city.
RWC	8	10%	cos.	8	1	city.
SolosDb	16	10%	cos.	16	16	cos.
Digit	16	10%	city.	8	4	eucl.
Insects	8	50%	city.	16	16	cos.

TABLE 6.13 – Rappel des paramètres utilisés pour les représentations scattering selon la base considérée

	ALEA	*BOF*	*DTW*	*CLSo*1	*CLSo*2	*COo*1	*COo*2
gygi	5.1	31.8	25.8	23.9	39.3	30.0	**44.4***
gygiExt	3.6	20.9	19.3	19.4	28.4	20.9	**38.9**
*houix*1	43.6	54.6	55.5	54.8	53.4	52.0	**59.0**
*houix*21	39.4	**46.3**	40.6	39.7	40.0	41.7	38.8
*houix*22	20.6	**27.6**	20.9	22.5	22.3	23.3	20.6
*houix*23	14.7	**25.2**	14.5	15.0	14.7	15.7	13.0
iowa	8.4	29.8	32.0	47.0	**50.4**	35.7	39.9
rwc	8.9	30.0	30.2	38.6	**44.8**	40.5	39.5
solosDb	8.7	**39.7**	33.8	32.3	35.8	29.4	28.0
digit	10.3	45.9	**70.1**	30.1	50.0	53.4	60.2
insects	20.8	0	42.1	69.8	**74.2***	62.5	65.9

TABLE 6.14 – Récapitulatif des résultats par base. Les résultats suivis d'un astérisque indiquent que la distance de Spearman donne de meilleurs résultats

quence ou en temps et vont appeler alors des représentations invariantes ou très stables par translation fréquentielle ou temporelle.

Concernant les bases environnementales, il est intéressant d'étudier les résultats de manière plus fine, notamment en se comparant aux matrices de distances issues d'expériences psycho-acoustiques.

Considérons la base Houix1. Afin de mieux percevoir la qualité des différences entre les algorithmes et la matrice de distances perceptive, la figure 6.2 montre une visualisation de ces matrices sous forme de diagramme de Voronoï (voir Annexe A). Sur les 4 figures, chaque couleur correspond à une classe. On remarque que l'aspect général des figures correspond à la valeur correspondante du MAP. Ainsi la matrice de distances « humain » organise correctement les classes et à l'autre extrême, l'algorithme DTW montre une organisation plus mélangée.

En analysant la figure associée à la représentation combined, on s'aperçoit que certains éléments apparemment mal placés peuvent être expliqués. Sur cette figure, trois éléments en particulier sont notés : « Gouttes » (classe Liquides), « Allumette » (classe Gaz), et « Bip Bip » (classe Machines). Ces trois éléments sont visiblement placés au milieu des sons de solides et en les écoutant on comprend bien que le bruit d'impacts répétés des gouttes ou l'attaque de

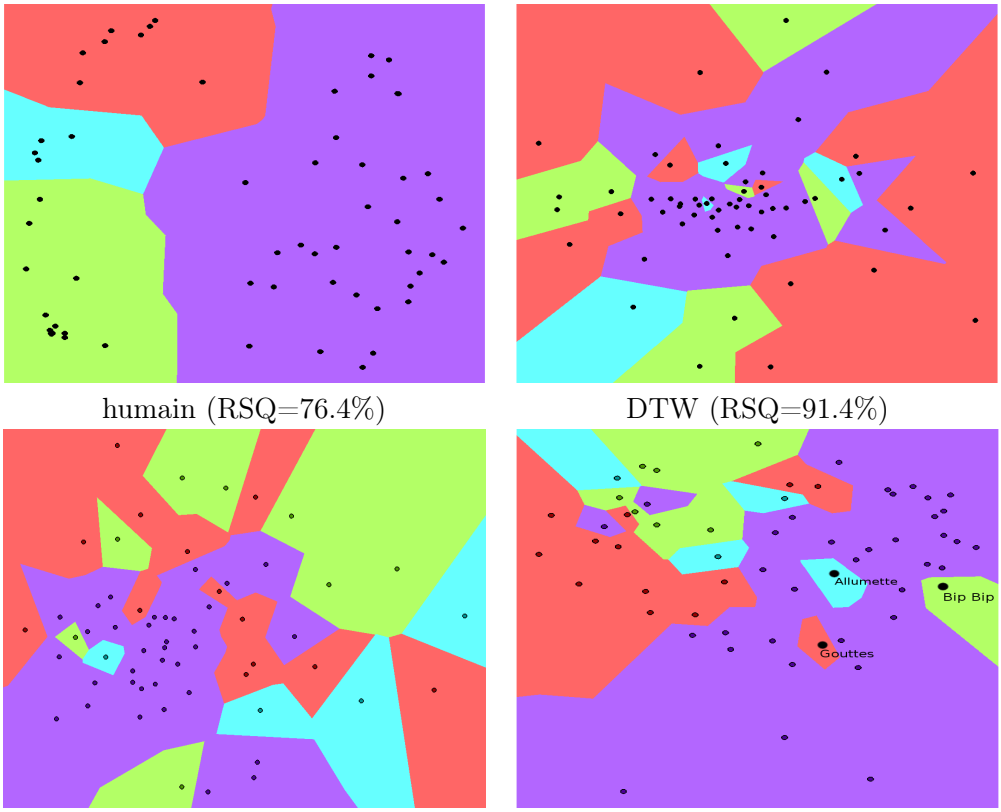


FIGURE 6.2 – Visualisations des matrices de distances issues de l'expérience perceptive sur la base Houix1 (voir section 2.2.2) et des algorithmes DTW, CLS et combined sur la base Houix1. Les couleurs représentent les classes : Liquides (Rouge), Solides (Violet), Gaz (Bleu), Machines (Vert). Les valeurs RSQ représente la précision de la représentation par rapport à la matrice de distances originale (voir Annexe A).

l'allumette en train d'être allumée ou encore les « Bip » répétés d'un micro-onde ont conduit l'algorithme à les rapprocher de sons de Solides.

Ceci illustre l'importance d'une analyse plus fine, lorsque cela est possible, des résultats afin de mieux comprendre le comportement des algorithmes utilisés et de mitiger les résultats.

Conclusion

À travers cette étude sur les représentations et distances entre sons qui posent le cadre formel de la modélisation de la perception auditive, nous avons pu percevoir l'arbitrage de complexité entre algorithme de représentation et métrique de distance. Plus l'on s'approche d'une représentation idéale, c'est-à-dire épurée de toute information inutile et structurée de telle sorte à décorréler l'information utile, plus la métrique de distance peut être simple : l'effort est concentré dans le calcul de la représentation. C'est ce que nous avons vu avec les représentations scattering où des distances simples comme la distance cosinus donnent de bons résultats. À l'inverse, des représentations plus simples comme le spectrogramme ou le Mel-spectre nécessitent la mise en place de métriques de distances plus complexes, comme l'algorithme DTW pour tenir compte des mêmes informations.

En comparaison aux autres types de représentation, les algorithmes scattering semblent apporter une représentation pertinente des sons au travers de trois principales contributions.

Premièrement, la représentation sous forme de fréquences des modulations fréquentielles. En effet les autres représentations classiques font apparaître les modulations fréquentielles de manière indirecte et nécessitent une analyse du signal dans le temps (comme DTW) pour les extraire.

Deuxièmement, les propriétés d'invariance et de stabilité en temps et en fréquence. La représentation scattering originale peut garantir une invariance par translation dans le temps sans perdre en résolution fréquentielle, ainsi qu'une stabilité par translation en fréquence. La représentation scattering combined propose une forme d'invariance par translation fréquentielle.

Enfin la représentation CLS permet également de décorréler les coefficients scattering et concentrer ainsi l'information sur moins de coefficients.

Ce que semblent nous montrer nos résultats est une pertinence des représentations scattering dans l'étude de la similarité entre sons. Les résultats montrent une amélioration des performances par rapport à l'état de l'art, ce qui conforte l'idée que les apports des représentations scattering sont utiles dans notre contexte.

Cette étude ouvre également à d'autres questions concernant aussi bien la représentation scattering en soi que les mesures de distance. Il s'agit principalement de s'intéresser aux différences entre l'ordre 1 et l'ordre 2. En effet notre manière de percevoir le timbre d'un son n'est pas la même que notre perception des évolutions temporelles de ce timbre. Il serait par exemple intéressant de changer et adapter le banc de filtres en ondelettes entre l'ordre 1 et l'ordre 2 de scattering ou d'adapter le paramètre T_{co} entre l'application du deuxième algorithme scattering sur les coefficients d'ordre 1 et son application sur ceux d'ordre 2. Enfin les distances que nous avons utilisées tiennent compte arbitrairement des ordres 1 et 2 de manière équivalente.

Bibliographie

[ADP07] J.-J Aucouturier, B. Defreville, and F. Pachet. The bag-of-frame approach to audio pattern recognition : A sufficient model for urban soundscapes but not for polyphonic music. *Journal of the Acoustical Society of America*, 122(2) :881–891, 2007.

[AM11] J. Andén and S. Mallat. Multiscale scattering for audio classification. In *ISMIR*, pages 657–662, 2011.

[AV07] H Abdi and D. Valentin. DISTATIS How to analyze multiple distance matrices. 2007.

[CKB12] Y. Chen, E. Keogh, and G. Batista. Ucr insect classification contest. Disponible sur internet à l'adresse http://www.cs.ucr.edu/~eamonn/CE/contest.htm, 2012.

[DKJ+07] Jason V. Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S. Dhillon. Information-theoretic metric learning. In *ICML*, pages 209–216, 2007.

[ERD06a] S. Essid, G. Richard, and B. David. Instrument recognition in polyphonic music based on automatic taxonomies. In *IEEE Transactions on Speech and Audio Processing*, pages 68–80, 2006.

[ERD06b] S. Essid, G. Richard, and B. David. Musical instrument recognition by pairwise classification strategies. In *IEEE Transactions on Audio Speech and Language Processing*, pages 1401–14012, 2006.

[Fri97] L. Fritts. Musical instrument samples. Disponible sur internet à l'adresse http://theremin.music.uiowa.edu/MIS.html, 1997.

[Gav93] W. W. Gaver. What in the World Do We Hear ? : An Ecological Approach to Auditory Event Perception. *Ecological Psychology*, 5(1) :1–29, March 1993.

[GKW07] B. Gygi, Gary R Kidd, and Charle S Watson. Similarity and categorization of environmental sounds. *Perception And Psychophysics*, 69(6) :839–855, 2007.

[GN03] M. Goto and T. Nishimura. Rwc music database : Music genre database and musical instrument sound database. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pages 229–230, 2003.

[HLM+12] O. Houix, G. Lemaitre, N. Misdariis, P. Susini, and I. Urdapilleta. A lexical analysis of environmental sound categories. *Journal of Experimental Psychology : Applied*, 2012.

[LD91] R. Gary Leonard and George R. Doddington. A speaker-independent connected-digit database. Disponible sur Internet à l'adresse http://www.ldc.upenn.edu/Catalog/docs/LDC93S10/tidigits.readme.html, February 1991.

[Log00] B. Logan. Mel frequency cepstral coefficients for music modeling. In *In International Symposium on Music Information Retrieval*, 2000.

[MSS06] N. Mesgarani, M. Slaney, and S. A. Shamma. Discrimination of speech from nonspeech based on multiscale spectro-temporal modulations. In *IEEE Transactions on Audio, Speech and Language Processing*, pages 920–930, 2006.

[PKA10] Y Panagakis, C Kotropoulos, and G R Arce. Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification, 2010.

[Rey09] Douglas A. Reynolds. Gaussian mixture models. In *Encyclopedia of Biometrics*, pages 659–663. 2009.

[SML06a] P Susini, N Misdariis, and G Lemaitre. Closing the loop of sound evaluation and design. *Perceptual Quality of Systems*, 2006.

[SML06b] P Susini, Nicolas Misdariis, and Guillaume Lemaitre. Closing the loop of sound evaluation and design. *Perceptual Quality of Systems*, 2(4), 2006.

[TE03] Robert J. Turetsky and Daniel P. W. Ellis. Ground-truth transcriptions of real music from force-aligned midi syntheses. In *ISMIR*, 2003.

[ToCS02] G. Tzanetakis and Princeton University. Dept. of Computer Science. *Manipulation, Analysis and Retrieval Systems for Audio Signals*. Princeton University, 2002.

[WS95] K. Wang and S. A. Shamma. Spectral Shape analysis in the Central Auditory System. *sap*, 3, 1995.

Annexe A

Outil de visualisation

Pendant le stage nous avons développé un outil de visualisation de matrices de distances à l'aide d'une analyse MDS et d'un diagramme de Voronoï. En effet les mesures de performances ne donnent qu'une vision partielle des résultats et il était intéressant de pouvoir visualiser plus finement les matrices de distances afin de les évaluer qualitativement, en particulier en repérant les sons mal classés ou en jugeant visuellement de l'organisation spatiale des classes entre elles.

Cet outil se présente sous la forme d'une application graphique matlab se présentant comme sur la figure A.1. En chargeant un fichier contenant des matrices de distance et certaines informations complémentaires comme le nom des points ou la classe d'appartenance de chaque point, l'outil tente de recréer un nuage de points dans un espace à 2 ou 3 dimensions où les distances entre les points correspondent à celles de la matrice de distance.

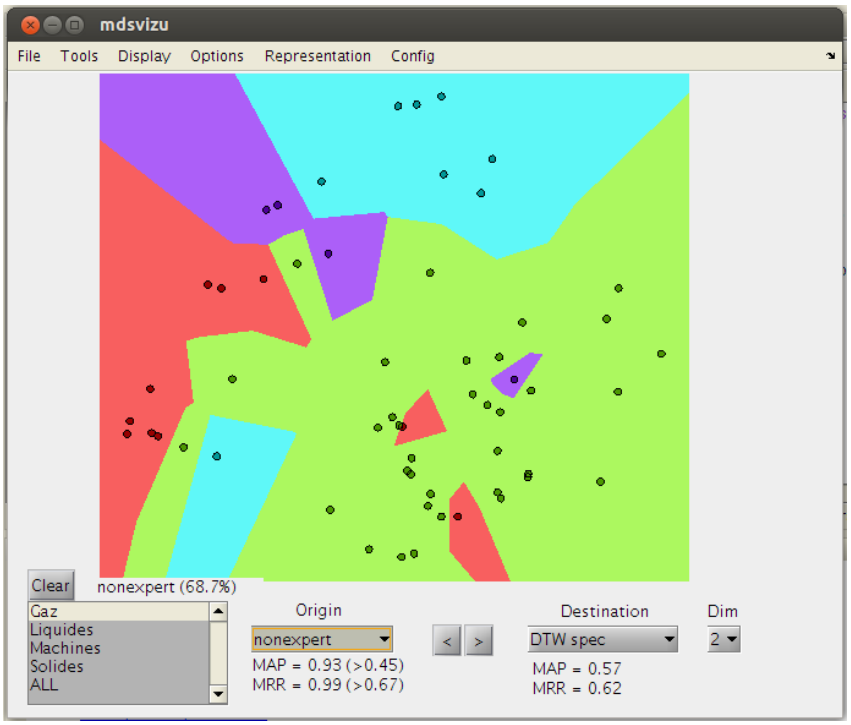


FIGURE A.1 – Interface de l'outil de visualisation de matrices de distances.

A.1 Calcul de la représentation

Pour représenter visuellement une matrice de distances, nous avons procédé en deux étapes. Nous projetons tout d'abord la matrice de distances dans un espace de dimension fixée (2 ou 3) à l'aide d'une Multi-Dimensional Scaling (MDS), puis nous représentons le nuage de points résultant sous forme d'un diagramme de Voronoï.

L'analyse MDS est une technique de visualisation de données de similarité ou dissimilarité. À partir d'une matrice de distances entre N points, l'algorithme MDS renvoie les coordonnées de N points dans un espace de dimension fixée. Dans notre cas, pour que l'on puisse représenter ces points graphiquement, la dimension de l'espace d'arrivée est 2 ou 3.

Afin de visualiser les classes de chaque point au même temps que les distances, nous assignons une couleur à chaque classe et représentons chaque point avec la couleur de la classe à laquelle il appartient comme montré sur la figure A.2.

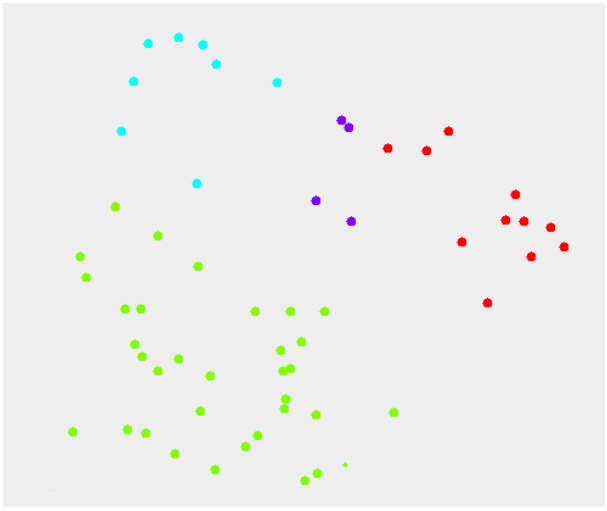


FIGURE A.2 – Exemple de visualisation MDS d'une matrice de distances. Chaque couleur correspond à une classe.

L'analyse MDS ne rend compte des distances de la matrice d'origine qu'avec une certaine précision. Il est donc important de pouvoir mesurer la fidélité de la représentation MDS à la matrice de distances. Pour cela nous calculons et affichons sur l'outil de visualisation la valeur R-Squared Correlation (RSQ) qui mesure la proportion de variance prise en compte dans la représentation MDS. Si l'on note $d = (d_i)_{i\in[1,M]}$ le vecteur des $M = N(N-1)/2$ distances de la matrice de distances et $\hat{d} = (\hat{d}_i)_{i\in[1,M]}$ le vecteur des distances entre les points calculés par MDS, alors RSQ s'écrit :

$$\text{RSQ} = 1 - \frac{\sum_{i=1}^{M}(d_i - \hat{d}_i)^2}{\sum_{i=1}^{M}(d_i - \bar{d})^2}$$

En dimension 2 nous avons ajouté la possibilité de représenter le nuage de point sous forme d'un diagramme de Voronoï. Étant donné un ensemble de N points $\mathcal{P} = (p_i)_{i\in[1,N]}$, un diagramme de Voronoï est un pavage du plan en N zones (ou cellules) $(c_i)_{i\in[1,N]}$ où c_i est l'ensemble des points plus proches de p_i que de tout autre point de $\mathcal{P}$, soit :

$$c_i = \{x \in \mathbb{R}^2, d(x, p_i) = min_{k\in[1,N]} d(x, p_k)\}$$

où d est la distance euclidienne. Chaque cellule c_i est alors représentée avec la couleur de la classe à laquelle appartient le point p_i. On obtient ainsi une visualisation telle que celle de la figure A.1.

A.2 Fonctionnalités

L'outil de visualisation propose certaines fonctionnalités facilitant l'analyse de cette représentation. Les plus importantes sont résumées ci-dessous :

- Afficher la représentation en 2 ou 3 dimensions du nuage de points (MDS simple ou MDS+Voronoï).
- Outils de rotation, zoom, navigation dans la représentation.
- Afficher les noms des points ou leur classe.
- Lire un fichier son associé à un point.
- Afficher la valeur RSQ de la représentation MDS.
- Afficher le MAP et Mean Reciprocal Rank (MRR) de la matrice de distances affichée.
- Ajouter un filtre sur les classes affichées.
- Tracer des liens entre chaque point et son plus proche voisin (selon la matrice de distances d'origine).
- Afficher une animation interpolant linéairement deux nuages de points.
- Exporter la visualisation sous forme d'image.

Enfin, nous une préversion de l'outil de visualisation est disponible en ligne à l'adresse : http://recherche.ircam.fr/equipes/analyse-synthese/lagrange/research/vizuMds/.

Annexe B

Détails des résultats pour la représentation CLS

Les résultats sont présentés par base. La métrique d'évaluation est le MAP (voir chapitre 3) et les performances sont écrites en pourcentage. La valeur maximale pour chaque ordre de CLS est notée en gras. Pour les résultats impliquant la distance de spearman, les cases notées « 0 » indiquent que la matrice de distances contenaient un nombre trop important de

B.1 Gygi CLS Ordre 1

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	11.2	14.0	21.5	23.3	24.1	9.6	14.2	21.1	**26.8**	23.9
	0.25	11.3	20.8	24.7	24.0	24.1	11.3	20.4	25.9	24.7	24.8
	0.5	14.2	22.6	24.6	24.0	24.5	16.7	23.0	25.4	25.6	24.5
	0.75	17.2	22.6	25.2	24.0	24.5	20.9	23.9	25.9	25.3	25.3
	1	17.2	22.7	24.8	24.0	24.5	21.3	23.7	26.1	25.3	25.3
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	6.1	9.9	19.5	21.6	23.3	5.6	5.1	0.0	12.3	19.6
	0.25	8.6	20.8	23.4	23.5	23.5	4.9	0.0	17.4	22.9	18.7
	0.5	14.1	21.2	23.7	23.0	23.3	0.0	16.1	22.0	20.5	16.8
	0.75	15.4	21.4	24.2	23.4	23.3	16.0	18.8	22.6	18.0	15.4
	1	16.2	21.6	24.1	23.4	23.3	19.3	22.3	20.6	14.3	13.3

B.2 Gygi CLS Ordre 2

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	15.4	18.6	29.7	31.9	31.7	15.0	24.1	34.7	37.2	**39.3**
	0.25	14.3	27.9	31.0	33.1	31.5	16.0	33.5	37.7	36.9	35.0
	0.5	19.6	27.2	30.7	32.1	31.4	25.7	34.2	35.4	34.5	34.1
	0.75	20.5	27.8	30.6	32.2	31.6	26.6	35.3	35.5	33.8	33.1
	1	21.5	27.8	30.6	32.3	31.3	26.9	35.4	35.4	33.2	32.5
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	10.9	15.4	28.4	33.1	34.3	0.0	0.0	38.9	37.7	32.1
	0.25	12.9	28.6	32.7	34.3	33.8	15.8	37.5	38.2	28.5	21.9
	0.5	19.3	29.5	32.2	34.2	33.8	27.9	34.5	33.0	21.9	22.0
	0.75	21.6	30.0	32.1	34.2	34.3	28.0	32.5	27.0	25.1	24.4
	1	22.7	29.1	32.1	34.2	34.5	25.3	32.0	32.4	24.2	23.1

B.3 Gygi Extended CLS Ordre 1

d		euclidean				cityblock			
Q		1	2	4	8	1	2	4	8
k	0.1	6.9	11.9	15.9	17.6	6.8	11.6	16.0	19.4
	0.25	10.1	16.6	17.5	17.8	10.1	17.4	19.7	**20.3**
	0.5	12.4	17.0	17.7	17.8	13.1	19.1	20.2	20.0
	0.75	12.6	17.2	17.7	17.8	13.9	19.6	19.9	19.7
	1	12.6	17.2	17.8	17.8	14.2	19.5	19.7	19.7
d		cosine				spearman			
Q		1	2	4	8	1	2	4	8
k	0.1	4.5	10.1	18.0	19.4	7.2	6.8	9.4	15.8
	0.25	9.8	19.5	20.3	19.8	10.3	11.7	16.7	19.9
	0.5	14.2	20.1	20.2	19.8	11.5	16.6	19.3	**20.4**
	0.75	14.9	20.2	20.2	19.8	14.2	18.1	19.4	19.6
	1	14.9	20.2	20.2	19.7	15.1	18.1	18.4	17.5

B.4 Gygi Extended CLS Ordre 2

d		euclidean				cityblock			
Q		1	2	4	8	1	2	4	8
k	0.1	9.5	16.1	20.1	23.0	10.1	18.9	24.8	**28.4**
	0.25	12.0	21.3	22.4	23.5	13.1	26.0	28.4	28.4
	0.5	14.0	21.9	22.7	23.5	16.1	27.5	28.2	28.0
	0.75	14.2	21.9	22.7	23.4	16.8	27.4	27.7	27.2
	1	14.3	22.0	22.7	23.4	17.0	27.3	27.6	26.9
d		cosine				spearman			
Q		1	2	4	8	1	2	4	8
k	0.1	9.8	16.5	21.8	25.4	6.7	22.1	27.9	30.3
	0.25	12.8	23.2	25.0	26.3	14.7	30.1	31.7	26.5
	0.5	15.5	23.6	25.4	26.3	19.8	**32.3**	28.8	23.4
	0.75	15.9	23.6	25.3	26.3	20.2	31.1	26.6	21.7
	1	16.1	23.6	25.3	26.2	19.9	29.6	24.8	20.7

B.5 Houix1 CLS Ordre 1

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	53.0	56.5	55.1	54.8	54.6	53.3	56.6	54.5	55.0	54.8
	0.25	53.7	56.1	55.6	55.3	54.9	53.5	55.4	56.2	56.0	55.2
	0.5	53.9	56.3	55.8	55.4	55.0	53.1	56.1	**56.7**	55.9	56.2
	0.75	53.9	56.3	55.8	55.4	55.0	53.1	56.2	56.3	56.0	56.1
	1	53.9	56.3	55.8	55.4	55.0	53.4	56.3	56.2	56.3	56.0
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	49.3	55.2	54.6	55.0	55.0	39.1	40.0	42.7	45.0	44.5
	0.25	51.8	56.0	55.6	55.5	55.3	44.0	42.1	43.9	46.1	45.9
	0.5	53.3	56.2	56.0	55.7	55.4	42.0	43.5	44.6	44.8	45.4
	0.75	53.4	**56.2**	56.0	55.7	55.4	42.9	42.9	43.4	44.3	46.3
	1	53.4	56.2	56.0	55.7	55.4	42.9	42.8	43.7	44.8	47.0

B.6 Houix1 CLS Ordre 2

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	54.6	53.6	51.6	52.0	51.6	54.8	53.8	52.3	53.2	53.4
	0.25	51.9	53.1	52.6	52.5	52.0	52.3	54.0	54.5	54.9	53.6
	0.5	52.5	53.6	52.9	52.4	52.1	53.3	55.3	55.4	55.0	53.3
	0.75	52.6	53.6	53.0	52.7	52.1	54.7	**55.6**	55.6	54.8	52.9
	1	52.7	53.6	53.0	52.7	52.1	54.6	55.6	55.6	54.7	52.7
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	53.4	51.9	51.5	52.6	52.5	52.0	52.4	51.6	51.0	49.9
	0.25	51.1	53.0	52.8	53.1	52.7	**55.3**	51.4	50.2	48.2	49.1
	0.5	52.4	53.4	53.0	53.0	52.8	52.1	51.8	48.6	48.1	49.1
	0.75	52.6	53.4	53.0	53.0	52.8	51.5	51.7	47.7	47.9	49.0
	1	52.7	53.4	53.0	53.1	52.8	49.5	50.9	47.2	48.0	49.2

B.7 Iowa CLS Ordre 1

d		euclidean				cityblock			
Q		1	2	4	8	1	2	4	8
k	0.1	20.7	23.8	36.4	43.0	20.5	23.1	35.5	**44.8**
	0.25	29.5	36.8	39.8	39.3	28.5	36.2	38.3	35.5
	0.5	36.2	36.6	38.6	36.6	35.9	34.7	33.8	29.4
	0.75	36.8	36.2	38.6	36.6	36.1	33.0	33.4	29.2
	1	36.2	36.1	38.6	36.6	34.4	32.2	33.3	29.2
d		cosine				spearman			
Q		1	2	4	8	1	2	4	8
k	0.1	18.0	21.6	38.0	**47.0**	17.0	17.0	19.0	33.1
	0.25	27.7	39.1	41.5	40.4	16.9	21.2	25.5	21.5
	0.5	37.1	38.6	39.3	36.9	22.3	22.5	20.5	17.5
	0.75	37.7	37.9	39.2	36.8	24.8	19.6	18.7	16.5
	1	36.8	37.7	39.2	36.8	23.6	17.8	17.7	16.3

B.8 Iowa CLS Ordre 2

d		euclidean				cityblock			
Q		1	2	4	8	1	2	4	8
k	0.1	25.5	28.0	38.1	43.8	27.2	29.8	39.7	**47.3**
	0.25	33.0	38.7	42.8	42.8	34.8	39.7	42.2	40.8
	0.5	40.0	39.3	42.3	40.7	41.4	38.3	37.6	33.2
	0.75	40.5	39.0	42.1	40.6	40.8	36.4	36.4	32.4
	1	40.0	39.0	42.2	40.6	38.6	35.6	36.1	32.3
d		cosine				spearman			
Q		1	2	4	8	1	2	4	8
k	0.1	28.0	30.1	42.4	**50.4**	18.0	22.6	30.9	39.6
	0.25	33.5	42.3	46.8	47.5	26.3	31.3	30.0	22.7
	0.5	43.2	42.3	45.5	44.0	32.2	26.6	21.9	17.8
	0.75	43.3	41.8	45.2	43.9	30.0	23.1	20.2	17.3
	1	42.2	41.7	45.2	43.8	27.4	21.4	19.5	17.3

B.9 RWC CLS Ordre 1

d		euclidean				cityblock			
Q		1	2	4	8	1	2	4	8
k	0.1	17.6	21.9	31.2	36.6	17.4	21.1	30.7	**38.6**
	0.25	24.7	28.3	34.3	34.7	24.2	27.6	34.0	33.2
	0.5	29.2	29.1	33.8	33.0	28.9	28.2	31.5	29.3
	0.75	30.0	29.0	33.8	33.0	29.6	27.6	31.3	29.2
	1	29.9	29.0	33.8	33.0	29.1	27.3	31.0	29.1
d		cosine				spearman			
Q		1	2	4	8	1	2	4	8
k	0.1	16.3	21.3	30.5	**37.1**	18.0	18.7	23.2	34.0
	0.25	23.7	29.2	33.9	33.4	18.1	19.6	26.2	23.4
	0.5	28.9	29.9	32.8	31.0	20.2	20.8	21.5	19.1
	0.75	29.7	29.7	32.7	31.0	21.4	19.5	19.4	17.9
	1	29.5	29.6	32.6	31.0	21.2	18.2	18.0	17.1

B.10 RWC CLS Ordre 2

d		euclidean				cityblock			
Q		1	2	4	8	1	2	4	8
k	0.1	22.3	26.8	36.3	41.6	23.3	28.1	36.8	**44.8**
	0.25	28.8	32.8	40.3	41.1	29.6	33.0	40.6	40.3
	0.5	34.3	33.9	40.1	39.9	34.8	33.6	37.6	35.1
	0.75	35.2	33.8	40.1	39.9	35.4	32.6	37.0	34.5
	1	35.1	33.8	40.1	39.9	34.5	32.2	36.7	34.3
d		cosine				spearman			
Q		1	2	4	8	1	2	4	8
k	0.1	22.8	27.0	35.8	**44.6**	18.2	24.0	29.3	41.1
	0.25	27.9	33.5	41.0	42.5	22.6	27.8	32.2	25.7
	0.5	34.2	34.6	40.0	40.2	26.6	25.9	24.5	20.1
	0.75	35.2	34.4	39.9	40.1	27.6	23.1	22.5	19.3
	1	34.9	34.3	39.9	40.1	26.1	21.6	21.4	19.0

B.11 SolosDb CLS Ordre 1

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	14.9	18.8	25.6	28.6	29.5	14.8	18.8	25.8	29.8	**30.5**
	0.25	19.6	24.7	28.7	29.4	29.5	19.8	25.0	30.0	30.0	29.3
	0.5	23.0	26.6	29.0	29.5	29.5	23.9	27.7	29.9	29.6	28.4
	0.75	23.4	26.7	29.1	29.5	29.5	24.4	27.7	29.9	29.6	28.4
	1	23.5	26.7	29.1	29.5	29.5	24.4	27.7	30.0	29.5	28.2
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	11.9	18.9	27.2	31.4	**32.3**	19.9	22.9	21.0	25.3	25.9
	0.25	19.8	26.3	31.1	32.0	32.2	20.3	21.9	25.1	25.5	22.5
	0.5	24.2	28.6	31.6	32.0	32.1	20.5	22.9	24.4	23.0	20.6
	0.75	24.6	28.7	31.6	32.1	32.1	21.5	22.4	23.1	21.7	19.9
	1	24.6	28.7	31.6	32.1	32.1	21.6	21.2	21.8	20.6	19.2

B.12 SolosDb CLS Ordre 2

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	15.2	21.9	28.0	30.9	32.5	15.6	23.0	30.2	33.9	**35.2**
	0.25	19.7	27.3	31.6	32.7	33.2	20.5	29.5	34.3	35.1	34.7
	0.5	23.6	29.7	32.3	33.0	33.3	25.0	32.4	34.8	34.6	33.5
	0.75	24.0	29.8	32.3	33.1	33.4	25.4	32.2	34.7	34.5	33.1
	1	24.0	29.8	32.3	33.1	33.4	25.3	32.1	34.7	34.3	33.1
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	15.6	23.7	30.2	33.9	35.8	16.3	20.9	25.8	32.3	31.0
	0.25	20.7	29.6	34.3	35.7	36.4	17.5	26.8	31.2	30.5	25.8
	0.5	24.9	32.0	35.0	36.0	36.5	23.4	29.2	29.2	26.3	22.9
	0.75	25.3	32.1	35.1	36.1	36.6	23.6	27.8	27.2	25.2	22.3
	1	25.2	32.2	35.1	36.1	**36.6**	23.5	26.7	26.1	24.1	22.1

B.13 Digit CLS Ordre 1

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	17.6	15.4	18.3	26.3	28.7	17.1	15.2	17.7	26.8	30.1
	0.25	17.7	19.8	29.0	28.4	26.3	17.0	19.5	**30.7**	29.5	25.8
	0.5	23.0	23.7	28.9	26.9	25.7	23.2	24.7	29.9	26.5	24.0
	0.75	24.2	23.6	28.7	26.8	25.6	25.0	24.6	28.9	25.9	23.8
	1	24.5	23.4	28.7	26.8	25.7	25.5	24.2	28.6	25.7	23.7
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	15.3	13.7	18.1	29.2	32.2	10.1	10.1	10.7	27.7	**33.0**
	0.25	16.2	20.9	32.9	31.8	28.9	10.1	13.4	30.9	29.4	22.8
	0.5	26.0	26.6	32.6	30.0	28.3	16.1	24.3	29.5	22.4	17.3
	0.75	27.5	26.4	32.3	29.8	28.2	22.3	23.8	25.9	19.4	15.6
	1	27.8	26.3	32.2	29.8	28.2	24.8	21.4	22.9	17.8	14.9

B.14 Digit CLS Ordre 2

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	25.1	22.1	20.7	36.7	47.8	25.4	22.8	21.9	39.4	**50.0**
	0.25	24.3	26.2	35.9	42.4	47.4	24.8	27.9	39.6	44.9	45.7
	0.5	26.4	30.6	37.7	42.5	46.8	27.3	34.2	41.6	42.6	42.1
	0.75	27.7	30.8	38.0	42.4	46.7	28.9	34.0	41.5	41.5	40.9
	1	28.1	30.6	38.0	42.4	46.8	29.6	33.2	40.9	41.0	40.5
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	24.8	21.3	19.7	37.0	48.6	18.2	11.7	21.2	43.2	**48.6**
	0.25	24.0	26.1	36.9	43.5	47.5	20.9	26.5	42.4	42.1	35.5
	0.5	27.2	31.2	39.2	43.6	46.8	28.0	33.1	39.9	33.6	26.0
	0.75	28.9	31.5	39.5	43.4	46.7	30.0	30.4	37.1	29.0	22.9
	1	29.3	31.3	39.4	43.4	46.7	29.2	26.9	33.6	26.9	21.6

B.15 Insects CLS Ordre 1

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	24.8	25.0	31.1	37.1	49.9	24.8	25.0	31.1	38.0	55.6
	0.25	25.3	33.7	39.4	54.1	61.5	25.2	34.2	42.0	62.8	**69.8**
	0.5	28.3	36.9	49.1	59.1	61.4	28.8	39.8	59.8	69.3	67.0
	0.75	29.3	39.1	50.3	59.2	61.4	30.9	46.2	63.0	69.0	66.3
	1	30.1	40.1	50.4	59.2	61.4	33.2	49.9	63.4	68.7	66.1
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	22.7	23.2	31.9	39.5	54.2	20.8	20.8	25.9	39.4	63.5
	0.25	23.1	34.7	42.1	59.0	66.4	20.8	31.0	43.6	70.7	78.2
	0.5	29.1	38.8	53.6	64.1	66.1	31.0	42.7	67.8	**78.5**	73.7
	0.75	30.7	41.7	55.1	64.2	66.0	33.2	54.6	74.0	77.6	69.8
	1	32.0	43.0	55.2	64.2	66.0	38.7	61.5	75.3	75.8	67.3

B.16 Insects CLS Ordre 2

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	24.6	27.4	35.5	42.9	55.3	24.7	28.8	37.0	45.9	60.7
	0.25	25.3	37.7	45.7	62.2	66.4	25.2	40.1	51.1	70.4	72.1
	0.5	29.5	41.5	56.9	66.5	65.4	30.0	46.7	68.6	**74.2**	67.1
	0.75	30.9	44.8	58.3	66.6	65.3	32.7	55.0	71.3	73.2	65.9
	1	32.0	46.2	58.5	66.6	65.2	35.5	58.9	71.4	72.6	65.5
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
k	0.1	24.0	28.6	37.5	45.9	61.7	23.6	28.1	33.2	44.5	67.7
	0.25	24.5	39.2	48.5	67.5	72.3	24.8	37.1	54.0	76.1	76.7
	0.5	30.4	43.6	61.2	71.4	70.7	34.6	48.5	75.7	**77.9**	67.5
	0.75	32.2	47.5	62.7	71.4	70.6	37.0	62.9	77.8	75.7	64.1
	1	33.8	49.1	63.0	71.4	70.6	41.9	67.9	77.2	74.2	62.7

Annexe C

Détails des résultats pour la représentation scattering combined

C.1 Gygi CO Ordre 1

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	18.7	20.0	20.8	21.1	20.2	18.1	18.5	19.9	22.4	24.5
	2	20.3	21.9	23.5	24.0	24.6	22.1	23.8	22.3	28.4	**32.0**
	4	20.2	21.2	24.3	24.6	23.9	20.4	23.6	25.7	28.7	30.5
	8	20.2	20.2	23.4	23.7	24.2	20.4	23.6	26.8	28.6	29.4
	16	20.2	20.2	23.3	20.6	22.1	20.4	23.6	28.0	25.0	25.8
	32	20.2	20.2	23.3	20.6	21.1	20.4	23.6	28.0	25.0	22.5
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	29.0	29.4	30.0	28.3	30.8	22.9	22.1	19.4	27.8	30.0
	2	27.7	28.9	28.2	27.4	27.8	25.9	23.6	24.4	26.4	28.8
	4	26.7	26.4	27.3	24.2	26.8	16.7	20.3	22.5	26.4	27.5
	8	26.7	25.1	26.2	23.4	27.0	16.7	16.3	23.6	27.3	30.6
	16	26.7	25.1	24.9	22.9	22.9	16.7	16.3	24.8	25.3	26.1
	32	26.7	25.1	24.9	22.9	21.8	16.7	16.3	24.8	25.3	27.2

C.2 Gygi CO Ordre 2

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	33.1	34.1	35.4	34.9	33.7	33.2	34.4	35.8	34.8	35.8
	2	31.5	33.8	36.5	34.0	33.9	33.2	35.1	34.7	34.4	34.9
	4	31.0	32.7	33.3	33.8	33.7	33.5	34.5	35.8	33.7	34.1
	8	31.0	31.6	33.2	31.8	30.7	33.5	33.2	35.6	34.3	33.5
	16	31.0	31.6	32.7	30.4	28.9	33.5	33.2	34.2	33.9	34.0
	32	31.0	31.6	32.7	30.4	25.2	33.5	33.2	34.2	33.9	32.7
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	42.2	44.7	44.4	43.2	43.7	**45.9**	43.9	44.2	41.6	40.2
	2	41.4	43.7	43.1	40.9	41.5	39.6	39.8	37.1	37.0	35.7
	4	39.8	39.7	40.2	39.3	39.7	32.1	36.0	37.7	34.6	34.6
	8	39.8	38.5	38.7	39.0	37.5	32.1	33.3	36.7	34.3	35.3
	16	39.8	38.5	39.8	36.5	35.1	32.1	33.3	36.4	37.1	37.8
	32	39.8	38.5	39.8	36.5	34.5	32.1	33.3	36.4	37.1	38.0

C.3 Gygi Extended CO Ordre 1

d		euclidean				cityblock			
Q		1	2	4	8	1	2	4	8
Tco	1	18.7	18.5	20.7	22.6	18.9	19.0	20.9	23.4
	2	20.9	20.4	22.4	24.1	22.2	22.2	22.3	**25.8**
	4	22.3	22.1	23.3	23.9	23.5	23.3	24.4	25.7
	8	22.3	22.1	23.2	22.8	23.5	23.6	25.1	25.3
	16	22.3	22.1	22.4	21.5	23.5	23.6	24.6	24.0
	32	22.3	22.1	22.4	21.5	23.5	23.6	24.6	24.0
d		cosine				spearman			
Q		1	2	4	8	1	2	4	8
Tco	1	18.5	20.0	20.9	21.8	18.7	0.0	0.0	0.0
	2	18.8	19.6	21.1	**21.9**	0.0	0.0	0.0	0.0
	4	19.0	19.2	20.9	21.0	15.4	16.3	0.0	18.9
	8	19.0	19.0	20.1	20.7	15.4	14.5	17.6	19.9
	16	19.0	19.0	20.0	19.5	15.4	14.5	17.6	20.0
	32	19.0	19.0	20.0	19.5	15.4	14.5	17.6	20.0

C.4 Gygi Extended CO Ordre 2

d		euclidean				cityblock			
Q		1	2	4	8	1	2	4	8
Tco	1	30.6	31.6	32.2	32.4	33.0	34.7	**35.6**	34.6
	2	31.1	31.6	32.6	32.1	33.4	34.7	35.1	34.1
	4	30.4	30.8	31.1	30.5	32.9	33.9	33.9	32.7
	8	30.4	30.3	30.2	28.3	32.9	33.5	32.8	31.4
	16	30.4	30.3	29.2	26.6	32.9	33.5	32.2	30.3
	32	30.4	30.3	29.2	26.6	32.9	33.5	32.2	30.3
d		cosine				spearman			
Q		1	2	4	8	1	2	4	8
Tco	1	39.1	**39.8**	38.9	38.6	35.6	33.8	31.5	30.0
	2	37.7	37.7	37.4	36.6	30.6	30.2	28.8	26.7
	4	35.1	35.6	35.6	34.6	25.8	26.0	25.4	23.5
	8	35.1	34.8	34.0	32.9	25.8	23.5	24.9	25.8
	16	35.1	34.8	33.5	31.0	25.8	23.5	24.9	26.8
	32	35.1	34.8	33.5	31.0	25.8	23.5	24.9	26.8

C.5 Houix1 CO Ordre 1

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	57.9	58.2	57.3	57.0	57.1	57.8	58.1	57.4	57.0	57.0
	4	58.5	58.6	57.8	57.3	56.9	58.8	**59.0**	58.3	57.7	57.1
	8	58.5	58.4	58.1	57.6	57.4	58.8	58.9	58.6	57.8	57.6
	12	58.5	58.4	58.1	57.6	57.4	58.8	58.9	58.6	57.8	57.6
	16	58.5	58.4	58.2	57.9	58.6	58.8	58.6	58.6	58.1	58.1
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	51.2	51.8	52.0	51.5	51.9	**55.6**	54.6	53.3	52.4	47.9
	4	49.4	49.6	48.7	48.7	49.9	48.6	48.9	48.9	48.7	48.9
	8	49.7	49.6	49.7	49.5	49.6	50.7	50.6	49.5	48.9	49.4
	12	49.7	49.6	49.7	49.5	49.6	50.7	50.6	49.5	48.9	49.4
	16	49.6	49.5	49.5	50.0	50.0	48.2	45.4	42.6	43.4	49.5

C.6 Houix1 CO Ordre 2

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	57.1	56.6	56.8	56.1	56.3	56.7	56.2	56.1	55.1	54.4
	4	57.4	57.2	56.9	56.3	56.1	57.1	56.7	56.4	55.2	54.5
	8	57.6	57.5	57.0	56.4	56.1	57.4	56.9	56.5	55.2	54.8
	12	57.6	57.5	57.0	56.4	56.1	57.4	56.9	56.5	55.2	54.8
	16	**58.1**	57.8	57.4	56.6	56.5	57.6	56.9	56.5	55.4	55.0
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	58.4	57.4	**59.0**	58.8	58.6	51.8	50.9	50.9	49.9	48.5
	4	58.0	56.8	58.3	58.2	58.2	54.4	52.4	51.1	50.5	49.1
	8	57.4	56.6	58.1	58.4	58.5	54.2	52.6	51.8	50.4	49.4
	12	57.4	56.6	58.1	58.4	58.5	54.2	52.6	51.8	50.4	49.4
	16	57.5	56.7	58.0	57.9	58.3	48.8	45.0	43.3	43.6	49.6

C.7 Iowa CO Ordre 1

d		euclidean				cityblock			
Q		1	2	4	8	1	2	4	8
Tco	1	31.1	32.3	**36.0**	35.8	31.4	32.5	35.5	35.7
	2	31.9	33.0	36.0	34.3	32.0	33.0	35.7	34.2
	4	31.1	32.0	34.3	33.5	31.6	32.6	34.0	32.6
	8	31.1	31.6	33.5	28.8	31.6	32.4	33.0	29.2
	16	31.1	31.6	29.8	24.7	31.6	32.4	30.6	26.2
d		cosine				spearman			
Q		1	2	4	8	1	2	4	8
Tco	1	23.9	25.9	27.2	25.6	24.5	26.0	26.4	25.3
	2	23.7	25.2	26.3	24.8	23.8	23.9	26.1	26.8
	4	23.2	24.2	25.2	24.8	25.7	25.8	26.8	24.6
	8	23.2	24.4	25.3	22.2	25.7	19.6	26.7	25.6
	16	23.2	24.4	23.4	19.7	25.7	19.6	27.0	**30.2**

C.8 Iowa CO Ordre 2

d		euclidean				cityblock			
Q		1	2	4	8	1	2	4	8
Tco	1	36.2	38.7	38.0	37.6	38.0	40.4	**40.6**	39.9
	2	35.1	37.3	37.3	35.3	36.0	37.9	38.6	36.6
	4	33.2	35.2	35.2	34.0	34.2	35.9	35.7	34.2
	8	33.2	34.1	33.8	29.1	34.2	34.8	33.8	30.1
	16	33.2	34.1	30.3	25.0	34.2	34.8	31.5	27.3
d		cosine				spearman			
Q		1	2	4	8	1	2	4	8
Tco	1	32.7	35.7	**36.0**	34.3	30.4	30.3	33.9	34.9
	2	30.2	33.0	33.5	30.7	30.2	32.2	33.9	32.6
	4	28.1	30.8	30.6	29.2	31.5	32.0	33.7	28.9
	8	28.1	29.8	29.1	24.7	31.5	29.5	32.6	30.8
	16	28.1	29.8	26.1	21.2	31.5	29.5	31.9	34.8

C.9 RWC CO Ordre 1

d		euclidean				cityblock			
Q		1	2	4	8	1	2	4	8
Tco	1	38.8	38.8	38.7	38.6	39.0	38.9	38.9	39.0
	2	39.1	39.0	39.1	38.9	39.4	39.4	39.4	39.5
	4	38.9	38.8	38.9	39.1	39.0	39.1	39.0	39.2
	8	38.9	38.9	39.3	39.2	39.0	39.0	39.3	39.4
	16	38.9	38.9	39.2	39.4	39.0	39.0	39.1	39.4
d		cosine				spearman			
Q		1	2	4	8	1	2	4	8
Tco	1	41.7	41.5	**41.7**	41.5	10.1	10.1	40.6	41.2
	2	41.5	41.5	41.4	41.2	41.1	40.5	40.8	40.5
	4	40.6	40.5	40.3	40.3	40.5	40.1	39.3	40.5
	8	40.6	40.4	40.2	40.1	40.5	39.4	39.9	40.7
	16	40.6	40.4	40.2	40.2	40.5	39.4	39.6	39.2

C.10 RWC CO Ordre 2

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	38.7	38.7	38.6	38.5	38.7	38.9	38.8	39.1	39.8	40.4
	2	38.8	38.8	38.8	38.8	39.0	39.1	39.1	39.6	40.0	40.6
	4	38.7	38.6	38.9	39.2	39.4	39.0	39.1	39.4	39.9	**40.6**
	8	38.7	38.9	39.3	39.2	39.2	39.0	39.0	39.4	39.9	40.3
	16	38.7	38.9	39.3	39.0	38.9	39.0	39.0	39.3	39.3	39.5
	32	38.7	38.9	39.3	39.0	38.7	39.0	39.0	39.3	39.2	39.0
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	39.0	39.1	38.8	38.6	38.9	39.4	39.5	40.1	40.5	40.4
	2	39.1	39.2	38.8	38.9	38.7	40.0	40.4	**40.9**	40.7	40.1
	4	39.3	39.5	39.2	39.0	38.8	39.9	39.3	39.9	39.2	38.9
	8	39.3	39.5	39.3	38.9	38.8	39.9	39.0	39.4	38.9	38.7
	16	39.3	39.5	39.4	39.0	38.4	39.9	39.0	39.5	39.0	38.8
	32	39.3	39.5	39.4	38.8	38.4	39.9	39.0	39.5	39.1	39.1

C.11 SolosDb CO Ordre 1

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	12.0	14.1	14.1	13.9	13.7	11.9	13.6	13.6	13.4	13.2
	4	14.8	17.3	17.6	17.5	17.5	15.0	17.5	17.9	17.9	18.1
	8	16.7	19.4	19.7	19.9	20.5	16.7	20.1	20.6	20.9	21.4
	12	16.7	19.4	19.7	19.9	20.5	16.7	20.1	20.6	20.9	21.4
	16	18.4	21.8	22.8	23.2	23.7	18.2	22.7	24.0	24.4	**24.9**
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	24.1	26.7	27.1	26.0	24.0	18.0	22.4	23.2	21.4	19.2
	4	23.0	25.9	26.5	25.5	24.8	21.8	25.3	25.5	24.9	23.7
	8	23.9	26.0	26.2	26.6	26.5	24.5	26.3	28.0	28.9	27.0
	12	23.9	26.0	26.2	26.6	26.5	24.5	26.3	28.0	28.9	27.0
	16	25.4	27.7	28.7	28.7	29.4	24.5	27.6	30.1	**30.4**	30.3

C.12 SolosDb CO Ordre 2

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	14.1	17.6	15.6	14.9	15.2	13.9	19.4	17.8	16.4	16.5
	4	15.5	20.2	18.6	17.6	17.4	14.8	21.3	19.9	18.2	17.8
	8	17.0	22.0	20.7	20.1	20.2	15.7	23.0	21.8	20.1	19.4
	12	17.0	22.0	20.7	20.1	20.2	15.7	23.0	21.8	20.1	19.4
	16	18.2	23.6	23.3	23.0	22.9	16.2	**24.7**	23.9	22.3	21.2
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	24.6	23.7	19.6	18.9	19.5	25.3	25.6	25.0	24.5	26.9
	4	25.0	26.3	23.2	22.4	22.4	25.1	27.6	27.0	26.3	28.3
	8	25.3	27.6	25.2	24.8	25.3	25.2	28.7	27.9	27.5	29.8
	12	25.3	27.6	25.2	24.8	25.3	25.2	28.7	27.9	27.5	29.8
	16	26.4	28.9	28.1	27.8	28.0	25.1	29.1	29.5	29.3	**31.1**

C.13 Digit CO Ordre 1

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	34.5	40.2	42.9	41.6	39.7	37.4	41.9	43.2	41.7	40.2
	2	39.1	43.2	46.5	44.9	43.3	39.8	45.5	46.9	44.9	43.2
	4	39.1	49.2	**53.4**	52.6	51.0	39.8	48.1	50.1	48.3	46.8
	8	39.1	49.2	49.2	45.9	43.9	39.8	48.1	49.4	45.0	42.9
	16	39.1	49.2	49.2	39.0	35.8	39.8	48.1	49.4	41.9	38.1
	32	39.1	49.2	49.2	39.0	32.2	39.8	48.1	49.4	41.9	35.5
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	36.5	33.2	35.9	34.8	32.8	37.3	37.4	34.8	33.7	33.3
	2	40.3	35.0	38.5	37.6	36.5	30.6	38.3	41.9	42.9	42.4
	4	40.3	40.7	**45.5**	45.3	44.6	30.6	18.5	24.0	26.0	28.3
	8	40.3	40.7	42.0	40.2	39.8	30.7	18.5	19.9	21.9	23.4
	16	40.3	40.7	42.0	34.4	32.8	30.6	18.5	19.9	24.1	23.7
	32	40.3	40.7	42.0	34.4	29.9	30.6	18.5	19.9	24.1	26.0

C.14 digit CO Ordre 2

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	41.4	48.4	50.3	50.4	49.1	44.1	51.5	52.1	51.7	50.7
	2	43.7	50.4	53.3	54.9	56.2	45.7	52.9	54.5	56.0	56.9
	4	43.7	55.2	59.6	**60.2**	56.5	45.7	55.3	57.7	58.0	56.4
	8	43.7	55.2	54.0	51.0	46.1	45.7	55.3	55.4	53.2	51.4
	16	43.7	55.2	54.0	42.9	37.8	45.7	55.3	55.4	49.3	46.4
	32	43.7	55.2	54.0	42.9	35.3	45.7	55.3	55.4	49.3	44.7
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	45.5	40.3	41.8	41.9	41.5	43.7	43.8	42.2	45.3	52.8
	2	46.7	42.5	45.3	47.2	49.4	38.6	45.6	48.4	50.8	**53.5**
	4	46.7	47.6	52.2	53.4	50.9	38.6	28.6	38.0	39.6	40.7
	8	46.7	47.6	47.9	45.8	42.1	38.6	28.6	29.8	36.3	41.4
	16	46.7	47.6	47.9	38.9	34.9	38.6	28.6	29.8	37.4	44.9
	32	46.7	47.6	47.9	38.9	32.9	38.6	28.6	29.8	37.4	45.3

C.15 Insects CO Ordre 1

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	35.9	34.0	35.9	36.0	36.3	39.8	38.2	40.6	41.5	42.0
	2	37.9	36.0	38.6	40.0	40.8	40.2	39.3	42.9	46.8	48.4
	4	37.9	37.8	41.3	43.7	45.3	40.2	39.0	43.5	49.5	52.0
	8	37.9	37.8	49.5	56.7	59.1	40.2	39.0	47.5	56.8	60.4
	16	37.9	37.8	49.5	59.5	**62.3**	40.2	39.0	47.5	56.2	60.0
	32	37.9	37.8	49.5	59.5	61.7	40.2	39.0	47.5	56.2	57.5
	64	37.9	37.8	49.5	59.5	61.7	40.2	39.0	47.5	56.2	57.5
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	36.0	34.1	36.1	36.1	36.4	0.0	40.0	42.5	38.3	39.4
	2	38.0	36.1	38.7	40.2	41.0	37.1	40.3	42.9	53.6	52.4
	4	38.0	37.9	41.4	43.9	45.5	37.1	43.6	45.1	48.6	47.7
	8	38.0	37.9	49.7	56.9	59.3	37.1	43.6	51.4	51.3	45.7
	16	38.0	37.9	49.7	59.8	**62.5**	37.1	43.6	51.4	61.6	44.6
	32	38.0	37.9	49.7	59.8	61.9	37.1	43.6	51.4	61.6	59.8
	64	38.0	37.9	49.7	59.8	61.9	37.1	43.6	51.4	61.6	59.8

C.16 Insects CO Ordre 2

d		euclidean					cityblock				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	46.4	44.4	47.8	47.5	48.0	48.9	47.5	49.9	51.1	51.7
	2	47.8	45.7	49.7	51.0	53.2	49.2	49.1	52.5	55.5	56.5
	4	47.8	47.1	54.6	61.7	64.3	49.2	48.8	56.0	62.7	63.4
	8	47.8	47.1	59.2	66.3	66.5	49.2	48.8	58.4	65.5	64.1
	16	47.8	47.1	59.2	**66.8**	65.4	49.2	48.8	58.4	65.0	62.6
	32	47.8	47.1	59.2	**66.8**	65.2	49.2	48.8	58.4	65.0	62.3
	64	47.8	47.1	59.2	**66.8**	65.2	49.2	48.8	58.4	65.0	62.3
d		cosine					spearman				
Q		1	2	4	8	16	1	2	4	8	16
Tco	1	46.6	44.6	48.0	47.8	48.5	57.4	57.1	61.0	65.2	71.5
	2	48.0	45.8	49.9	51.4	53.7	54.3	59.1	63.8	70.5	**73.3**
	4	48.0	47.2	54.8	62.1	64.9	54.3	59.2	64.0	70.7	67.8
	8	48.0	47.2	59.4	66.8	67.0	54.3	59.2	65.0	71.5	66.4
	16	48.0	47.2	59.4	67.3	65.9	54.3	59.2	65.0	71.3	63.0
	32	48.0	47.2	59.4	67.3	65.7	54.3	59.2	65.0	71.3	62.3
	64	48.0	47.2	59.4	67.3	65.7	54.3	59.2	65.0	71.3	62.3

Annexe D

Détails de la base Gygi Extended

- Tous les sons sont échantillonnés a 44100 Hz
- Nombre total d'éléments : 294
- Nombre total de classes : 50
- Statistiques sur les durées (en secondes) (max,min,moy,écart-type) : 4.80 0.21 2.40 0.92
- Statistiques sur les classes (max,min,moy,écart-type) : 12 4 5.88 1.77

NOM	CLASSE	SOURCE
AIRPLAN	AIRPLAN	www.freesound.org
AIRPLAN	AIRPLAN	www.freesound.org
AIRPLANA	AIRPLAN	www.freesound.org
AIRPLANA	AIRPLAN	www.freesound.org
AIRPLAN_1_1	AIRPLAN	www.freesound.org
AIRPLAN_1_1	AIRPLAN	www.freesound.org
AIRPLAN_1_2	AIRPLAN	www.freesound.org
AIRPLAN_1_2	AIRPLAN	www.freesound.org
AIRPLAN_2_1	AIRPLAN	www.freesound.org
AIRPLAN_2_1	AIRPLAN	www.freesound.org
AIRPLAN_3_1	AIRPLAN	www.freesound.org
AIRPLAN_3_1	AIRPLAN	www.freesound.org
AIRPLAN_4_1	AIRPLAN	www.freesound.org
AIRPLAN_4_1	AIRPLAN	www.freesound.org
AIRPLAN_5_1	AIRPLAN	www.freesound.org
AIRPLAN_5_1	AIRPLAN	www.freesound.org
AIRPLAN_6_1	AIRPLAN	www.freesound.org
AIRPLAN_6_1	AIRPLAN	www.freesound.org
AXE	AXE	www.freesound.org
AXE	AXE	www.freesound.org
AXEA	AXE	www.freesound.org
AXEA	AXE	www.freesound.org
AXE_1_1	AXE	www.freesound.org
AXE_1_1	AXE	www.freesound.org
AXE_1_2	AXE	www.freesound.org
AXE_1_2	AXE	www.freesound.org
AXE_1_3	AXE	www.freesound.org
AXE_1_3	AXE	www.freesound.org
AXE_2_1	AXE	www.freesound.org
AXE_2_1	AXE	www.freesound.org
AXE_2_2	AXE	www.freesound.org
AXE_2_2	AXE	www.freesound.org
BABYA	BABY	www.freesound.org
BABYA	BABY	www.freesound.org
BABYCRY	BABY	www.freesound.org
BABYCRY	BABY	www.freesound.org
BABY_1_1	BABY	www.freesound.org
BABY_1_1	BABY	www.freesound.org
BABY_1_2	BABY	www.freesound.org
BABY_1_2	BABY	www.freesound.org
BABY_2_1	BABY	www.freesound.org
BABY_2_1	BABY	www.freesound.org
BABY_3_1	BABY	www.freesound.org
BABY_3_1	BABY	www.freesound.org
BABY_4_1	BABY	www.freesound.org
BABY_4_1	BABY	www.freesound.org
BABY_4_2	BABY	www.freesound.org
BABY_4_2	BABY	www.freesound.org
BABY_5_1	BABY	www.freesound.org
BABY_5_1	BABY	www.freesound.org
BBALL	BBALL	www.freesound.org
BBALL	BBALL	www.freesound.org
BBALLA	BBALL	www.freesound.org
BBALLA	BBALL	www.freesound.org
BBALL_1_1	BBALL	www.freesound.org
BBALL_1_1	BBALL	www.freesound.org
BBALL_2_1	BBALL	www.freesound.org
BBALL_2_1	BBALL	www.freesound.org
BBALL_3_1	BBALL	www.freesound.org
BBALL_3_1	BBALL	www.freesound.org
BBALL_4_1	BBALL	www.freesound.org
BBALL_4_1	BBALL	www.freesound.org
BBALL_5_1	BBALL	www.freesound.org
BBALL_5_1	BBALL	www.freesound.org
BELLS	BELLS	www.freesound.org
BELLS	BELLS	www.freesound.org
BELLSA	BELLS	www.freesound.org
BELLSA	BELLS	www.freesound.org
BELL_1_1	BELLS	www.freesound.org
BELL_1_1	BELLS	www.freesound.org
BELL_2_1	BELLS	www.freesound.org
BELL_2_1	BELLS	www.freesound.org
BELL_3_1	BELLS	www.freesound.org
BELL_3_1	BELLS	www.freesound.org
BIRD	BIRD	www.freesound.org
BIRD	BIRD	www.freesound.org
BIRDA	BIRD	www.freesound.org
BIRDA	BIRD	www.freesound.org
BIRD_1_1	BIRD	www.freesound.org
BIRD_1_1	BIRD	www.freesound.org
BIRD_1_2	BIRD	www.freesound.org
BIRD_1_2	BIRD	www.freesound.org
BIRD_2_1	BIRD	www.freesound.org
BIRD_2_1	BIRD	www.freesound.org
BIRD_2_2	BIRD	www.freesound.org
BIRD_2_2	BIRD	www.freesound.org
BIRD_3_1	BIRD	www.freesound.org
BIRD_3_1	BIRD	www.freesound.org
BIRD_4_1	BIRD	www.freesound.org
BIRD_4_1	BIRD	www.freesound.org
BIRD_5_1	BIRD	www.freesound.org
BIRD_5_1	BIRD	www.freesound.org
BOWLIN	BOWLIN	www.freesound.org
BOWLIN	BOWLIN	www.freesound.org
BOWLINA	BOWLIN	www.freesound.org
BOWLINA	BOWLIN	www.freesound.org
BOWLIN_1_1	BOWLIN	www.freesound.org
BOWLIN_1_1	BOWLIN	www.freesound.org
BOWLIN_2_1	BOWLIN	www.freesound.org
BOWLIN_2_1	BOWLIN	www.freesound.org
BOWLIN_3_1	BOWLIN	www.soundbible.com
BUBBLESA	BUBBLES	www.freesound.org
BUBBLESA	BUBBLES	www.freesound.org
BUBBLES_1_1	BUBBLES	www.freesound.org
BUBBLES_1_1	BUBBLES	www.freesound.org
BUBBLES_2_1	BUBBLES	www.freesound.org
BUBBLES_2_1	BUBBLES	www.freesound.org
BUBBLES_3_1	BUBBLES	www.freesound.org
BUBBLES_3_1	BUBBLES	www.freesound.org
BUBBLES_4_1	BUBBLES	www.freesound.org
BUBBLES_4_1	BUBBLES	www.freesound.org
BUBBLES_5_1	BUBBLES	www.freesound.org
BUBBLES_5_1	BUBBLES	www.freesound.org
BUBBLE_1_1	BUBBLES	www.freesound.org
BUBBLE_1_1	BUBBLES	www.freesound.org
BUBBLE_2_1	BUBBLES	www.freesound.org
BUBBLE_2_1	BUBBLES	www.freesound.org
BUBBLE_3_1	BUBBLES	www.freesound.org
BUBBLE_3_1	BUBBLES	www.freesound.org
BUBBLE_4_1	BUBBLES	www.freesound.org
BUBBLE_4_1	BUBBLES	www.freesound.org
BUBBLE_5_1	BUBBLES	www.freesound.org
BUBBLE_5_1	BUBBLES	www.freesound.org
BUBLES	BUBBLES	www.freesound.org
BUBLES	BUBBLES	www.freesound.org
CARSTART	CARST	www.freesound.org
CARSTART	CARST	www.freesound.org
CARSTRTA	CARST	www.freesound.org
CARSTRTA	CARST	www.freesound.org
CARST_1_1	CARST	www.freesound.org
CARST_1_1	CARST	www.freesound.org
CARST_2_1	CARST	www.freesound.org
CARST_2_1	CARST	www.freesound.org
CARST_3_1	CARST	www.freesound.org
CARST_3_1	CARST	www.freesound.org
CARST_4_1	CARST	www.freesound.org
CARST_4_1	CARST	www.freesound.org
CARST_5_1	CARST	www.freesound.org
CARST_5_1	CARST	www.freesound.org
CARST_6_1	CARST	www.freesound.org
CARST_6_1	CARST	www.freesound.org
CARST_7_1	CARST	www.freesound.org
CARST_7_1	CARST	www.freesound.org
CATA	CAT	www.freesound.org
CATA	CAT	www.freesound.org
CATX	CAT	www.freesound.org
CATX	CAT	www.freesound.org
CAT_1_1	CAT	www.freesound.org
CAT_1_1	CAT	www.freesound.org
CAT_2_1	CAT	www.freesound.org
CAT_2_1	CAT	www.freesound.org
CAT_3_1	CAT	www.freesound.org
CAT_3_1	CAT	www.freesound.org
CAT_4_1	CAT	www.freesound.org
CAT_4_1	CAT	www.freesound.org
CAT_5_1	CAT	www.freesound.org
CAT_5_1	CAT	www.freesound.org
CAT_5_2	CAT	www.freesound.org
CAT_5_2	CAT	www.freesound.org
CAT_6_1	CAT	www.freesound.org
CAT_6_1	CAT	www.freesound.org
CLAP	CLAP	www.freesound.org
CLAP	CLAP	www.freesound.org

CLAPSA	CLAP	www.freesound.org
CLAPSA	CLAP	www.freesound.org
CLAP_1_1	CLAP	www.freesound.org
CLAP_1_1	CLAP	www.freesound.org
CLAP_2_1	CLAP	www.freesound.org
CLAP_2_1	CLAP	www.freesound.org
CLAP_3_1	CLAP	www.freesound.org
CLAP_3_1	CLAP	www.freesound.org
CLAP_4_1	CLAP	www.freesound.org
CLAP_4_1	CLAP	www.freesound.org
CLAP_5_1	CLAP	www.freesound.org
CLAP_5_1	CLAP	www.freesound.org
CLOCK	CLOCK	www.freesound.org
CLOCK	CLOCK	www.freesound.org
CLOCKA	CLOCK	www.freesound.org
CLOCKA	CLOCK	www.freesound.org
CLOCK_1_1	CLOCK	www.freesound.org
CLOCK_1_1	CLOCK	www.freesound.org
CLOCK_2_1	CLOCK	www.freesound.org
CLOCK_2_1	CLOCK	www.freesound.org
CLOCK_3_1	CLOCK	www.freesound.org
CLOCK_3_1	CLOCK	www.freesound.org
CLOCK_4_1	CLOCK	www.freesound.org
CLOCK_4_1	CLOCK	www.freesound.org
CLOCK_5_1	CLOCK	www.freesound.org
CLOCK_5_1	CLOCK	www.freesound.org
CLOCK_6_1	CLOCK	www.freesound.org
CLOCK_6_1	CLOCK	www.freesound.org
CLOCK_7_1	CLOCK	www.freesound.org
CLOCK_7_1	CLOCK	www.freesound.org
CLOCK_8_1	CLOCK	www.freesound.org
CLOCK_8_1	CLOCK	www.freesound.org
COPTER	COPTER	www.freesound.org
COPTER	COPTER	www.freesound.org
COPTERA	COPTER	www.freesound.org
COPTERA	COPTER	www.freesound.org
COPTER_1_1	COPTER	www.freesound.org
COPTER_1_1	COPTER	www.freesound.org
COPTER_2_1	COPTER	www.freesound.org
COPTER_2_1	COPTER	www.freesound.org
COPTER_3_1	COPTER	www.freesound.org
COPTER_3_1	COPTER	www.freesound.org
COPTER_4_1	COPTER	www.freesound.org
COPTER_4_1	COPTER	www.freesound.org
COPTER_5_1	COPTER	www.freesound.org
COPTER_5_1	COPTER	www.freesound.org
COPTER_6_1	COPTER	www.freesound.org
COPTER_6_1	COPTER	www.freesound.org
COUGH	COUGH	www.freesound.org
COUGH	COUGH	www.freesound.org
COUGHA	COUGH	www.freesound.org
COUGHA	COUGH	www.freesound.org
COUGH_1_1	COUGH	www.freesound.org
COUGH_1_1	COUGH	www.freesound.org
COUGH_2_1	COUGH	www.freesound.org
COUGH_2_1	COUGH	www.freesound.org
COUGH_3_1	COUGH	www.freesound.org
COUGH_3_1	COUGH	www.freesound.org
COWA	COW	www.freesound.org
COWA	COW	www.freesound.org
COWX	COW	www.freesound.org
COWX	COW	www.freesound.org
COW_1_1	COW	www.freesound.org
COW_1_1	COW	www.freesound.org
COW_2_1	COW	www.soundbible.com
COW_3_1	COW	www.weblust.com
CYMBALA	CYMBAL	www.freesound.org
CYMBALA	CYMBAL	www.freesound.org
CYMBAL_1_1	CYMBAL	www.freesound.org
CYMBAL_1_1	CYMBAL	www.freesound.org
CYMBAL_2_1	CYMBAL	www.freesound.org
CYMBAL_2_1	CYMBAL	www.freesound.org
CYMBAL_3_1	CYMBAL	www.freesound.org
CYMBAL_3_1	CYMBAL	www.freesound.org
CYMBOL	CYMBAL	www.freesound.org
CYMBOL	CYMBAL	www.freesound.org
DOGA	DOG	www.freesound.org
DOGA	DOG	www.freesound.org
DOGX2	DOG	www.freesound.org
DOGX2	DOG	www.freesound.org
DOG_1_1	DOG	www.freesound.org
DOG_1_1	DOG	www.freesound.org
DOG_1_2	DOG	www.freesound.org
DOG_1_2	DOG	www.freesound.org
DOG_2_1	DOG	www.freesound.org
DOG_2_1	DOG	www.freesound.org
DOOROC	DOOROC	www.freesound.org
DOOROC	DOOROC	www.freesound.org
DOOROCA	DOOROC	www.freesound.org
DOOROCA	DOOROC	www.freesound.org
DOOROC_1_1	DOOROC	www.freesound.org
DOOROC_1_1	DOOROC	www.freesound.org
DOOROC_2_1	DOOROC	www.freesound.org
DOOROC_2_1	DOOROC	www.freesound.org
DOOROC_3_1	DOOROC	www.freesound.org
DOOROC_3_1	DOOROC	www.freesound.org
DRUMS	DRUMS	www.freesound.org
DRUMS	DRUMS	www.freesound.org
DRUMSA	DRUMS	www.freesound.org
DRUMSA	DRUMS	www.freesound.org
DRUMS_1_1	DRUMS	www.freesound.org
DRUMS_1_1	DRUMS	www.freesound.org
DRUMS_2_1	DRUMS	www.freesound.org
DRUMS_2_1	DRUMS	www.freesound.org
DRUMS_3_1	DRUMS	www.freesound.org
DRUMS_3_1	DRUMS	www.freesound.org
ELECSAW	ELECSAW	www.freesound.org
ELECSAW	ELECSAW	www.freesound.org
ELECSAWA	ELECSAW	www.freesound.org
ELECSAWA	ELECSAW	www.freesound.org
ELECSAW_1_1	ELECSAW	www.freesound.org
ELECSAW_1_1	ELECSAW	www.freesound.org
ELECSAW_2_1	ELECSAW	www.weblust.com
FOOTSTP	FOOTSTP	www.freesound.org
FOOTSTP	FOOTSTP	www.freesound.org
FOOTSTPA	FOOTSTP	www.freesound.org
FOOTSTPA	FOOTSTP	www.freesound.org
FOOTSTP_1_1	FOOTSTP	www.freesound.org
FOOTSTP_1_1	FOOTSTP	www.freesound.org
FOOTSTP_2_1	FOOTSTP	www.freesound.org
FOOTSTP_2_1	FOOTSTP	www.freesound.org
FOOTSTP_3_1	FOOTSTP	www.freesound.org
FOOTSTP_3_1	FOOTSTP	www.freesound.org
GLASSBR	GLASSBR	www.freesound.org
GLASSBR	GLASSBR	www.freesound.org
GLASSBRA	GLASSBR	www.freesound.org
GLASSBRA	GLASSBR	www.freesound.org
GLASSBR_1_1	GLASSBR	www.freesound.org
GLASSBR_1_1	GLASSBR	www.freesound.org
GLASSBR_2_1	GLASSBR	www.freesound.org
GLASSBR_2_1	GLASSBR	www.freesound.org
GLASSBR_3_1	GLASSBR	www.freesound.org
GLASSBR_3_1	GLASSBR	www.freesound.org
GUN	GUN	www.freesound.org
GUN	GUN	www.freesound.org
GUNA	GUN	www.freesound.org
GUNA	GUN	www.freesound.org
GUN_1_1	GUN	www.freesound.org
GUN_1_1	GUN	www.freesound.org
GUN_2_1	GUN	www.freesound.org
GUN_2_1	GUN	www.freesound.org
GUN_3_1	GUN	www.freesound.org
GUN_3_1	GUN	www.freesound.org
HARP	HARP	www.freesound.org
HARP	HARP	www.freesound.org
HARPA	HARP	www.freesound.org
HARPA	HARP	www.freesound.org
HARP_1_1	HARP	www.soundjay.com
HARP_1_2	HARP	www.soundjay.com
HARP_2_1	HARP	archive.cnmat.berkeley.edu
HORSERNA	HORSERN	www.freesound.org
HORSERNA	HORSERN	www.freesound.org
HORSERN_1_1	HORSERN	www.freesound.org
HORSERN_1_1	HORSERN	www.freesound.org
HORSERN_3_1	HORSERN	cd.textfiles.com
HORSERUN	HORSERN	www.freesound.org
HORSERUN	HORSERN	www.freesound.org
HORSEWIN	HORSEWN	www.freesound.org
HORSEWIN	HORSEWN	www.freesound.org
HORSEWNA	HORSEWN	www.freesound.org
HORSEWNA	HORSEWN	www.freesound.org
HORSEWN_1_1	HORSEWN	www.freesound.org
HORSEWN_1_1	HORSEWN	www.freesound.org
HORSEWN_2_1	HORSEWN	www.freesound.org

HORSEWN_2_1	HORSEWN	www.freesound.org
HORSEWN_3_1	HORSEWN	www.weblust.com
ICEDROP	ICEDROP	www.freesound.org
ICEDROP	ICEDROP	www.freesound.org
ICEDROPA	ICEDROP	www.freesound.org
ICEDROPA	ICEDROP	www.freesound.org
ICEDROP_1_1	ICEDROP	www.freesound.org
ICEDROP_1_1	ICEDROP	www.freesound.org
ICEDROP_1_2	ICEDROP	www.freesound.org
ICEDROP_1_2	ICEDROP	www.freesound.org
ICEDROP_2_1	ICEDROP	www.freesound.org
ICEDROP_2_1	ICEDROP	www.freesound.org
KEYBRD	KEYBRD	www.freesound.org
KEYBRD	KEYBRD	www.freesound.org
KEYBRDA	KEYBRD	www.freesound.org
KEYBRDA	KEYBRD	www.freesound.org
KEYBRD_1_1	KEYBRD	www.freesound.org
KEYBRD_1_1	KEYBRD	www.freesound.org
KEYBRD_1_2	KEYBRD	www.freesound.org
KEYBRD_1_2	KEYBRD	www.freesound.org
KEYBRD_1_3	KEYBRD	www.freesound.org
KEYBRD_1_3	KEYBRD	www.freesound.org
LAUGH	LAUGH	www.freesound.org
LAUGH	LAUGH	www.freesound.org
LAUGHA	LAUGH	www.freesound.org
LAUGHA	LAUGH	www.freesound.org
LAUGH_1_1	LAUGH	www.freesound.org
LAUGH_1_1	LAUGH	www.freesound.org
LAUGH_2_1	LAUGH	www.freesound.org
LAUGH_2_1	LAUGH	www.freesound.org
LAUGH_3_1	LAUGH	www.freesound.org
LAUGH_3_1	LAUGH	www.freesound.org
MATCH	MATCH	www.freesound.org
MATCH	MATCH	www.freesound.org
MATCHA	MATCH	www.freesound.org
MATCHA	MATCH	www.freesound.org
MATCH_1_1	MATCH	www.freesound.org
MATCH_1_1	MATCH	www.freesound.org
MATCH_2_1	MATCH	www.freesound.org
MATCH_2_1	MATCH	www.freesound.org
MATCH_3_1	MATCH	www.freesound.org
MATCH_3_1	MATCH	www.freesound.org
PEELOUT	PEELOUT	www.freesound.org
PEELOUT	PEELOUT	www.freesound.org
PEELOUTA	PEELOUT	www.freesound.org
PEELOUTA	PEELOUT	www.freesound.org
PEELOUT_1_1	PEELOUT	www.freesound.org
PEELOUT_1_1	PEELOUT	www.freesound.org
PEELOUT_2_1	PEELOUT	www.freesound.org
PEELOUT_2_1	PEELOUT	www.freesound.org
PEELOUT_3_1	PEELOUT	www.freesound.org
PEELOUT_3_1	PEELOUT	www.freesound.org
PHONE	PHONE	www.freesound.org
PHONE	PHONE	www.freesound.org
PHONEA	PHONE	www.freesound.org
PHONEA	PHONE	www.freesound.org
PHONE_1_1	PHONE	www.freesound.org
PHONE_1_1	PHONE	www.freesound.org
PHONE_2_1	PHONE	www.freesound.org
PHONE_2_1	PHONE	www.freesound.org
PHONE_3_1	PHONE	www.freesound.org
PHONE_3_1	PHONE	www.freesound.org
PINGP	PINGP	www.freesound.org
PINGP	PINGP	www.freesound.org
PINGPA	PINGP	www.freesound.org
PINGPA	PINGP	www.freesound.org
PINGP_1_1	PINGP	www.freesound.org
PINGP_1_1	PINGP	www.freesound.org
PINGP_2_1	PINGP	www.freesound.org
PINGP_2_1	PINGP	www.freesound.org
PINGP_3_1	PINGP	www.freesound.org
PINGP_3_1	PINGP	www.freesound.org
POUR2	POUR	www.freesound.org
POUR2	POUR	www.freesound.org
POURA	POUR	www.freesound.org
POURA	POUR	www.freesound.org
POUR_1_1	POUR	www.freesound.org
POUR_1_1	POUR	www.freesound.org
POUR_2_1	POUR	www.freesound.org
POUR_2_1	POUR	www.freesound.org
POUR_3_1	POUR	www.freesound.org
POUR_3_1	POUR	www.freesound.org
RAIN	RAIN	www.freesound.org
RAIN	RAIN	www.freesound.org
RAINA	RAIN	www.freesound.org
RAINA	RAIN	www.freesound.org
RAIN_1_1	RAIN	www.freesound.org
RAIN_1_1	RAIN	www.freesound.org
RAIN_2_1	RAIN	www.freesound.org
RAIN_2_1	RAIN	www.freesound.org
RAIN_3_1	RAIN	www.freesound.org
RAIN_3_1	RAIN	www.freesound.org
RAIN_4_1	RAIN	www.freesound.org
RAIN_4_1	RAIN	www.freesound.org
RAIN_5_1	RAIN	www.freesound.org
RAIN_5_1	RAIN	www.freesound.org
ROOSTER	ROOSTER	www.freesound.org
ROOSTER	ROOSTER	www.freesound.org
ROOSTERA	ROOSTER	www.freesound.org
ROOSTERA	ROOSTER	www.freesound.org
ROOSTER_1_1	ROOSTER	www.freesound.org
ROOSTER_1_1	ROOSTER	www.freesound.org
ROOSTER_2_1	ROOSTER	www.freesound.org
ROOSTER_2_1	ROOSTER	www.freesound.org
ROOSTER_3_1	ROOSTER	www.freesound.org
ROOSTER_3_1	ROOSTER	www.freesound.org
SCISSOR	SCISSOR	www.freesound.org
SCISSOR	SCISSOR	www.freesound.org
SCISSORA	SCISSOR	www.freesound.org
SCISSORA	SCISSOR	www.freesound.org
SCISSOR_1_1	SCISSOR	www.freesound.org
SCISSOR_1_1	SCISSOR	www.freesound.org
SCISSOR_1_2	SCISSOR	www.freesound.org
SCISSOR_1_2	SCISSOR	www.freesound.org
SCISSOR_1_3	SCISSOR	www.freesound.org
SCISSOR_1_3	SCISSOR	www.freesound.org
SHEEPA	SHEEP	www.freesound.org
SHEEPA	SHEEP	www.freesound.org
SHEEPBAA	SHEEP	www.freesound.org
SHEEPBAA	SHEEP	www.freesound.org
SHEEP_1_1	SHEEP	www.freesound.org
SHEEP_1_1	SHEEP	www.freesound.org
SHEEP_2_1	SHEEP	www.freesound.org
SHEEP_2_1	SHEEP	www.freesound.org
SHEEP_3_1	SHEEP	mixonline.com
SIREN	SIREN	www.freesound.org
SIREN	SIREN	www.freesound.org
SIRENA	SIREN	www.freesound.org
SIRENA	SIREN	www.freesound.org
SIREN_1_1	SIREN	www.freesound.org
SIREN_1_1	SIREN	www.freesound.org
SIREN_2_1	SIREN	www.freesound.org
SIREN_2_1	SIREN	www.freesound.org
SIREN_3_1	SIREN	www.freesound.org
SIREN_3_1	SIREN	www.freesound.org
SNEEZE	SNEEZE	www.freesound.org
SNEEZE	SNEEZE	www.freesound.org
SNEEZEA	SNEEZE	www.freesound.org
SNEEZEA	SNEEZE	www.freesound.org
SNEEZE_1_1	SNEEZE	www.freesound.org
SNEEZE_1_1	SNEEZE	www.freesound.org
SNEEZE_2_1	SNEEZE	www.freesound.org
SNEEZE_2_1	SNEEZE	www.freesound.org
SNEEZE_3_1	SNEEZE	www.freesound.org
SNEEZE_3_1	SNEEZE	www.freesound.org
SPLASH	SPLASH	www.freesound.org
SPLASH	SPLASH	www.freesound.org
SPLASHA	SPLASH	www.freesound.org
SPLASHA	SPLASH	www.freesound.org
SPLASH_1_1	SPLASH	www.freesound.org
SPLASH_1_1	SPLASH	www.freesound.org
SPLASH_1_2	SPLASH	www.freesound.org
SPLASH_1_2	SPLASH	www.freesound.org
SPLASH_2_1	SPLASH	www.freesound.org
SPLASH_2_1	SPLASH	www.freesound.org
THUNDER	THUNDER	www.freesound.org
THUNDER	THUNDER	www.freesound.org
THUNDERA	THUNDER	www.freesound.org
THUNDERA	THUNDER	www.freesound.org
THUNDER_1_1	THUNDER	www.freesound.org
THUNDER_1_1	THUNDER	www.freesound.org
THUNDER_2_1	THUNDER	www.freesound.org
THUNDER_2_1	THUNDER	www.freesound.org
THUNDER_3_1	THUNDER	www.freesound.org

THUNDER_3_1	THUNDER	www.freesound.org
TOILET	TOILET	www.freesound.org
TOILET	TOILET	www.freesound.org
TOILETA	TOILET	www.freesound.org
TOILETA	TOILET	www.freesound.org
TOILET_1_1	TOILET	www.freesound.org
TOILET_1_1	TOILET	www.freesound.org
TOILET_2_1	TOILET	www.freesound.org
TOILET_2_1	TOILET	www.freesound.org
TOILET_3_1	TOILET	www.freesound.org
TOILET_3_1	TOILET	www.freesound.org
TRAFJAM	TRAFJAM	www.freesound.org
TRAFJAM	TRAFJAM	www.freesound.org
TRAFJAMA	TRAFJAM	www.freesound.org
TRAFJAMA	TRAFJAM	www.freesound.org
TRAFJAM_1_1	TRAFJAM	www.freesound.org
TRAFJAM_1_1	TRAFJAM	www.freesound.org
TRAFJAM_2_1	TRAFJAM	www.freesound.org
TRAFJAM_2_1	TRAFJAM	www.freesound.org
TRAFJAM_3_1	TRAFJAM	www.soundbible.com
TRAFJAM_3_2	TRAFJAM	www.soundbible.com
TRAIN	TRAIN	www.freesound.org
TRAIN	TRAIN	www.freesound.org
TRAINA	TRAIN	www.freesound.org
TRAINA	TRAIN	www.freesound.org
TRAIN_1_1	TRAIN	www.freesound.org
TRAIN_1_1	TRAIN	www.freesound.org
TRAIN_1_2	TRAIN	www.freesound.org
TRAIN_1_2	TRAIN	www.freesound.org
TRAIN_2_1	TRAIN	www.freesound.org
TRAIN_2_1	TRAIN	www.freesound.org
TRAIN_3_1	TRAIN	www.freesound.org
TRAIN_3_1	TRAIN	www.freesound.org
TRAIN_3_2	TRAIN	www.freesound.org
TRAIN_3_2	TRAIN	www.freesound.org
TRAIN_4_1	TRAIN	www.freesound.org
TRAIN_4_1	TRAIN	www.freesound.org
TYPEWRI	TYPEWRI	www.freesound.org
TYPEWRI	TYPEWRI	www.freesound.org
TYPEWRIA	TYPEWRI	www.freesound.org
TYPEWRIA	TYPEWRI	www.freesound.org
TYPEWRI_1_1	TYPEWRI	www.freesound.org
TYPEWRI_1_1	TYPEWRI	www.freesound.org
TYPEWRI_2_1	TYPEWRI	www.freesound.org
TYPEWRI_2_1	TYPEWRI	www.freesound.org
TYPEWRI_3_1	TYPEWRI	www.freesound.org
TYPEWRI_3_1	TYPEWRI	www.freesound.org
WAVES	WAVES	www.freesound.org
WAVES	WAVES	www.freesound.org
WAVESA	WAVES	www.freesound.org
WAVESA	WAVES	www.freesound.org
WAVES_1_1	WAVES	www.freesound.org
WAVES_1_1	WAVES	www.freesound.org
WAVES_2_1	WAVES	www.freesound.org
WAVES_2_1	WAVES	www.freesound.org
WAVES_3_1	WAVES	www.freesound.org
WAVES_3_1	WAVES	www.freesound.org
WHISTLE	WHISTLE	www.freesound.org
WHISTLE	WHISTLE	www.freesound.org
WHISTLEA	WHISTLE	www.freesound.org
WHISTLEA	WHISTLE	www.freesound.org
WHISTLE_1_1	WHISTLE	www.freesound.org
WHISTLE_1_1	WHISTLE	www.freesound.org
WHISTLE_2_1	WHISTLE	www.freesound.org
WHISTLE_2_1	WHISTLE	www.freesound.org
WHISTLE_2_2	WHISTLE	www.freesound.org
WHISTLE_2_2	WHISTLE	www.freesound.org
WIPERS	WIPERS	www.freesound.org
WIPERS	WIPERS	www.freesound.org
WIPERSA	WIPERS	www.freesound.org
WIPERSA	WIPERS	www.freesound.org
WIPERS_1_1	WIPERS	www.freesound.org
WIPERS_1_1	WIPERS	www.freesound.org
WIPERS_2_1	WIPERS	www.freesound.org
WIPERS_2_1	WIPERS	www.freesound.org
ZIPPER	ZIPPER	www.freesound.org
ZIPPER	ZIPPER	www.freesound.org
ZIPPERA	ZIPPER	www.freesound.org
ZIPPERA	ZIPPER	www.freesound.org
ZIPPER_1_1	ZIPPER	www.freesound.org
ZIPPER_1_1	ZIPPER	www.freesound.org
ZIPPER_1_2	ZIPPER	www.freesound.org
ZIPPER_1_2	ZIPPER	www.freesound.org
ZIPPER_2_1	ZIPPER	www.freesound.org
ZIPPER_2_1	ZIPPER	www.freesound.org