UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE
APPLIQUÉES

PAR

LOUIS ROMPRÉ

VERS UNE MÉTHODE DE CLASSIFICATION DE FICHIERS SONORES

NOVEMBRE 2007

Résumé

Ce mémoire présente une généralisation de la classification numérique appliquée à l'audio. L'approche présentée est une combinaison de concepts liés au traitement des langues naturelles et à l'analyse fréquentielle. Les données brutes des documents audio sont transposées en chaîne de caractères alphanumérique propice à l'analyse numérique. Ainsi un traitement analogue à celui voué à la classification de documents textuels peut être appliqué. L'analyse de texte couplée à l'analyse audio donne lieu à la réalisation d'un système capable de classifier de larges corpus hétérogènes.

Vers une méthode de classification de fichiers sonores
Louis Rompré
30024804

TABLE DES MATIÈRES

1. Introduction

Nous sommes présentement dans une ère de création et de consommation d'information. Le volume de données numériques croît constamment et ce à un rythme de plus en plus important. Initialement composés quasi totalement de données textuelles, les réseaux regorgent désormais de données multimédias de toutes sortes. Parmi ces documents, les fichiers sonores sont des plus convoités. En 2004, l'Organisation de Coopération et de Développement Économique (OCDE) estimait à près de 10 millions le nombre d'internautes branchés sur les principaux réseaux point à point[1] à la recherche de fichiers sonores. Même si ces pratiques demeurent contestées, elles démontrent un intérêt croissant pour la recherche de données musicales.

Les données recherchées sont généralement désordonnées et réparties. Une recherche parmi ces millions de documents peut facilement devenir une tâche complexe qui demande un temps considérable. Pour favoriser la recherche et l'obtention des documents souhaités, l'utilisation de diverses techniques est nécessaire. L'indexation est souvent priorisée. Cette technique consiste à définir des descripteurs spécifiques aux documents afin de les représenter et les identifier avec un minimum d'informations. Lors d'une recherche, seuls les index sont consultés ce qui réduit le nombre de données à traiter et par conséquent accélère le processus. La qualité des résultats obtenus est donc fortement liée à la pertinence des descripteurs utilisés comme index. Ceux-ci sont généralement définis en terme de mots-clés. Communément, le nom du document est utilisé. Autrement, l'index d'un document audio correspond à une description saisie par un

[1] Réseaux de partage de données numériques qui ne nécessitent pas de serveur central.

utilisateur. L'augmentation constante du volume de données contraint leur gestion et requiert l'automatisation de certains processus.

L'indexation automatisée de documents sonores pose plusieurs problèmes. Le son est un phénomène continu composé d'une infinité de valeurs. Par conséquent, la notion de mot-clé ne peut être appliquée directement. La diversité des sources sonores et la fiabilité variable de la numérisation sont des facteurs qui complexifient l'automatisation. Les documents doivent subir des prétraitements afin d'être uniformisés. La classification est un prétraitement qui peut améliorer l'indexation. Elle permet de créer une hiérarchie qui facilite la consultation. Elle correspond à une réduction de la complexité de l'environnement par la création de classes de similarités.

La classification repose sur une tâche précise qui est l'identification des caractéristiques communes. Ces caractéristiques peuvent être considérées comme descripteurs génériques à partir desquels des index peuvent être bâtis. La classification inclut un processus fondamental qui est la catégorisation. Ce processus se définit comme l'action d'associer un nouvel élément à une classe en examinant les similarités entre les caractéristiques de cet élément à celles qui définissent les classes existantes. Plusieurs propriétés résultent de cette faculté [Bruner et al, 1956]:

- Réduit la complexité de l'environnement;
- Permet l'identification d'objets et de concepts;
- Diminue le besoin d'apprentissage continuel;
- Crée un ordonnancement et des associations;
- Favorise la prise de décision.

Les propriétés énumérées précédemment justifient le choix de la classification comme prétraitement à prioriser. Les bienfaits de la classification vont au-delà des besoins en matière de recherche de documents. Elle permet de créer une vue d'ensemble qui favorise la connaissance de l'environnement ciblé. Ce mémoire est consacré à l'étude de la classification de documents sonores dans l'objectif d'automatiser cette tâche.

La classification requiert des mécanismes d'analyse et de reconnaissance qui permettent de définir le niveau de ressemblance entre les différents documents. Jusqu'à présent, les travaux dans le domaine du texte sont considérablement en avance face aux domaines de l'image, du son et de la vidéo. L'idée d'élargir la portée de ces travaux devient donc une solution intéressante.

1.1. Approche

L'approche préconisée repose sur la combinaison de concepts liés au traitement des langues naturelles (TALN) et de concepts liés à l'analyse fréquentielle. La classification est réalisée à partir d'une évaluation statistique de séquences particulières contenues dans les documents. Ces unités d'information, nommées n-gramme, sont constituées de n caractères descriptifs consécutifs. L'accent a été porté sur la transformation des données en caractères descriptifs sujets à l'analyse numérique. Le modèle de découpage en n-grammes est traditionnellement associé au traitement automatisé des langues naturelles. Cependant, il possède des propriétés intéressantes pour la recherche de similarités musicales :

- Opère sur des données dont la nature peut varier [Dunning, 1994];
- Tolère un certain ratio de déformation [Miller et al, 1999].

Les propriétés liées au modèle de découpage en n-grammes permettent de tenir compte de la diversité des sons et de la fiabilité variable des enregistrements audio.

Le processus qui mène à la création des n-grammes est subdivisé en quelques étapes. Dans un premier temps, les données brutes sont lues puis segmentées. Un traitement est ensuite appliqué à chacun des segments afin de prélever un aspect prédéterminé. Deux caractéristiques sont considérées : la forme et le contenu fréquentiel du signal. Les renseignements acquis sont quantifiés, traduits en caractères descriptifs puis stockés dans une liste en respectant leur ordre d'apparition dans le signal. Les n-grammes sont formés à partir de la liste générée.

Les n-grammes extraits sont comptabilisés et insérés dans une matrice de manière à représenter leur fréquence d'apparitions à l'intérieur des documents soumis à la classification [Biskri et Meunier, 2002]. Cette représentation vectorielle est présentée à un classifieur externe. Dans le cadre de ce mémoire, le réseau de neurones ART a été utilisé [Carpenter et Grossberg, 2003].

1.2. Domaine d'application

L'intérêt pour un outil de classification automatisée de données sonores s'étend à plusieurs domaines d'application entre autres l'amélioration des moteurs de recherche, la création d'outils d'indexation de documents audiovisuels et la mise en place d'une chaîne de traitements multiformats.

a. Amélioration des moteurs de recherche

L'essor du réseau Internet a largement contribué à l'émergence de nouvelles sources d'information. Les données disponibles se présentent sous plusieurs formats. Dans cet environnement où l'information de masse est généralement peu structurée, l'utilisation de moteurs de recherche est nécessaire pour obtenir efficacement la partie de l'information qui correspond aux critères de recherche établis. Cette recherche doit s'effectuer indépendamment du format des fichiers. Les moteurs de recherche actuels ne semblent pas, à un niveau intéressant, tenir compte de la diversité de formes que peut prendre l'information. Le contenu hétérogène du réseau Internet engendre le besoin de développer des outils de recherche capables de traiter le texte mais également l'image, le son et la vidéo. La classification de documents audio se présente comme une partie de la solution à cette problématique.

b. Création d'outils d'indexation de documents audiovisuels

L'indexation de documents audiovisuels est une tâche complexe. Le temps de calcul nécessaire à l'analyse audio est moindre que celui de la vidéo [Brück et al, 2004]. Certains travaux soutiennent que l'analyse audio d'un document audiovisuel permet une caractérisation qui peut être utilisée comme référence sur le contenu de ce document. La trame sonore est utilisée pour générer une signature plus compacte. L'analyse des fréquences contenues dans le signal sonore est généralement retenue.

c. Mise en place d'une chaîne de traitements multiformats

La portabilité du modèle de découpage en n-grammes ouvre la porte à la mise en place d'une chaîne de traitements unique pour l'analyse de documents multiformats [Bisk■ri et al, 2006]. L'avantage de l'uniformisation réside dans l'accroissement de la capacité

d'adaptation des systèmes face à l'évolutivité des nouveaux médias. Seule la préparation des données diffère selon le format.

1.3. Organisation du mémoire

Le présent chapitre est voué à une brève introduction dans laquelle sont donnés les motifs qui ont guidé la mise en œuvre de ce projet. Les lignes directrices de cet ouvrage sont également abordées.

Le chapitre 2 est consacré à la description des processus de classification et de catégorisation de documents numériques. Au cours de ce chapitre, les concepts clés sont abordés avec l'objectif de mettre en évidence l'influence de l'unité d'information dans ces processus. La notion de n-gramme est présentée et suggérée comme alternative intéressante.

Le chapitre 3 est subdivisé en deux volets : le premier expose les différents aspects de la musique tandis que le second est dédié à un survol de l'état de l'art en matière de classification de documents sonores. L'étude du signal acoustique est nécessaire à l'identification des caractères descriptifs susceptibles d'engendrer les unités d'information les mieux adaptées à la classification de documents sonores. Une analyse critique des approches existantes conclut ce chapitre.

Le chapitre 4 illustre l'architecture du système développé pour remédier aux lacunes décrites lors de l'analyse critique effectuée au chapitre 3. Le rôle des différentes composantes du système et les raisons qui ont motivé leur conception sont également expliqués.

Les résultats des expérimentations effectuées sont présentés au chapitre 5. Ces expérimentations ont été réalisées de manière à valider le système développé. Des analyses qualitatives et subjectives des classes de similarité obtenues sont également abordées.

La conclusion est présentée au chapitre 6. Une évaluation globale des travaux effectués est dressée.

2. La classification numérique

Une classification est une distribution ordonnée au sein de classes de similarité. L'objectif est d'organiser et de hiérarchiser l'information afin de favoriser la recherche, la consultation mais également de dresser un portrait de l'ensemble des données. Le partitionnement des classes sur un ensemble de documents est couramment exprimé par [Bouveyron, 2006]:

> L'ensemble $P = \{C_1, C_2, ..., C_k\}$ est une partition de l'ensemble E en k classes si et seulement si :
>
> $$\begin{aligned} &i. \quad C_i \neq \emptyset \text{ pour } i = 1, \ldots, k \\ &ii. \quad \bigcup_{i=1}^{k} C_i = E \\ &iii. \quad C_i \cap C_l = \emptyset \text{ pour tout } i \neq l \end{aligned} \qquad \textbf{(2.1)}$$

Certains modèles omettent les contraintes (ii) et (iii) de manière à, d'une part, exclure les documents jugés impropres à la classification et, d'autre part, autoriser l'appartenance à plus d'une classe.

La pertinence d'une classification est habituellement jugée en fonction de l'homogénéité des classes qui en résulte. La notion d'homogénéité est mesurée de manière différente d'un individu à l'autre et par conséquent ce critère d'évaluation demeure relatif. L'examen d'une classe par un intervenant est accompli à partir de ses connaissances du domaine abordé. Cette subjectivité complexifie l'évaluation lorsque celle-ci est effectuée par un processus automatisé. La problématique majeure est de déterminer les frontières des classes. Un processus automatisé peut générer des associations constantes qui permettent de faire ressortir des caractéristiques communes inattendues. Il est donc essentiel de considérer le contexte avant de mesurer la valeur de la qualité des résultats obtenus.

2.1. L'automatisation

Les systèmes de classification automatisés les plus récents reposent sur des modèles probabilistes qui formalisent la définition d'une classe. La catégorisation est réalisée à l'aide de règles de décision bâties à partir de données statistiques puisées à même les documents. Le défi consiste à formuler une règle qui associe à une classe spécifique un document lorsque celui-ci répond à un seuil de similitude prédéfini. La recherche d'affinités forme l'essentiel de cette tâche. Les parties semblables sont identifiées et regroupées en classes de similarité. Lorsque le degré de ressemblance n'est pas atteint et ce pour la totalité des classes existantes, une nouvelle classe est créée. On distingue deux étapes importantes lors de la catégorisation [Harb, 2003]:

i. L'extraction des caractéristiques;

ii. L'application d'une méthode de classification ou la sélection d'un classifieur qui assigne une probabilité d'appartenance.

Explicitement le processus de classification automatisée consiste à :

- Déterminer une unité d'information à considérer;
- Prendre en entrée des documents ou des portions de document;
- Comptabiliser les unités d'informations;
- Créer un vecteur de cooccurrence pour chacun des documents;
- Ordonner les vecteurs de manière à retourner en sortie des classes de similarités.

La figure 2.1 illustre ce mécanisme. L'ensemble des documents forme le domaine d'information tandis que l'union des différentes unités répertoriées détermine l'ensemble des unités d'information.

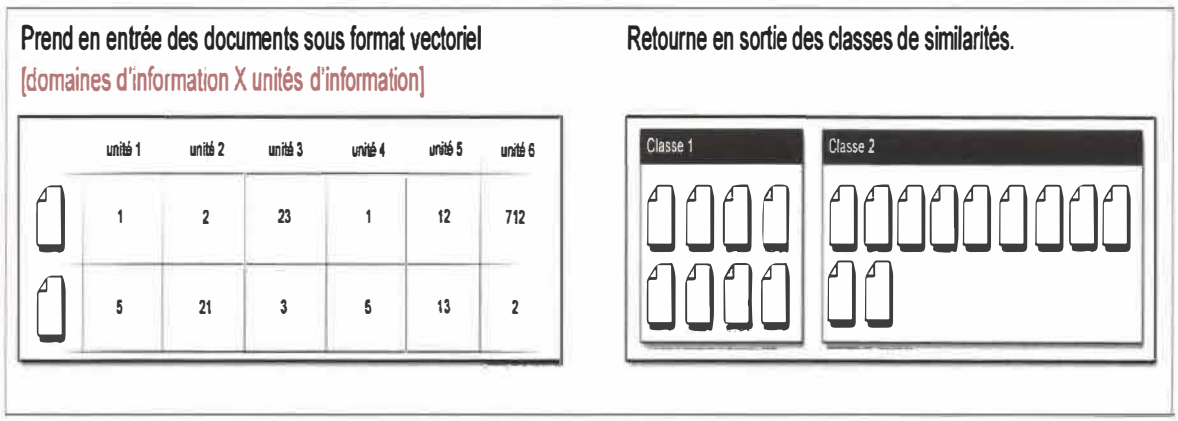


Figure 2.1 Processus de classification illustré.

En simplifiant, on peut dire que des documents sont jugés similaires lorsqu'ils sont constitués des mêmes unités d'information à des fréquences presque identiques. Par conséquent, le choix de l'unité d'information a un impact direct sur les résultats de la classification.

2.2. L'unité d'information

Dans une perspective d'une classification automatisée de fichiers sonores une question se pose :

Quelle est donc l'unité d'information la plus adéquate pour représenter une onde sonore?

La définition de l'unité d'information joue un rôle décisif dans la mise en œuvre de la classification. Une unité d'information doit satisfaire quatre contraintes pour assurer un niveau d'adaptabilité intéressant [Biskri et al, 2002] :

i. Elle doit être répartie dans le domaine d'information;

ii. Il doit être simple de la repérer sur le plan informatique;

iii. Elle doit être statistiquement comparable de sorte qu'il soit possible de calculer sa fréquence d'apparitions, d'estimer sa distribution et sa régularité;

iv. Sa nature doit dépendre des objectifs du traitement. La définition de l'unité d'information dépend de l'objectif de lecture et de compréhension.

La contrainte (i) spécifie que l'unité d'information doit être une partie intégrante du contenu, de manière à représenter une caractéristique réelle et non une interprétation du document. L'exigence (ii) relève du domaine algorithmique et vise à minimiser le coût computationnel. Les systèmes de classification automatisés reposent sur des modèles probabilistes. Cette particularité architecturale engendre l'énoncé (iii). La dernière contrainte énumérée permet l'analyse de documents de formats différents impliquant des modifications minimes. Cette orientation permet la couverture d'un large éventail de formats audio moyennant un effort restreint.

Dans le cadre d'une analyse numérique en fonction d'une classification, que l'unité d'information soit comprise hors de son contexte n'est pas une contrainte en soit [Biskri et al, 2002]. Il suffit que l'unité d'information soit suffisamment discriminante pour faire ressortir des distinctions et des associations entre les documents.

2.3. Les n-grammes comme unité d'information

L'unité d'information est généralement liée au domaine auquel elle est attachée. Cette singularité engendre l'utilisation de plusieurs variantes afin de couvrir un environnement complexe. L'utilisation de dictionnaires devient nécessaire et réduit la capacité d'adaptation face au changement. Or, les progrès réalisés dans le domaine de la compression numérique entraînent l'apparition de nouveaux formats de fichiers. Ce contexte est peu favorable à l'utilisation d'une unité d'information peu flexible.

Un n-gramme est une séquence de n caractères descriptifs consécutifs. Le caractère qui le compose correspond à l'élément atomique le plus descriptif contenu dans le document traité. Cette structure a été instaurée en 1951 par l'article « Prediction and Entropy of Printed English » [Shannon, 1951]. L'idée sous-jacente repose sur deux concepts : l'entropie et la redondance des langages. L'entropie est un paramètre statistique qui

mesure la contribution d'un caractère dans une chaîne de caractères. Elle permet d'estimer le niveau d'incertitude de la nature d'une donnée à partir des données qui la précèdent. L'entropie mesure le désordre. L'entropie est nulle lorsqu'il y a absence d'incertitude. La redondance, quant à elle, quantifie les contraintes imposées par la structure d'un langage. À titre d'exemple, dans la langue anglaise, les probabilités que la lettre « h » succède à la lettre « t » (the, width, thanks etc.) ou que la lettre « u » succède à la lettre « q » (quote, question, equal etc.) sont élevées. L'absence ou la présence de combinaisons particulières permet donc de déduire certaines caractéristiques de la source de données.

La notion de n-gramme suscite un intérêt grandissant depuis le milieu des années 90, principalement dans le domaine du traitement automatisé des langues naturelles. Des recherches sur l'identification de la langue [Greffenstette, 1995] et sur le calcul de similarités [Damashek, 1995] ont démontré que même si les n-grammes ne représentent pas une entité concrète tel que le mot pour le domaine du texte, ils n'entraînent pas de perte d'information.

À l'instar des autres techniques d'analyse numérique fondées sur la comparaison de chaînes de caractères, le modèle de découpage en n-grammes permet d'exercer un contrôle sur la taille du lexique et de le maintenir à un seuil raisonnable. Un lexique obtenu suite à ce découpage ne peut surpasser la taille de l'alphabet à la puissance 2.

2.4. La classification à partir de n-grammes

La classification de documents à partir du modèle de découpage en n-grammes est réalisée en six étapes [Damashek, 1995] :

i. Définir les préconditions qui uniformisent ou compressent les données;

ii. Extraire les n-grammes des différents fichiers;

iii. Assigner une clé d'identification aux n-grammes;

iv. Calculer la répartition et créer une table de hachage pour chacun des fichiers;

v. Normaliser les tables de hachage en fonction de la taille des fichiers;

vi. Comparer les tables de hachages.

La définition de préconditions a comme objectif de réduire le nombre de combinaisons possibles. Certains caractères sont peu descriptifs, par conséquent, les éliminer peut se traduire par un gain de performance.

Extraire les n-grammes d'un document consiste à glisser une fenêtre de taille n sur les données constituant ce document. Ce déplacement est effectué un caractère à la fois. La figure 2.2 illustre le procédé pour le domaine du texte.

Données: « *Exemple du procédé* »

	Tri-gramme
Exemple du procédé	Exe
E**xem**ple du procédé	xem
Ex**emp**le du procédé	emp
...	...
Exemple du proc**édé**	édé

Figure 2.2 : Extraction de n-grammes dans le domaine textuel où $n = 3$.

L'assignation d'une clé d'identification revient à lister les différents n-grammes contenus dans le domaine d'information. Un n-gramme doit être représenté par la même clé pour

l'ensemble des documents. Cependant, une clé identique peut être attribuée à deux n-grammes jugés similaires. Selon Damashek, cette collision n'affecte que très peu les résultats obtenus.

Le calcul de la répartition des n-grammes permet de générer une représentation vectorielle propice à la majorité des classifieurs. La table de hachage proposée peut être remplacée par tout autre structure de données équivalente.

Lorsque tous les n-grammes sont extraits d'un document, le nombre d'occurrences d'un n-gramme est divisé par le nombre total de n-grammes extraits pour ce document. Ce qui signifie que le nombre absolu d'occurrences est remplacé par la fréquence relative d'apparitions. La motivation derrière cette opération est de pouvoir comparer des fichiers de taille variable de manière équitable.

Il existe plusieurs modes de comparaison des résultats obtenus. Plusieurs travaux de recherche sont dédiés à cette tâche [Hsu et Lin, 2002] [Carpenter et al 1991]. L'utilisation de classifieur existant permet donc de tirer profit de l'expertise établie.

2.5. Sommaire du chapitre 2

La classification automatisée œuvre à partir de données statistiques sur les unités d'information dont les documents sont formés. Dans ces systèmes, l'unité d'information occupe un rôle substantiel. Le choix de l'unité d'information considérée a un impact direct sur les résultats obtenus. La richesse du son et la qualité variable de la numérisation impliquent une unité d'information flexible pouvant s'adapter aux variations. Le modèle de découpage répond à ces critères. Même si les n-grammes ne sont pas des entités concrètes ils ne causent pas de perte d'information. Au contraire, ils peuvent rehausser certaines associations. Par exemple, pour le domaine du texte, les mots « ANIMAL » et

« ANIMALERIE » ont un degré de similarité élevé car l'ensemble des bi-grammes $P_1 = \{AN, NI, IM, MA, AL\}$ du premier mot est un sous ensemble de l'ensemble des bi-grammes $P_2 = \{AN, NI, IM, MA, AL, LE, ER, RI, IE\}$ du deuxième mot.

Les modèles de classification basés sur la notion de n-gramme recherchent les séquences discriminantes au sein du contenu des documents à classifier. Ce concept statistique axé sur l'occurrence de combinaisons particulières peut s'étendre au-delà du domaine du texte. Bien que généralement utilisé pour le traitement de corpus textuel, ce découpage représente une avenue prometteuse pour la recherche de similarités musicales.

L'intérêt réside dans la portabilité de ce mode de découpage. La possibilité d'appliquer la même chaîne de traitements à des documents de formats différents laisse entrevoir l'émergence d'outils de classification capables de suivre l'évolution résultant de l'intérêt croissant pour les documents multimédias.

3. Revue de la littérature

Au cours de ce chapitre, les nombreux aspects de la musique seront abordés afin de clarifier la nature du contenu des fichiers sonores. Par la suite, un survol de l'état de l'art sera dressé afin de faire valoir les solutions proposées et les manques en matière de recherche de similarités et de classification automatisée de fichiers sonores.

3.1. Les différentes facettes de la musique

La musique est un art qui agence sons et silences. Ce n'est pas un élément tangible mais plutôt un phénomène physique perceptible. Derrière cet art, il existe une théorie exprimée en termes de durée, de hauteur, d'intensité, de timbre, etc.

a. Le son

Le son est produit par des variations de pressions diffusées dans l'air. Il peut être pur (figure 3.1) ou complexe (figure 3.2).

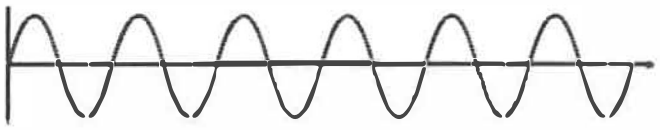


Figure 3.1 : Son pur.

Un son pur est périodique, c'est-à-dire, qu'il possède une forme sinusoïdale qui se répète selon un intervalle fixe nommé période. La vitesse de cette répétition est appelée fréquence et est mesurée en hertz (Hz). La fréquence est la réciproque de la période. Cette relation est donnée par la formule 3.1 où la variable P représente la période donnée en seconde et la variable F la fréquence exprimée en hertz [Thomas L. Floyd, 2001].

$$P = \frac{1}{F} \quad , \quad F = \frac{1}{P} \tag{3.1}$$

Un son complexe (figure 3.2) est composé de plusieurs sinusoïdes de fréquences et d'amplitudes différentes.

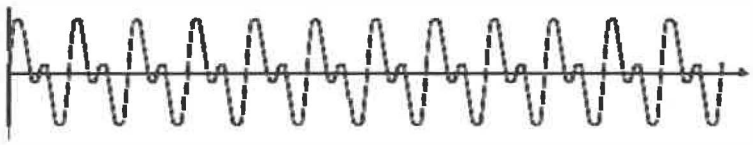


Figure 3.2: Son complexe.

b. La durée

La durée est le laps de temps pendant lequel un son est perçu. Le tempo est un indicateur de la durée. Plus le tempo est rapide plus la durée d'une note est courte.

c. La hauteur

La hauteur correspond à la fréquence du signal acoustique. Une fréquence rapide génère un son aigu (haut), tandis qu'une fréquence lente génère un son grave (bas). L'oreille humaine perçoit les sons dont la fréquence est comprise entre 20 et 20 000 hertz mais cet intervalle décroît avec l'âge.

La parole se limite à un intervalle d'environ 8 000 hertz.

d. Les basses fréquences

Les basses fréquences se situent entre 20 et 200 hertz.

e. Les hautes fréquences

Les hautes fréquences se situent entre 2 000 et 20 000 hertz.

f. La fréquence du diapason

La fréquence du diapason est une référence absolue. Sa valeur a changé à travers le temps. Actuellement, la fréquence du diapason est fixée à 440 hertz.

g. L'intervalle

La différence de hauteur entre deux notes se nomme intervalle. Un intervalle peut être ascendant (deuxième son plus aigu que le premier) ou descendant (deuxième son plus grave que le premier).

h. Les notes

Les notes de musique sont des indicateurs conventionnels de la hauteur et de la durée d'un son. Lorsqu'inscrite sur la portée[2], la position de la note indique la hauteur du son tandis que la forme indique sa durée relative.

Il existe sept notes fondamentales : *do*, *ré*, *mi*, *fa*, *sol*, *la*, *si*.

i. L'octave

On nomme octave l'intervalle qui sépare deux sons dont les fréquences fondamentales ont un rapport de fréquence égal à 2 (880, 440, 220, 110, 55 etc.). Ces sons sont perçus par l'oreille de façon comparable et sont identifiés par la même note.

j. Le ton

Le ton est un rapport entre les hauteurs de deux notes conjointes.

k. La gamme

Une gamme est une succession de notes à l'échelle d'une octave. La gamme majeure illustrée à la figure 3.3 est la gamme de référence. La majorité des autres gammes découlent de celle-ci.

Les notes d'une gamme ne sont pas nécessairement toutes séparées par le même intervalle.

---

[2] Série de cinq lignes horizontales, équidistantes et parallèles, utilisée pour noter la musique. Définition tirée du dictionnaire Petit Larousse Illustré 2005.

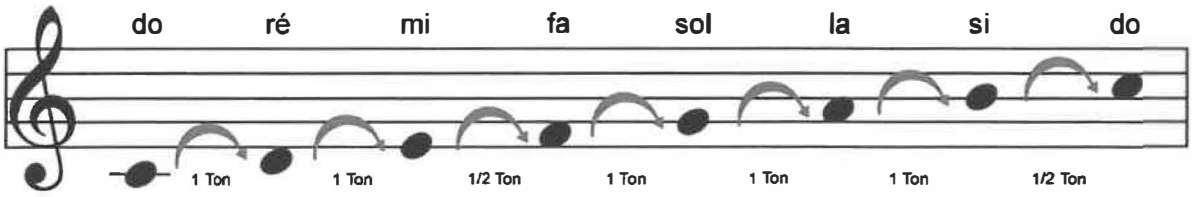


Figure 3.3 : Gamme majeure.

l. L'intensité

L'intensité est l'amplitude de l'onde sonore. Elle est à l'origine de la puissance du son. Une amplitude importante provoque un son fort et, inversement, une amplitude faible engendre un son moindre. La puissance est mesurée en décibels (dB). Le niveau sonore minimum que l'oreille humaine peut percevoir est de 1 décibel (seuil d'audibilité) tandis que le seuil maximum est autour de 120 décibels (seuil de douleur). Au delà du seuil de douleur, le son peut occasionner des dommages irréversibles à l'ouïe. La définition du décibel est donnée par la formule suivante :

$$dB = 10\log_{10}\left(\frac{I_1}{I_0}\right) \qquad \textbf{(3.2)}$$

La variable I_1 représente l'intensité mesurée et I_0 l'intensité de référence qui correspond au seuil d'audibilité.

m. Le timbre

Tout corps vibre de façon unique. Par conséquent, chaque source sonore texture de manière propre le son qu'elle émet. Cette distinction est le timbre. Il nous permet de différencier les différents instruments de musique même lorsque ceux-ci jouent la même mélodie simultanément. Le timbre est le résultat de l'ajout d'harmoniques dans le signal

sonore. Les figures 3.4 et 3.5 illustrent l'onde sonore de la note mi jouée respectivement à la guitare et à la base.

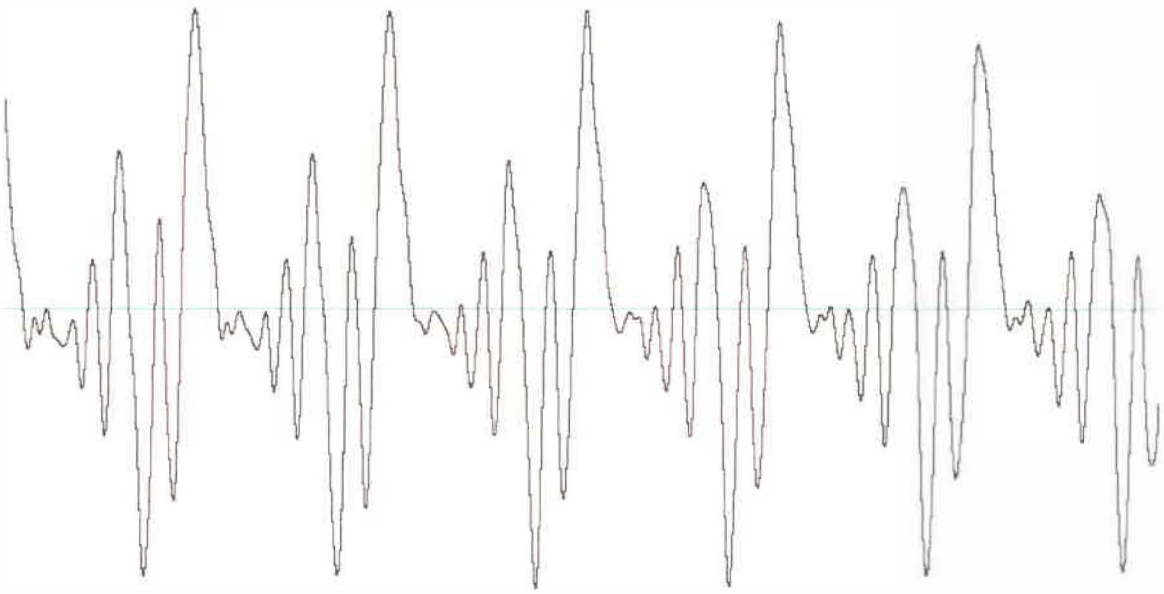


Figure 3.4 : L'onde sonore de la note mi jouée à la guitare.

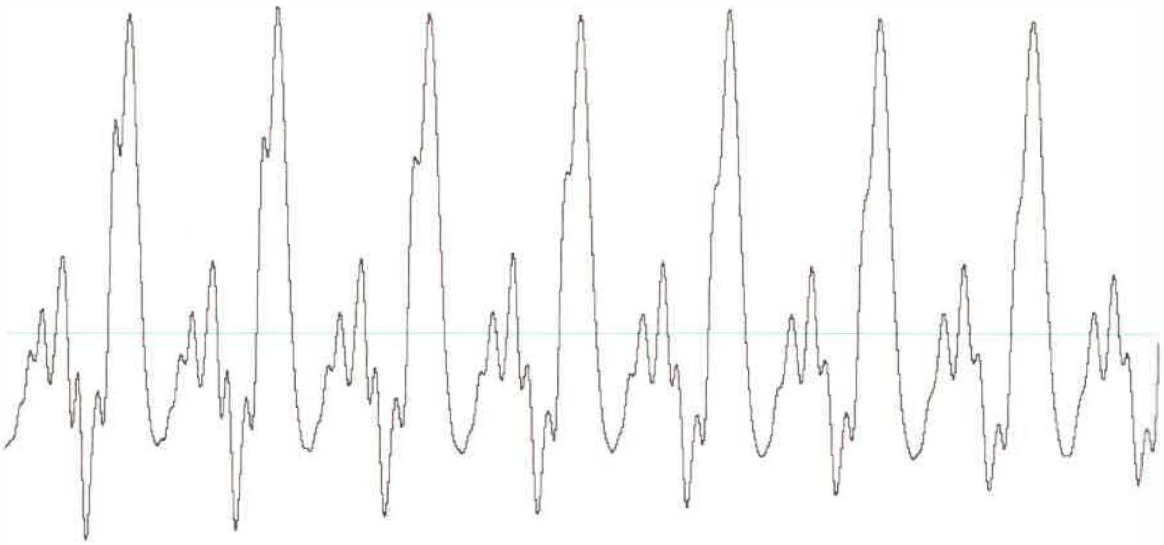


Figure 3.5 : L'onde sonore de la note mi jouée à la basse.

n. Le facteur humain de la musique

Le facteur humain est le plus complexe et le plus subjectif. Cette facette peut être fractionnée en cinq aspects :

i. L'époque;

ii. La géographie;

iii. La bibliographie;

iv. Le message;

v. La sémantique.

Chaque époque est tributaire d'une relation entre l'art et la société. Les mœurs et le contexte social façonnent l'art. Il est facile d'associer un courant musical à chaque époque de notre ère. À titre d'exemple, le rock'n roll des années 60 ou le disco des années 70. Ce phénomène est souvent lié à une innovation technologique telle l'apparition de l'amplification. Plusieurs tendances coexistent durant la même période et certaines d'entre elles chevauchent plus d'une décennie. La musique joue un rôle culturel important. Il n'existe pas de société sans musique. La musique façonne l'histoire des peuples du monde entier. Son usage est religieux (chants religieux), politique (hymnes nationaux) et folklorique (musiques traditionnelles). L'artiste et la musique à qui il donne vie sont habituellement indissociables. Même si certains artistes couvrent plusieurs genres musicaux, ils parviennent à se distinguer par le son, l'arrangement ou la structure qu'ils proposent. Le texte, le message d'une chanson lui procurent une valeur littéraire. La musique possède également une sémantique dont l'interprétation varie selon l'auditeur. Elle peut rappeler un souvenir ou un produit de consommation par exemple.

o. Les genres musicaux

Les genres musicaux sont une catégorisation de la musique. Ils résultent de pratiques musicales de même nature. Ils servent de référence pour les disquaires et les mélomanes. Il est possible de lister plus de 400 genres musicaux. La majorité de ces genres sont subdivisés en sous-catégories qui elles-mêmes sont subdivisées. L'association entre une musique et un genre musical ne fait pas toujours l'unanimité, par conséquent, il devient parfois difficile d'établir spécifiquement le genre d'un artiste.

Il n'existe pas de liste exhaustive des genres musicaux.

3.2. La conversion d'un signal analogique à numérique

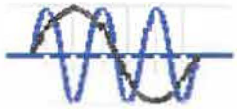
L'onde sonore est un signal analogique. Numériser un son consiste à prélever des points d'observations sur l'onde. L'échantillonnage est effectué à un intervalle de temps régulier nommé fréquence d'échantillonnage. Cette fréquence doit être suffisamment élevée afin de préserver la forme de l'onde sonore et ainsi reproduire le son le plus fidèlement possible. Selon le théorème de Shannon [Shannon, 1948], la fréquence d'échantillonnage doit être au moins égale au double de la plus haute fréquence contenue dans le signal analogique. En deçà de cette limite théorique, il n'est pas possible de reconstituer un signal à partir d'échantillons sans perte d'information.

Le théorème de Shannon est exprimé par la formule 3.3 où F_e représente la fréquence d'échantillonnage et $2F$ la fréquence maximale du signal analogique à numériser.

$$F_e \geq 2F \qquad \textbf{(3.3)}$$

Le processus de numérisation est démontré à la figure 3.6 où l'onde à numériser, la fréquence d'échantillonnage et les échantillons sont identifiés respectivement par (a), (b) et (c).

(a)



(b)

(c)

Figure 3.6 : Numérisation d'une onde analogique.

Le taux d'échantillonnage est exprimé en hertz et représente le nombre d'échantillons par seconde. L'oreille humaine perçoit les sons situés majoritairement entre 20 hertz et 20 000 hertz. Une fréquence d'échantillonnage de 40 000 hertz est donc nécessaire pour obtenir un enregistrement de qualité. Pour les enregistrements audio, la norme est fixée à 44 100 hertz. Généralement, pour quantifier les échantillons, 8 ou 16 bits sont utilisés. La fréquence d'échantillonnage et la profondeur de la quantification ont une influence directe sur le volume des données numérisées ainsi que sur la fidélité de la numérisation.

Les fichiers sonores sont volumineux et complexes à analyser.

3.3. État de l'art

La section suivante est dédiée à la revue de la littérature

3.3.1. La recherche d'information (*Information Retrieval*)

L'informatisation engendre l'apparition de nouvelles méthodes de recherche. La recherche de similitudes entre différents documents est un domaine en pleine effervescence. La surconsommation d'information, l'importance de celle-ci dans plusieurs champs d'activités de même que sa nature évolutive créent des attentes élevées face aux outils de recherche.

3.3.2. Le système SMART

SMART est un système de recherche d'informations développé dans les années soixante par une équipe de chercheurs de l'université Cornell.

La robustesse de trois concepts est démontrée:

i. La sémantique vectorielle (*vector space model*);

ii. La rétroaction (*revelance feedback*);

iii. L'heuristique TF*IDF (*Term Frequency by Inverse Document Frequency*).

SMART est un système précurseur. Les concepts qui y ont été appliqués ont été réutilisés par de nombreux modèles de classification.

a. La sémantique vectorielle

La sémantique vectorielle est un modèle algébrique qui représente un document de manière formelle par l'entremise de vecteurs dans un espace linéaire multidimensionnel. Les vecteurs sont composés de mots-clés [G. Salton et al, 1975]. La sélection des termes utilisés comme mots-clés est l'opération la plus complexe. L'extraction automatique de ces descripteurs est sujette à l'exploration.

b. La rétroaction

La rétroaction consiste à tenir compte de la pertinence des résultats obtenus lors d'une recherche afin d'améliorer les résultats de recherches futures. Le contenu des documents jugés pertinents est utilisé pour pondérer chacun des termes qui composent une requête. Trois catégories de rétroaction sont définies :

i. La rétroaction explicite;

ii. La rétroaction implicite. Les documents consultés sont jugés plus pertinents que les documents non consultés;

iii. La pseudo rétroaction.

La rétroaction explicite implique directement l'acteur de la recherche. Celui-ci attribue un degré de pertinence aux documents qu'il consulte. Cette évaluation affecte un seuil de désirabilité considéré lors d'analyses ultérieures.

L'utilisateur peut également contribuer indirectement (ii). Les documents consultés sont jugés plus pertinents que ceux non consultés. Cette rétroaction implicite génère un taux d'insatisfaction plus élevé.

La pseudo rétroaction revient à appliquer un filtre sur les résultats dévoilés. Les n premiers documents concordants sont jugés pertinents.

c. TF*IDF (*Term Frequency by Inverse Document Frequency*)

Introduit en 1972 [Spärck Jones, 1972], l'heuristique détermine le poids d'un terme en fonction de sa fréquence d'apparitions. Un terme contenu dans un nombre élevé de documents est moins discriminant qu'un terme contenu dans peu de documents (*Inverse Document Frequency* ou *IDF*). La forme communément utilisée pour quantifier cette intuition est donnée par l'équation 3.4 [Robertson, 2004] :

$$idf(t_i) = \log \frac{N}{n_i} \qquad \textbf{(3.4)}$$

La variable N représente le nombre total de documents contenus dans le corpus. Les variables t_i et n_i représentent respectivement le terme mesuré et le nombre de documents contenant ce terme.

La fréquence d'apparitions à l'intérieur même du document joue un tout autre rôle. Plus un terme est répété à l'intérieur d'un document, plus ce terme est représentatif de l'ensemble du document (*Term Frequency* ou *TF*).

Le couplage de ces deux principes noté TF*IDF constitue un heuristique qui permet de créer des associations cohérentes entre des documents.

3.3.3. La recherche de similarité musicale (*Music Information Retrieval*)

La recherche de similarités musicales est un domaine sous-jacent de la recherche de similarités. Les travaux qui y sont dédiés sont subdivisés en deux groupes [Downie, 2003]:

i. Les modèles d'analyse;

ii. Les modèles de localisation.

Les modèles d'analyse sont conçus dans l'objectif de fournir une représentation qui soit la plus complète possible. Les concepteurs de tels systèmes cherchent à acquérir le détail de tous les aspects théoriques de la musique au détriment de l'aspect humain. L'utilisation de ces systèmes nécessite un haut niveau d'expertise. Les utilisateurs des modèles d'analyse sont généralement des musicologues ou des compositeurs.

Contrairement aux modèles d'analyse, les modèles de localisation ne nécessitent pas d'expertise musicale. Ils sont conçus pour assister l'identification et la localisation des similarités [Downie, 1999]. Il existe principalement trois approches:

i. Le traitement du signal;

ii. La fouille de données contextuelles;

iii. Le filtrage collaboratif.

Le traitement du signal s'intéresse aux propriétés du signal acoustique. La recherche de descripteurs de bas niveau est réalisée à l'aide d'outils mathématiques telle la transformée

de Fourier. Le traitement du signal s'apparente au modèle d'analyse mais se limite à quelques aspects comme le contenu fréquentiel.

La fouille de données contextuelles repose sur les données textuelles qui décrivent la musique. L'efficacité de cette approche dépend de la qualité des mots-clés utilisés. Généralement, ils réfèrent au nom de l'artiste, au titre de la pièce, au genre etc.

La rétroaction forme l'essentiel du filtrage collaboratif. Cette approche consiste à déterminer des similarités à partir de pondérations recueillies auprès de différents individus. Cette technique est largement utilisée par les systèmes de recommandations [Lucille Tanquerel et Luigi Lancien, 2005].

Seules les approches basées sur le traitement du signal s'intéressent au contenu réel des fichiers et ne dépendent pas exclusivement de l'interprétation humaine. Le choix du signal sonore comme support à l'analyse s'impose comme méthode d'analyse par sa relation intrinsèque avec les données des fichiers.

3.3.4. L'approche de J. Stephen Downie

Downie est l'un des premiers à proposer l'utilisation d'approches traditionnellement associées au traitement automatisé des langues naturelles. En ce sens, son nom est souvent évoqué dans la littérature traitant de la recherche de similarités musicales. La méthode proposée consiste à fragmenter une mélodie en sous segments comparés à des « mot musicaux ».

Deux paradigmes sont abordés [Downie, 1999] :

i. Indexer les données à partir d'un modèle « full-text »;

ii. Préconiser le principe de parcimonie[3].

a. L'indexation « *full-text* »

Les documents sont généralement indexés à l'aide de deux types de descripteur : les descripteurs qui définissent le « à propos » et ceux qui définissent le « comment ». La première catégorie réfère à la facette bibliographique tandis que la seconde diffère selon l'intention et est généralement perçue comme un texte libre. L'index est indépendant de la représentation du contenu. Par exemple, le titre de la pièce, le genre, le nom l'auteur et le nom de l'interprète sont utilisés comme descripteurs pour la quasi-totalité des corpus musicaux. La résultante est que pour la majorité des outils de recherche, les index sont constitués de chaînes alphanumériques. Par conséquent, Downie conclut que l'indexation de documents audio doit être représentée sous forme de chaînes alphanumériques structurées de manière analogue au mot. Les « mots musicaux » doivent être bâtis à partir

[3] Avec une mesure extrême, en s'en tenant au strict minimum. Définition tirée du dictionnaire Petit Larousse Illustré 2005.

du contenu des documents afin de le résumer. Ils ont comme mandat de décrire la musique de la même manière et aussi précisément que le font les mots dans le domaine textuel.

b. La parcimonie

L'application du principe de parcimonie repose sur la notion de coût versus bénéfice. Un modèle simple est préféré à un modèle complexe s'il peut fournir des résultats jugés satisfaisants. La mise en application de ce principe implique la simplification de la représentation des données.

Pour simplifier les données, trois règles majeures sont adoptées:

i. Traiter des mélodies monophoniques uniquement;

ii. S'intéresser uniquement à l'intervalle;

iii. Utiliser les n-grammes comme « mots musicaux ».

c. Le projet MusicFind

Le projet MusicFind est un système de catégorisation automatisée de fichiers sonores de format MIDI (*Musical Instrument Digital Interface*). Le format MIDI est destiné aux instruments de musique numérique tel le synthétiseur. La particularité de ce format est qu'il contient une description détaillée de la mélodie. Ainsi, un fichier MIDI fournit explicitement le ton, l'intensité, et la durée des notes.

L'hypothèse soutenue est qu'une simple représentation, basée sur les intervalles d'une mélodie, contient assez d'information pour réaliser des opérations de catégorisation. L'utilisation de méthodes d'évaluation peut contrebalancer la perte d'information et de précision résultant de la vulgarisation de données initiales. Il existe une certaine équivalence entre un n-gramme composé d'intervalles dans le domaine de la musique et un n-gramme composé de lettres dans le domaine du texte. Ces mots informent sur la nature de l'ensemble des données. Cependant, le lexique généré à partir d'intervalles est susceptible d'atteindre une taille considérable. Afin de réduire le nombre de combinaisons possibles, les intervalles sont groupés en classes d'équivalences.

Un ensemble d'intervalles I est partitionné en sous ensembles disjoints $I_1, I_2, I_3, ..., I_n$ tel que :

$$\bigcup_{j=1}^{n} I_j = I, \; I_i \cap I_j \neq 0 \rightarrow i = j \qquad \textbf{(3.5)}$$

Chaque I_i est appelé une équivalence ou une classe d'intervalles de l'intervalle I. À partir d'une liste ordonnée $p_1, p_2, p_3, ..., p_{n+1}$ représentant les notes d'une mélodie, les intervalles $i_1, i_2, i_3, ..., i_{n-1}$ sont créés de telle sorte que i_j équivaut à l'intervalle entre p_j et p_{j+1} où $j = [1...n-1]$. La figure 3.5 illustre les onze intervalles créés à partir d'une mélodie formée de douze notes.

On s'intéresse alors à la différence de hauteur entre p_{j+1} et p_j. Par exemple, l'intervalle I_1 est constant tandis que l'intervalle I_2 est descendant.

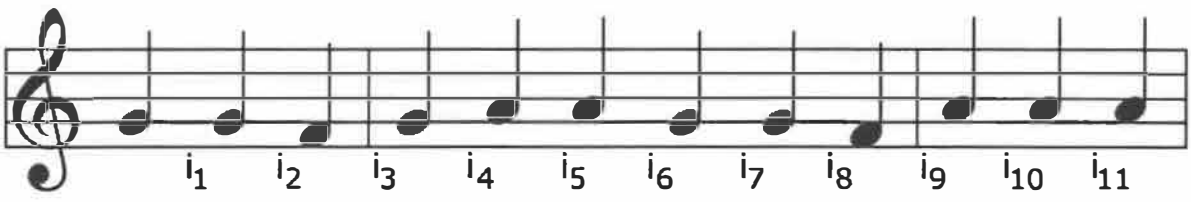


Figure 3.7 : Mélodie composée de 12 notes.

Une méthode de classification est appliquée aux intervalles de manière à déterminer leur appartenance à l'une ou l'autre des classes d'équivalence.

La méthode de classification *C*3 consiste à noter « a » les intervalles continus, à noter « b » les intervalles descendants et à noter « B » les intervalles ascendants. Le tableau 3.1 contient cette conversion pour les intervalles de la mélodie de la figure 3.7.

	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}
	→	↓	↑	↑	→	↓	→	↓	↑	→	→
C3	a	b	B	B	a	b	a	b	B	a	a

C3 : *Parson's Repeat, Up, Down*

Tableau 3.1: Interprétation de la mélodie de la figure 3.7.

Le lexique généré est ensuite segmenté en n-grammes. Le tableau 3.2 démontre le résultat de cette opération pour *n* variant de 4 à 6. Ainsi, les quadri-grammes « abBB » et « bBBa » représentent respectivement les intervalles 1 à 4 et 2 à 5.

4-gramme	5-gramme	6-gramme
abBB	abBBa	abBBab
bBBa	bBBab	bBBaba
BBab	BBaba	BBabab
Baba	Babab	BababB
abab	ababB	ababBa
babB	babBa	babBaa
abBa	abBaa	-
bBaa	-	-

Tableau 3.2 : Segmentation en n-grammes.

L'évaluation comparative des documents est effectuée à l'aide de l'algorithme TF*IDF [Spärck Jones, 1972] tandis que la pondération est réalisée à l'aide d'une métrique de normalisation de la précision nommée NPREC.

$$NPREC = 1 - \frac{\sum_{m=1}^{REL} \log \frac{Rank_m}{m}}{\log N!/(n - REL)!REL!} \qquad \textbf{(3.6)}$$

La normalisation de la précision est donnée par la formule 3.6 où N est le nombre de documents dans la base de données, REL le nombre de documents pertinents et $Rank_m$ le rang assigné au document m [Salton et al, 1983].

Downie s'est intéressé à l'impact de certains paramètres telle la taille des n-grammes. Il arrive à la conclusion qu'il n'existe pas de combinaison unique qui soit applicable à toutes les formes de recherche musicale.

Le technique de recherche de similarités proposée suscite un intérêt parce qu'elle propose l'utilisation d'une technique existante dont l'efficacité a déjà été démontrée pour le domaine du texte. Cependant, le son polyphonique n'est pas considéré, ce qui réduit la portée de la méthode.

3.3.5. L'approche de Nikunj Patel et Padma Mundur

Nikunj Patel et Padma Mundur [Nikunj Patel et Padma Mundur, 2004] suggèrent une technique de recherche de similarités musicales qui s'apparente à celle de Downie [Downie, 1999]. Leur recherche porte sur la détection de répétitions qui caractérisent une musique. Cependant, contrairement à Downie, ils s'intéressent au son polyphonique.

a. La répétition

Les fichiers musicaux sont volumineux. Par conséquent, pour optimiser les systèmes de recherche de données musicales, il est impératif que seuls les segments essentiels soient indexés. Ceci permet de réduire considérablement la taille des indexes et ainsi diminuer le temps nécessaire à la recherche. La répétition est définie de plusieurs manières dans la musique : une mélodie peut se répéter à plus d'une reprise sans changement, une mélodie peut être jouée à différentes octaves (ce qui équivaut à une transposition fréquentielle) et finalement une mélodie peut être une composante ou une partie d'une mélodie plus large. La répétition est une notion importante dans la composition de la musique. L'identification de ces événements sonores permet de générer un gabarit avec un minimum de données.

b. L'utilisation de n-grammes pour la détection des répétitions

Les données musicales sont transposées en une représentation textuelle composée d'un alphabet réduit à deux caractères D et U où D équivaut à un intervalle descendant (downward) et U un intervalle ascendant (upward). Un point important de divergence entre la technique proposée et celles issues du domaine textuel est que la taille des n-grammes n'est pas fixée. Cette variante permet de détecter des répétitions de longueur différente et de faire ressortir les séquences qui composent ces répétitions.

c. L'architecture

Le système de Patel et Mundur se compose de sept modules. La figure 3.8 illustre cette architecture. Le module d'analyse fréquentielle (Frequency Domain Data Analyzer), le module de conversion des données sonores en données textuelles DU (DU Pattern Generator), le module de gestion des n-grammes (N-Gram engine) et le module d'évaluation (Pattern Ranker) constituent les modules critiques.

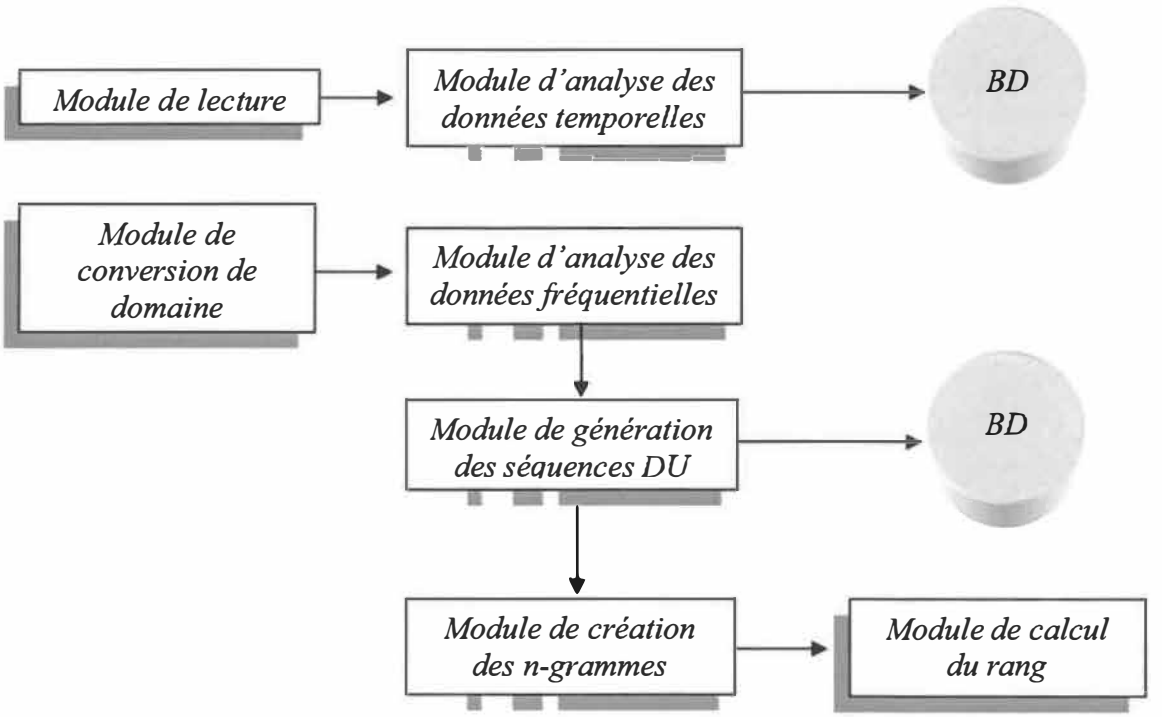


Figure 3.8 : Architecture du système[4].

[4] Image reproduite de l'article "*An n-gram based approach to finding the repeating pattern in muical data*" de. Nikunj Patel et Padma Mundur.

d. L'analyse fréquentielle

Le module d'analyse fréquentielle est responsable de la détection des notes. Les données initiales sont segmentées et soumises à une transformée de Fourier qui permet de détecter les fréquences dominantes. La technique génère pour chacun des segments une liste d'éléments correspondant aux maximums locaux considérés comme étant des notes. Pour éliminer les maximums pouvant résulter d'un bruit dans le signal sonore, un seuil de tolérance est défini. La valeur du seuil est calculée dynamiquement à partir de la formule suivante :

$$seuil = \left(\frac{top1 + top2}{\alpha} \right) + \beta \quad où \quad \alpha > 2, \beta > 0 \qquad \textbf{(3.7)}$$

Les variables top1 et top2 de la formule 3.7 représentent les deux maximums locaux les plus élevés du segment considéré. Le paramètre α est un facteur d'atténuation qui limite le nombre de notes détectées dans un segment tandis que le paramètre β est le facteur de bruit ajouté au seuil pour éliminer le bruit que peuvent contenir les silences. Les notes de même amplitude sont considérées lorsque leur valeur est supérieure au seuil de tolérance ce qui permet de traiter les sons polyphoniques.

e. La conversion des données sonores et l'extraction des n-grammes

Les modules de conversion et de gestion des n-grammes sont fortement liés. Le module de conversion traduit les données générées par le module d'analyse fréquentielle en un format reconnu par le module de gestion des n-grammes. Les intervalles descendants sont traduits par la lettre *D* tandis que les intervalles ascendants sont traduits par la lettre *U*. La répétition de la même fréquence est interprétée comme un intervalle descendant.

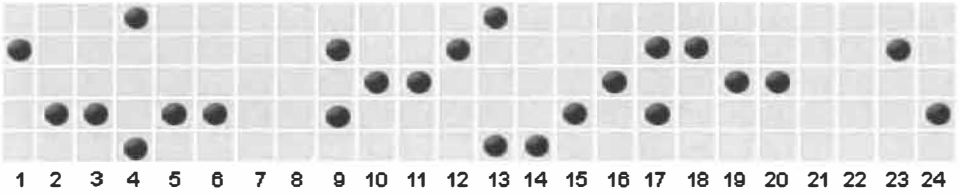


Figure 3.9 : Notes concurrentes

Les séquences *DDDU* et *DDUD* sont générées à partir des cinq premières notes illustrées à la figure 3.9. Deux séquences sont créées puisque deux notes sont jouées simultanément. Le processus d'extraction est itératif. La taille des n-grammes est incrémentée jusqu'à ce qu'aucune répétition ne soit trouvée. Ceci permet de détecter les répétitions de toutes les longueurs.

f. L'évaluation

Le rang d'une séquence est obtenu à partir de sa fréquence d'apparitions en fonction de sa taille et de la taille moyenne de l'ensemble des séquences de manière à ce que :

- La répétition la plus fréquente soit jugée la plus importante;
- Une séquence composée de plusieurs notes soit jugée plus importante qu'une séquence composée de quelques notes.

Les expérimentations effectuées [Patel et Munder, 2004] ont démontré que leur système est en mesure d'identifier les répétitions les plus significatives en leur accordant un rang élevé.

3.4. Sommaire du chapitre 3

Le son est un phénomène complexe composé d'ondes sinusoïdales. La hauteur d'un son réfère à la fréquence de l'onde qui le compose. Lorsque plusieurs fréquences interviennent simultanément, le son est dit polyphonique. Les notes de musique à partir desquelles les mélodies sont créées sont des fréquences particulières agréables pour l'oreille humaine. La différence de hauteur entre deux notes se nomme intervalle et correspond à un changement de fréquence. Ce changement peut être ascendant ou descendant. La notion d'intervalle occupe un rôle de premier plan pour les systèmes de recherche de similarités musicales. Des travaux [Downie, 1999] ont permis de démontrer qu'une représentation alphanumérique basée sur les intervalles peut être combinée à un modèle de découpage en n-grammes pour créer un support à l'analyse efficace. Cependant, la note n'est pas toujours clairement définie. Par conséquent, l'intervalle devient difficile à identifier. Les fichiers polyphoniques étant les plus répertoriés, il est impératif de prioriser une technique capable d'acquérir les caractéristiques d'une onde polyphonique. Des travaux récents effectués en reconnaissance de répétitions [Nikunj Patel et Padma Mundur, 2004] ont démontré qu'il est possible de répondre à cette contrainte par l'intermédiaire de l'analyse fréquentielle du signal.

4. Architecture du système proposé

L'approche préconisée lors de la conception de notre système de classification repose sur l'extraction de deux propriétés complémentaires du signal acoustique :

i. La forme de l'onde;

ii. Le contenu fréquentiel.

La forme de l'onde réfère au domaine spatial tandis que le contenu fréquentiel réfère au domaine spectral. L'analyse du signal acoustique donne une multitude d'indications sur la nature et la composition de l'onde sonore. Les données recueillies sont converties en caractères descriptifs de manière à créer une chaîne alphanumérique représentative du contenu des fichiers. L'ordonnancement respecte l'ordre d'apparition des caractères dans le signal d'origine. La classification est réalisée à partir d'une évaluation statistique de n-grammes extraits de l'équivalence alphanumérique.

4.1. Le caractère descriptif

Le système cherche à démontrer l'impact du caractère descriptif sur l'analyse numérique. Les résultats de la classification sont évalués en relation avec la nature du caractère utilisé. La première tâche consiste à déterminer l'élément atomique à considérer comme caractère descriptif.

4.1.1 Le point d'échantillonnage

Les fichiers sonores contiennent, pour la plupart, des points d'échantillonnage permettant de reconstituer le signal original. Ces données sont les plus élémentaires. L'étude de la conversion analogique/numérique indique que la fréquence d'échantillonnage doit être au minimum le double de la fréquence la plus élevée du signal à numériser [Shannon, 1948]. Un nombre important d'échantillons est donc nécessaire pour représenter avec fidélité le signal. Chacun d'eux équivaut à un événement sonore de très courte durée. Lorsque considéré dans son unicité ou dans son voisinage immédiat, le point d'échantillonnage est peu porteur d'information. Pour cette raison, il ne peut pas être considéré comme caractère descriptif destiné à l'analyse numérique.

4.1.2 Les paires amplitude/demie période

Le signal acoustique est caractérisé par une amplitude et une période. Nous avons donc posé l'hypothèse que les paires formées à l'aide de l'amplitude et d'une demie période permettaient de décrire de façon approximative la forme de l'onde. Ces couples forment des quadrilatères qui estiment la forme de l'onde. La figure 4.1 montre une partie d'un signal et son équivalence constituée de quadrilatères.

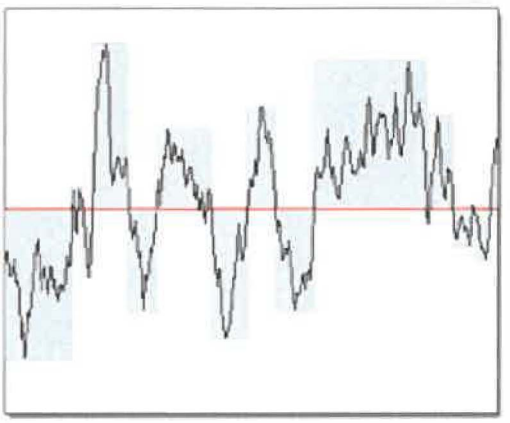


Figure 4.1 : Le caractère descriptif issu de la forme de l'onde.

4.1.3 Les fréquences dominantes

La musique est une combinaison de sons. Le résultat est un signal complexe composé de plusieurs sinusoïdes de fréquences et d'amplitudes différentes.

a. La série de Fourier

Le mathématicien français Joseph Fourier (1768 - 1830) a démontré en 1807 une relation entre un signal complexe et sa composition en sinusoïdes [Thomas J. Cavicchi, 2000].

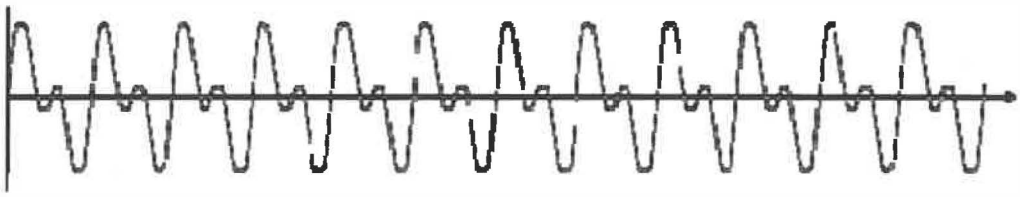


Figure 4.2 : Onde sonore complexe.

Le signal illustré à la figure 4.2 peut être décomposé sous la forme des signaux simples suivants des figures 4.3 et 4.4.

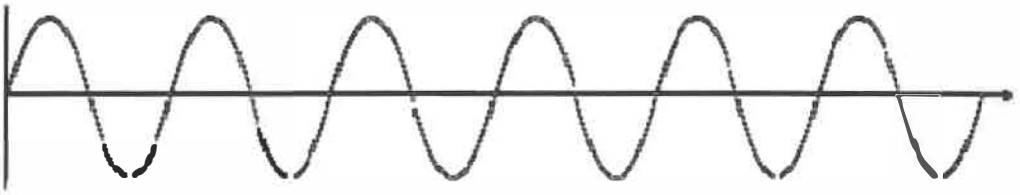


Figure 4.3 : Onde sinusoïdale basse fréquence.

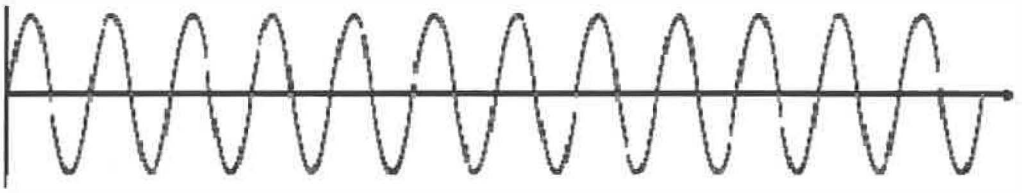


Figure 4.4 : Onde sinusoïdale haute fréquence.

Un signal peut être considéré comme une fonction continue f(t) représentant les harmoniques dans le temps.

Soit f un signal périodique de période $T = \frac{2\pi}{w}$, la série de Fourier équivaut au développement suivant :

$$f(t) = a_0 + \sum_{-\infty}^{+\infty} a_n \cos nwt + b_n \sin nwt \qquad \textbf{(4.1)}$$

On peut considérer f comme étant la somme d'un terme constant a_0 et d'un nombre infini de sinusoïdes appelés harmoniques. L'harmonique de rang n est donnée par l'équation 4.2 où n est le numéro de l'harmonique :

$$f_n = a_n \cos nwt + b_n \sin nwt \qquad \textbf{(4.2)}$$

L'harmonique de premier rang se nomme fréquence fondamentale et est notée f_0. La valeur de la fréquence fondamentale caractérise la hauteur de l'échantillon sonore. La connaissance de f_0 aide à différencier la source sonore. Par exemple, la connaissance de f_0 permet de faire la différence entre la voix d'un homme, d'une femme et d'un enfant. En moyenne, la fréquence fondamentale de la voix d'un homme se situe entre 100 et 150 hertz, entre 200 et 300 hertz pour la voix d'une femme et entre 350 et 400 hertz pour la voix d'un enfant [Bernard Caillaud et Mireille Leriche, 1999].

La représentation de l'amplitude des harmoniques en fonction de leurs fréquences illustre le spectre de fréquence d'un signal (figure 4.5).

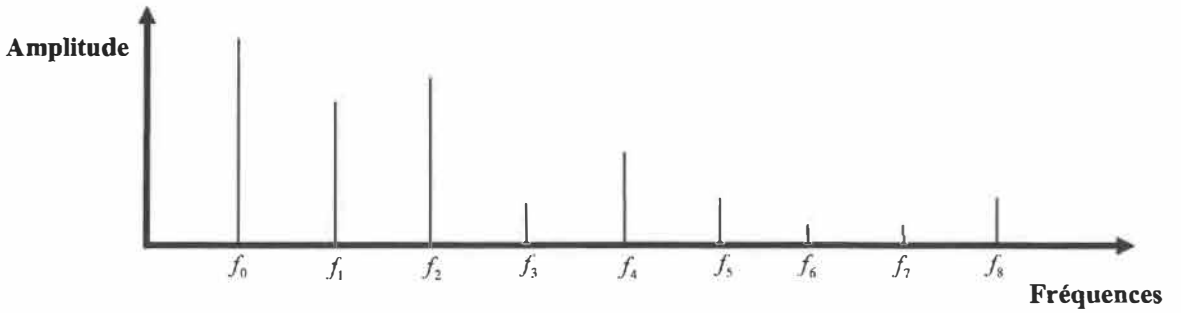


Figure 4.5 : Spectre d'un signal sonore.

b. La transformée de Fourier

La transformée de Fourier est une opération mathématique qui extrait les composantes d'un signal (fréquence) en fonction d'une variable (temps). Elle permet le passage du domaine spatial au domaine spectral. La transformée de Fourier est donnée par l'équation 4.3.

$$F(v) = \int_{-\infty}^{+\infty} f(x)\, e^{-ix(2\pi v)} dx \qquad \textbf{(4.3)}$$

La transformée de Fourier inverse (équation 4.4) permet le passage du domaine spectral au domaine spatial.

$$f(x) = \int_{-\infty}^{+\infty} F(t)\, e^{i2\pi xt} dx \qquad \textbf{(4.4)}$$

Le tableau 4.1 donne les représentations spatiale et spectrale d'un signal.

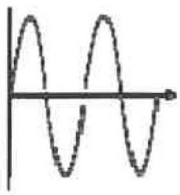
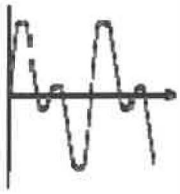
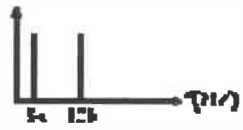
Description	Domaine spatial	Série de Fourier	Domaine spectral
Signal sinusoïdal 5k Hz		$f(t) = \sin(2\pi(5kHz))$	
Signal composé de 2 signaux sinusoïdaux de 5k Hz et 10k Hz		$f(t) = \sin(2\pi(5kHz)) + \sin(2\pi(10kHz))$	
Signal composé de 3 signaux sinusoïdaux de 5k Hz, 10k Hz et 20k Hz		$f(t) = \sin(2\pi(5kHz)) + \sin(2\pi(10kHz)) + \sin(2\pi(20kHz))$	

Tableau 4.1 : Signaux dans les domaines spatial et spectral[5].

c. Le caractère descriptif fréquentiel

Le contenu fréquentiel est un descripteur fidèle de la nature d'une onde sonore. Nous nous sommes intéressés à la bande de fréquence à laquelle appartiennent les fréquences dominantes du signal acoustique. Ces bandes de fréquences sont les indices importants sur la nature de l'onde sonore [Gonzales, 1992].

5 Tableau reproduit de la fiche technique DataQ de Melissa Ray Weimer.

4.2. La chaîne de traitements

Le processus de classification proposé est semi-automatique. Le choix de certains paramètres est fait par l'utilisateur en fonction de ses propres objectifs, de sa subjectivité et de son rôle dans l'interprétation des résultats. Les étapes de la classification sont :

- La lecture des données;
- La préparation des données;
- L'extraction des n-grammes et création des représentations vectorielles;
- La présentation des données au classifieur;
- L'évaluation des classes obtenues.

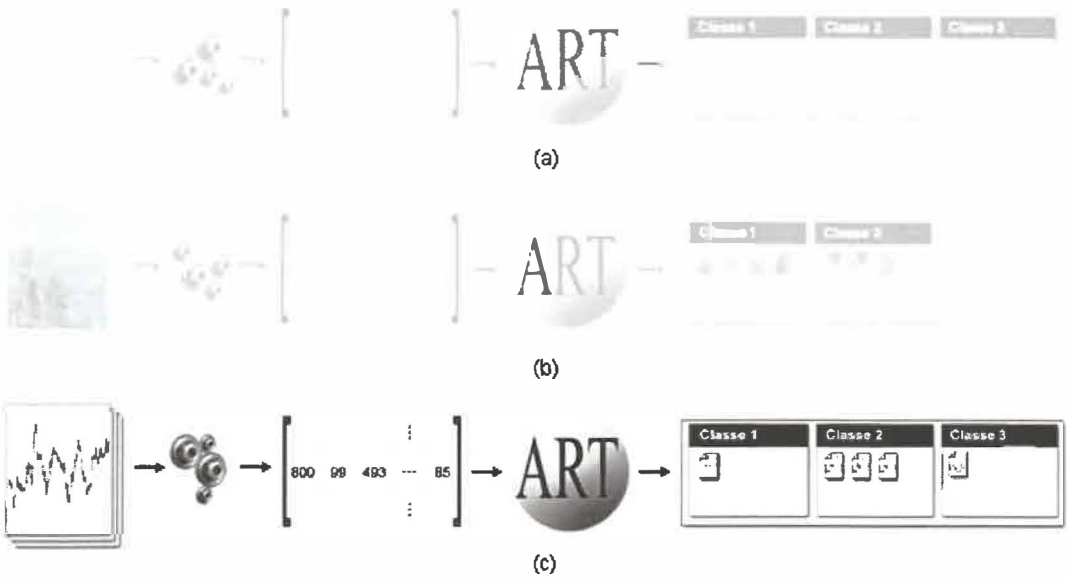


Figure 4.6 : Chaîne de traitements.

Comme l'illustre la figure 4.6, cette même chaîne de traitements peut être utilisée pour différents formats de fichiers. Seules la lecture et la préparation des données diffèrent d'un format à l'autre. Les domaines du texte, de l'image et du son sont respectivement identifiés par les chaînes (a), (b) et (c). Le détail de la chaîne (c) est donné à la figure 4.7.

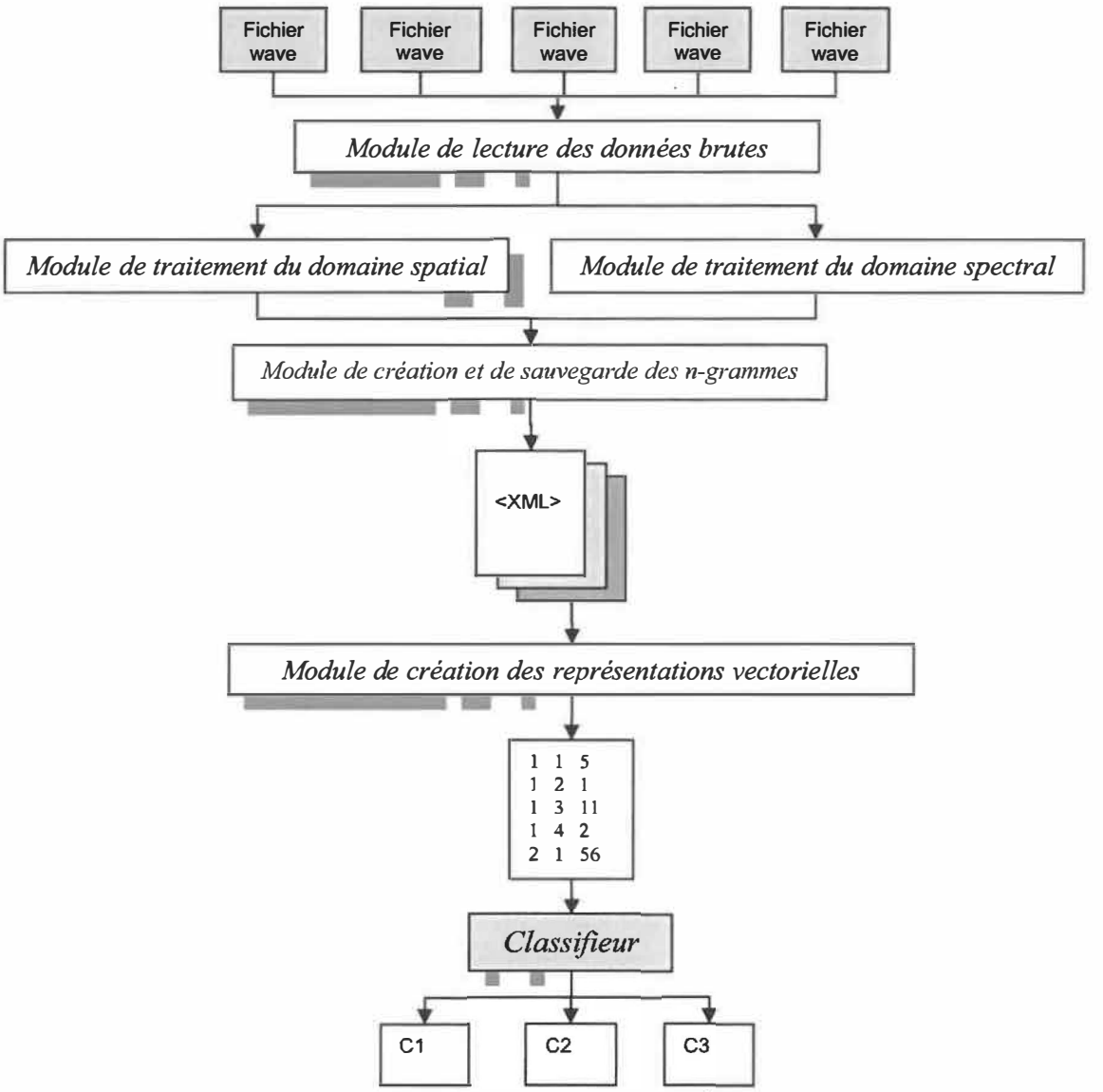


Figure 4.7 : Architecture du système proposé.

Les fichiers wave sont lus par un premier module puis les données brutes sont transmises à l'un des modules de traitements. Les données sont d'abord apprêtées afin d'être uniformisées (voir section 4.2.2). Selon l'analyse effectuée, les caractéristiques spatiales ou spectrales sont prélevées et transformées en caractères descriptifs. La liste est par la suite passée au module de création et de sauvegarde des n-grammes. Elle est dans un premier temps découpée en n-grammes puis l'inventaire est sauvegardé dans un fichier afin d'être réutilisé. Un fichier est créé par document soumis à la classification. Lorsque tous les documents sont traités, le module de création des représentations vectorielles est sollicité. Les fichiers inventaires sont décortiqués afin qu'une représentation vectorielle uniforme soit créée pour chacun des documents wave. Les vecteurs sont présentés au classifieur afin que les classes de similarités soient déduites et présentées à l'utilisateur.

4.2.1 La lecture des données

La première étape consiste à faire la lecture des données brutes. La lecture est réalisée à l'aide d'une librairie externe. Par conséquent, l'analyse de différents formats audio requiert uniquement l'ajout des librairies utiles à l'interprétation de ces formats et n'a aucun impact sur le restant de la chaîne de traitements. Cette flexibilité est nécessaire en raison de la nature évolutive de la représentation des données.

a. La lecture

Dans le cadre de nos expérimentations, l'input est constitué d'un ensemble de fichiers sonores de type WAVE PCM mono. Deux facteurs ont motivé cette décision :

i. Le format WAVE PCM est le format audio non compressé le plus répandu;

ii. Le traitement d'un seul canal est suffisant pour effectuer des observations.

b. Le format WAVE PCM

Le format WAVE est utilisé pour stocker des données audionumériques. Il est dérivé de la spécification RIFF (*The Resource Interchange File Format*) développée conjointement par les entreprises IBM et Microsoft. L'information est emmagasinée par octet dont l'ordonnancement correspond à la norme d'Intel 80x86 (*little-endian*) c'est-à-dire que l'octet de poids faible est stocké avant l'octet de poids fort. Les fichiers qui utilisent la spécification RIFF sont composés de différents segments étiquetés nommés chunk. La liste et l'ordonnancement des ces éléments peuvent différer d'un fichier à l'autre. Cependant, trois sont communs à tous les fichiers WAVE : les chunks « riff », « fmt » et « data ». Les deux premiers forment l'entête du fichier : ils spécifient la structure et contiennent les paramètres fondamentaux tels l'échelle de quantification, le nombre d'échantillons par seconde, le nombre de canaux etc. Finalement le troisième renferme les points d'échantillonnage.

Une partie importante de l'interprétation des fichiers WAVE consiste à lire les échantillons. Un point d'échantillonnage est une valeur qui représente un fragment de l'onde sonore à un temps précis. Puisque la majorité des CPU effectuent les opérations de lecture et d'écrire un octet à la fois, il a été défini que les points d'échantillonnage soient quantifiés à l'aide de multiples de 8 afin de faciliter ces opérations. Des bits de remplissage dont la valeur est 0 sont parfois insérés afin d'obtenir un multiple de 8 (figure 4.8).

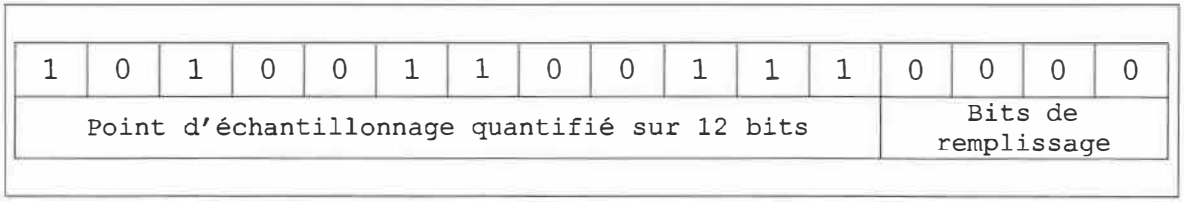


Figure 4.8: Représentation d'un point d'échantillonnage avec des bits de remplissage.

Puisque le format WAVE utilise la norme *little-endian* l'octet le moins significatif est stocké en premier dans le fichier. La figure 4.9 démontre la représentation binaire du point d'échantillonnage de la figure 4.8 selon la norme *little-endian*.

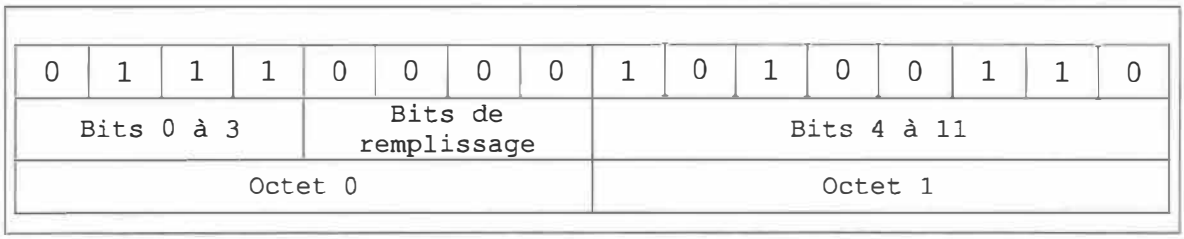


Figure 4.9 : Représentation d'un point d'échantillonnage selon la norme *little-endian*.

Pour les sons multicanaux comme la stéréo, les points d'échantillonnage sont entrelacés. Plutôt que de stocker tous les points d'échantillonnage du canal gauche puis tous ceux du canal droit, les points d'échantillonnage sont mixés : le premier point du canal gauche est stocké puis le premier point du canal droit ; le second point du canal gauche est stocké puis le second point du canal droit et ainsi de suite. La figure 4.10 démontre ce principe. La technique est la même pour les sons qui possèdent plus de deux canaux. L'avantage de cette approche est que les points d'échantillonnage qui doivent être interprétés simultanément sont stockés de manière continue dans le fichier.

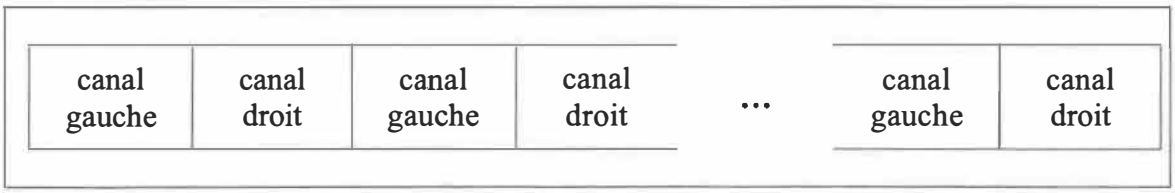


Figure 4.10 : Stockage des points d'échantillonnages entrelacés par alternance des canaux.

4.2.2 La préparation des données

C'est à l'étape de la préparation des données que l'usager intervient. Les paramètres lui permettent d'apprêter les données de sorte qu'elles correspondent à ses objectifs. Les paramètres offerts dépendent du mode d'analyse effectué.

Les prétraitements permettent de normaliser et de simplifier les données brutes de manière à faciliter la classification. Les paramètres sont :

- La taille des n-grammes;
- La taille du lexique;
- La projection des valeurs qui représente un facteur de compression;
- Le lissage.

a. La taille des n-grammes

La taille des n-grammes influence la probabilité d'apparition de ceux-ci. Une taille importante engendre un nombre élevé de n-grammes distincts et par conséquent une probabilité d'occurrences faible.

b. La taille du lexique

Un certain contrôle peut être exercé sur la taille du lexique.

La taille du lexique dans le domaine spatial

Le domaine spatial considère la forme de l'onde sonore. Les points d'échantillonnage permettent de reconstituer cette forme. La réduction de leur quantification offre donc la possibilité de limiter la taille de l'alphabet. Les échantillons codés sur 16 bits sont convertis sur 8 bits. Ce prétraitement permet de limiter l'alphabet à 256 éléments plutôt qu'à 65 535. Un même signal peut être quantifié de manière différente. Même si la qualité de la numérisation est affectée par la réduction de la quantification de chacun des échantillons, le signal peut tout de même être reproduit et reconnu.

La taille du lexique dans le domaine spectral

Pour le domaine spectral, ce sont les fréquences qui sont examinées. L'alphabet est déterminé par le nombre de fréquences considérées.

c. La projection des valeurs

La projection consiste à déterminer un seuil de tolérance permettant de regrouper certaines valeurs jugées similaires.

La projection des valeurs dans le domaine spatial

La projection est une réduction de l'échelle des valeurs admissibles. Les valeurs réelles sont projetées sur cette échelle.

$$(f_i - f_{\min} / f_{\max} - f_{\min}) * a \qquad \textbf{(4.5)}$$

La projection des valeurs est définie à partir de la formule 4.5 où f_i représente la i ième valeur, $f_{\min}$ la valeur minimum, $f_{\max}$ la valeur maximum et a une constante représentant la taille de l'intervalle de projection.

La réduction de la quantification des points d'échantillonnage est également une forme de projection.

La projection des valeurs dans le domaine spectral

Dans le domaine spectral, la projection équivaut à regrouper les fréquences du signal. Les découpages en 10 bandes (tableau 4.2), 19 bandes (tableau 4.3) et 31 bandes (tableau 4.4) peuvent être utilisés.

Bandes	Fréquences
1	0 - 31 Hz
2	32 - 63 Hz
3	64 - 125 Hz
4	126 - 250 Hz
5	251 - 500 Hz
6	501 - 1 000 Hz
7	1 001 - 2 000 Hz
8	2 001 - 4 000 Hz
9	4 001 - 8 000 Hz
10	16 000 Hz et plus

Tableau 4.2 : Découpage en 10 bandes.

Bandes	Fréquences
1	0 - 31 Hz
2	32 - 44 Hz
3	45 - 63 Hz
4	64 - 88 Hz
5	89 - 125 Hz
6	126 - 180 Hz
7	181 - 250 Hz
8	251 - 355 Hz
9	356 - 500 Hz
10	501 - 710 Hz
11	711 - 1 000 Hz
12	1 001 - 1 400 Hz
13	1 401 - 2 800 Hz
14	2 801 - 4 000 Hz
15	4 001 - 5 600 Hz
16	5 601 - 8 000 Hz
17	8 001 - 11 300 Hz
18	11 301 - 16 000 Hz
19	16 000 Hz et plus

Tableau 4.3 Découpage en 19 bandes.

Bandes	Fréquences
1	0 - 25 Hz
2	26 - 31 Hz
3	32 - 40 Hz
4	41 - 50 Hz
5	51 - 63 Hz
6	64 - 80 Hz
7	81 - 100 Hz
8	101 - 125 Hz
9	126 - 160 Hz
10	161 - 200 Hz
11	201 - 250 Hz
12	251 - 315 Hz
13	316 - 400 Hz
14	401 - 500 Hz
15	501 - 630 Hz
16	631 - 800 Hz
17	801 - 1 000 Hz
18	1 000 - 1 250 Hz
19	1 251 - 1 600 Hz
20	1 601 - 2 000 Hz
21	2 001 - 2 500 Hz
22	2 501 - 3 150 Hz
23	3 151 - 4 000 Hz
24	4 001 - 5 000 Hz
25	5 001 - 6 300 Hz
26	6 301 - 8 000 Hz
27	8 001 - 10 000 Hz
28	10 001 - 12 500 Hz
29	12 501 - 16 000 Hz
30	16 000 - 20 000 Hz
31	20 001 Hz et plus

Tableau 4.4 : Découpage en 31 bandes.

Ces découpages d'une onde échantillonnée à 44 100 hertz correspondent à ceux définis pour les appareils audionumériques commerciaux.

d. Le lissage

Le lissage est une opération qui élimine les détails et le bruit que peut contenir un signal. Le filtre gaussien 1-D est utilisé. Ce filtre est un opérateur de convolution qui élimine le détail et le bruit d'un signal [Meunier et al, 2001].

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \qquad \textbf{(4.6)}$$

La formule 4.6 donne le noyau du filtre gaussien 1-D où σ correspond à la dimension du filtre numérique. La figure 4.11 illustre la forme du filtre.

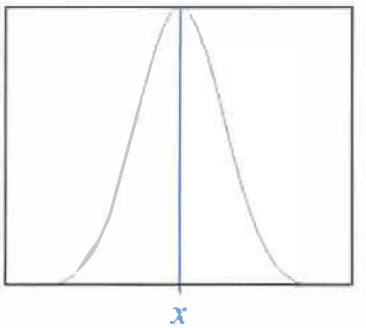


Figure 4.11 : Forme du filtre gaussien 1-D.

La convolution est une opération analogue de la multiplication spectrale. Le spectre du signal résultant est le produit terme à terme du spectre du signal original et de celui du filtre. Le filtre gaussien est donc un opérateur de lissage qui a comme effet d'éliminer certaines fréquences. Les figures 4.12, 4.13 et 4.14 illustrent l'effet de cette opération avec différentes valeurs de σ. Le signal en gris représente le signal d'origine tandis que le signal en bleu représente le signal lissé. Plus σ est grand plus le lissage est fort.

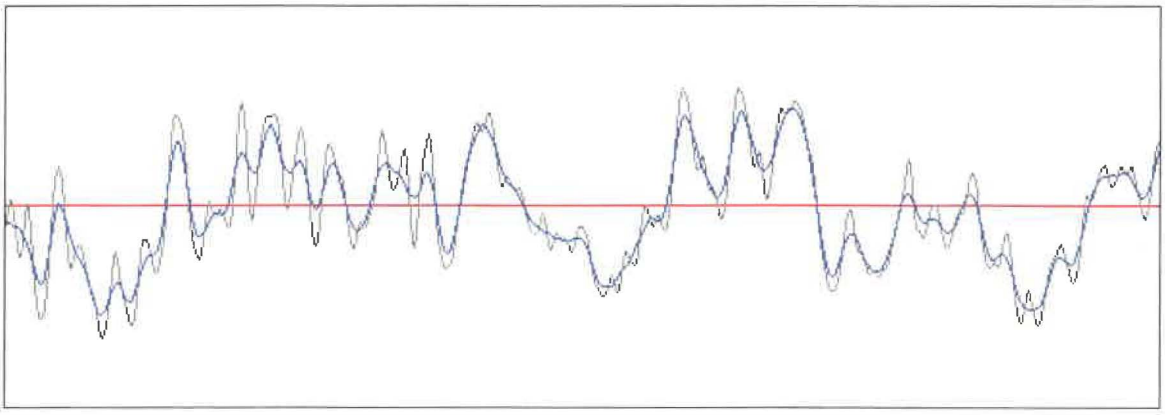


Figure 4.12 : Lissage d'une onde avec $\sigma = 3$.

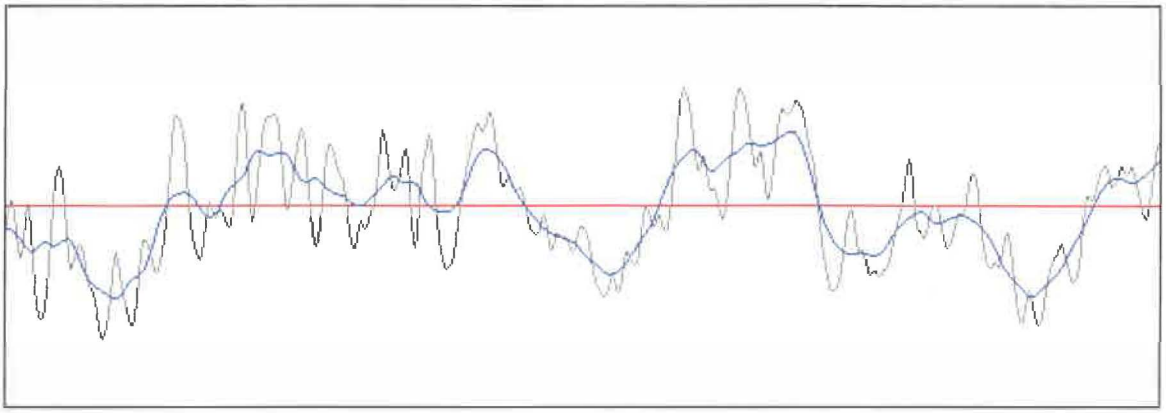


Figure 4.13 : Lissage d'une onde avec $\sigma = 7$.

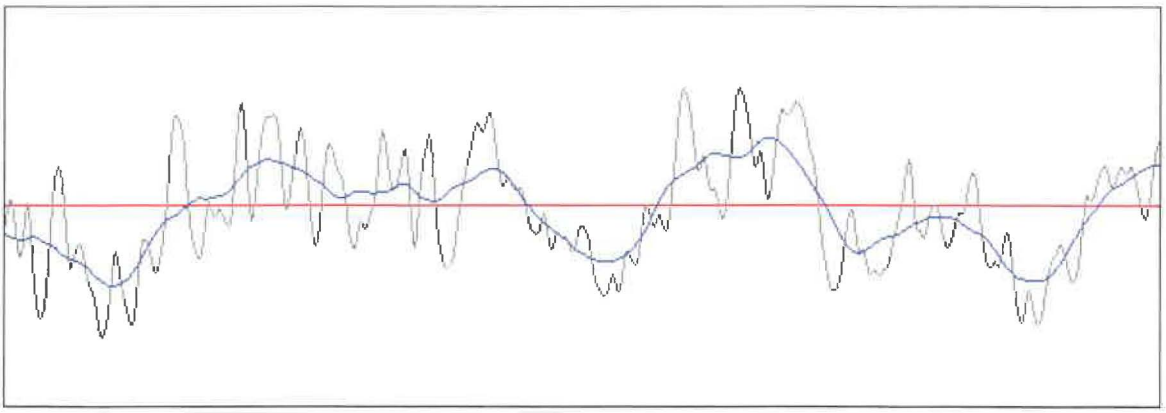


Figure 4.14 : Lissage d'une onde avec $\sigma = 11$.

4.2.3 L'extraction des n-grammes et création de la représentation vectorielle

Le découpage en n-grammes est effectué à partir de la liste de caractères informationnels qui correspond soit aux paires amplitude et demie-période soit aux bandes de fréquences auxquelles appartiennent les fréquences dominantes. La liste des n-grammes extraits est conservée dans un fichier XML. Lorsque tous les documents sont traités, le fichier XML est transposé en matrice afin de faire ressortir la distribution des différents n-grammes à l'intérieur des fichiers. La structure du fichier XML est donnée par le tableau 4.5.

```
<WaveFile>
   <FileName>1</FileName>
   <Url>C:\Documents and Settings\Louis\Desktop\UQTR\1.wav</Url>
   <RiffChunk>
          <sGroupID>RIFF</sGroupID>
          <dwFileLength>441046</dwFileLength>
          <sRiffType>WAVE</sRiffType>
  </RiffChunk>
  <FormatChunk>
          <sChunkID>fmt </sChunkID>
          <dwChunkSize>16</dwChunkSize>
          <wFormatTag>1</wFormatTag>
          <wChannels>1</wChannels>
          <dwSamplesPerSec>44100</dwSamplesPerSec>
          <dwAvgBytesPerSec>88200</dwAvgBytesPerSec>
          <wBlockAlign>2</wBlockAlign>
          <dwBitsPerSample>16</dwBitsPerSample>
  </FormatChunk>
  <DataChunk>
          <sChunkID>data</sChunkID>
          <dwChunkSize>441002</dwChunkSize>
          <lFilePosition>44</lFilePosition>
          <duree>0,05</duree>
          <dwNumSamples>220501</dwNumSamples>
  </DataChunk>
  <nGramme>
          <valeur>[2-3]</valeur>
          <occurence>21</occurence>
  </nGramme>
  <nGramme>
           <valeur>[2-4]</valeur>
          <occurence>5</occurence>
  </nGramme>
   .
   .
   .
</WaveFile>
```

Tableau 4.5 : Structure du fichier XML contenant l'inventaire des n-grammes extraits d'un document WAVE.

Le fichier XML est divisé en deux sections. La première réfère directement au document wave[6]. Le nom du document source et son emplacement sont les premières données contenues dans le fichier. Le contenu des chunks Riff et Format succède. L'entête du chunk Data conclut la première section. La deuxième partie du fichier XML est dédiée à l'inventaire des n-grammes. La valeur et le nombre d'occurrences rencontrées sont précisés.

L'extraction des n-grammes du domaine spatial

Les paires composées d'amplitudes et de demi-périodes forment les caractères descriptifs pour le domaine spatial. Pour obtenir ces paires, le signal est parcouru à la recherche des passages par zéro. Le nombre de points d'échantillonnage recensé entre deux passages par zéro est prélevé et considéré comme une demi-période. Ce nombre doit être supérieur à un sans quoi l'intervalle est ignoré. L'amplitude, quant à elle, équivaut au maximum local pour les valeurs supérieures à zéro ou au minimum local pour les valeurs inférieures à zéro. La figure 4.15 illustre les paires ainsi formées pour un signal quelconque.

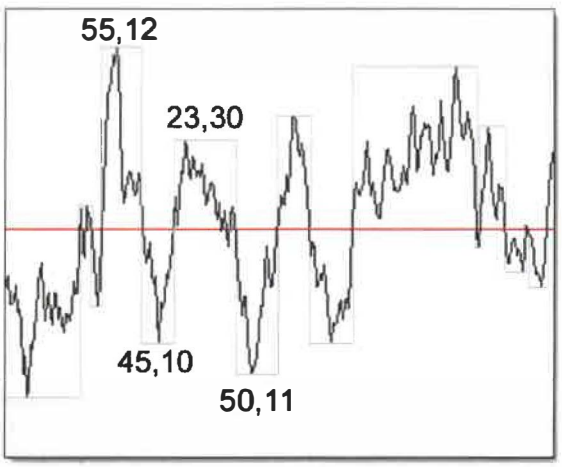


Figure 4.15 : Paires formées des amplitudes et des demies périodes.

[6] Voir annexe A pour obtenir la structure d'un fichier WAVE.

Les paires sont stockées en mémoire dans une liste en ordre d'apparition. Cette liste est par la suite découpée en n-grammes. Le tableau 4.6 montre le résultat d'un découpage en bi-grammes réalisé à partir de l'équivalence « amplitude – demi-période » du signal de la figure 4.15. Le premier élément du couple représente l'amplitude et le second la demi-période pour la partie du signal illustrée à la figure 4.15.

Bi-Gramme
(52-12)(45,10), (45,10)(23,30), (23,30)(50,11),...

Tableau 4.6 : N-grammes composés de paires d'amplitudes et de demi-périodes.

L'extraction des n-grammes du domaine spectral

Pour le domaine spectral, les bandes de fréquences auxquelles appartiennent les fréquences dominantes du signal sont définies comme caractère descriptif. Pour obtenir ces valeurs, le signal est dans un premier temps segmenté en tranches de 4, 8, 16 ou 32 octets avec chevauchement. La taille du segment influence la précision de l'opération. Une fenêtre de 32 octets donne une représentation plus fidèle du contenu fréquentiel mais nécessite un traitement plus lourd. Une transformée de Fourier est appliquée à chacun des segments afin d'obtenir leur contenu fréquentiel. La puissance est ensuite calculée à l'aide de la formule (4.7) où i est un nombre complexe représentant une fréquence du signal.

$$p = \sqrt{i.réel * i.réel + i.imaginaire * i.imaginaire} \qquad \textbf{(4.7)}$$

La puissance donne l'apport de chaque harmonique de fréquence différente dans le signal. La figure 4.16 représente graphiquement cette relation où la partie de gauche contient les basses fréquences et la partie de droite les hautes fréquences.

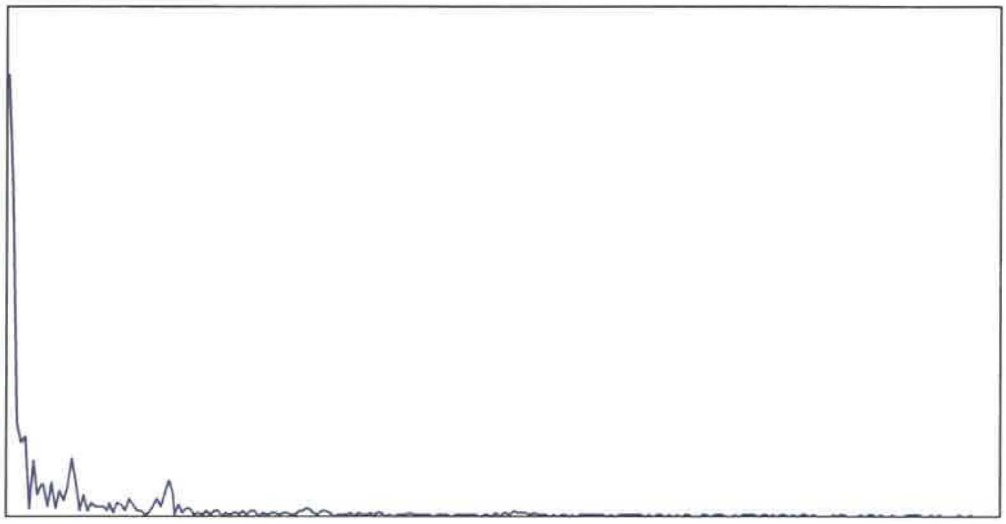


Figure 4.16 : Apport des fréquences dans un signal quelconque.

La pointe située à l'extrême gauche de la figure 4.16 représente la fréquence fondamentale qui équivaut à la moyenne de l'amplitude du signal.

La courbe résultante est filtrée afin d'isoler les maximums locaux. Cette opération est effectuée à l'aide d'un opérateur de convolution et du filtre de la dérivée première 1-D dont le noyau est donné par :

$$G'(x) = -(\frac{1}{2.50663 * \delta^3}) * x * e^{\frac{-x^2}{2 * \delta^2}} \qquad \textbf{(4.8)}$$

Le filtre de la dérivée première a comme caractéristique de permettre l'extraction des changements de pente d'une courbe. Ces changements représentent les maximums et les minimums locaux. Dans le contexte de cette recherche, seuls les maximums sont considérés (figure 4.17).

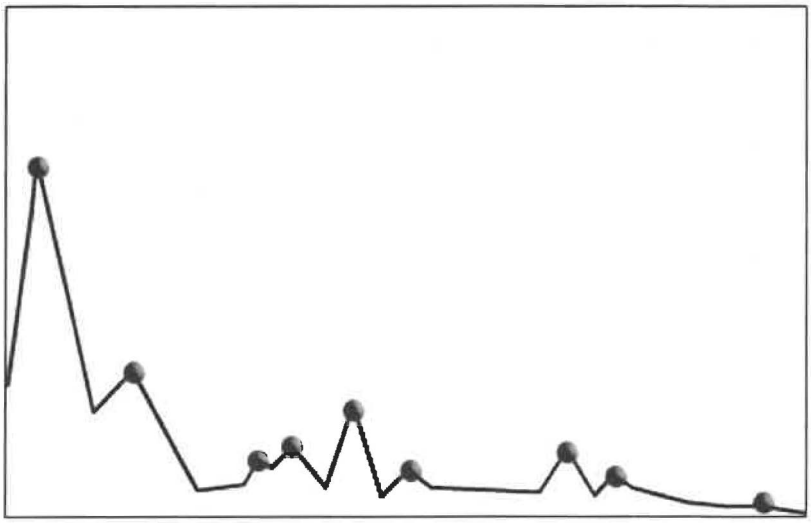


Figure 4.17 : Les maximums locaux associés aux fréquences dominantes

Les maximums donnent les fréquences dominantes du segment source. Ceux-ci sont triés en ordre décroissant et les n plus grands sont conservés. Les bandes de fréquences correspondantes sont définies comme caractère descriptif.

Cette association est illustrée aux figures 4.18, 4.19 et 4.20. Les plus hauts sommets représentent des fréquences dominantes.

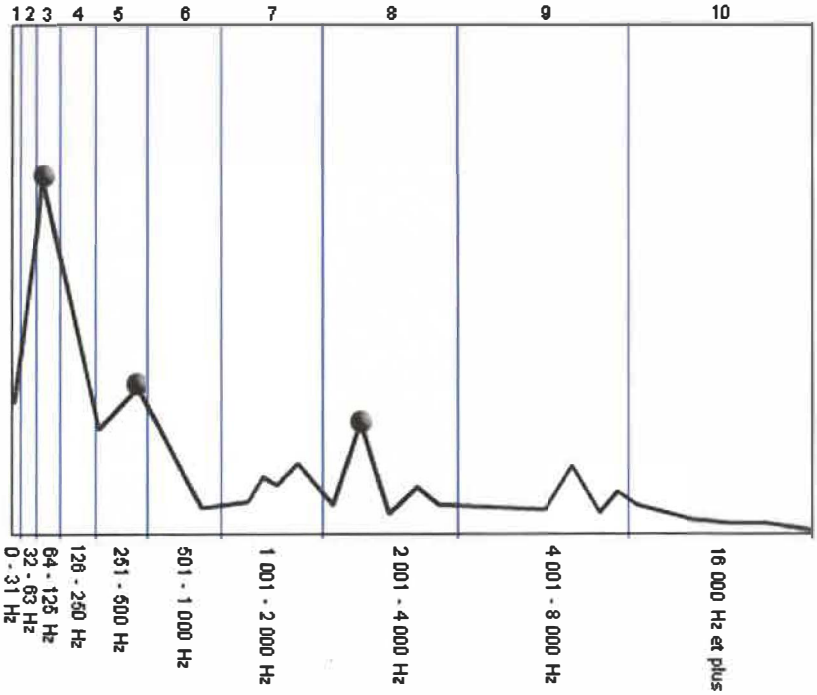


Figure 4.18 : Les bandes de fréquences des 3 fréquences dominantes du segment 1.

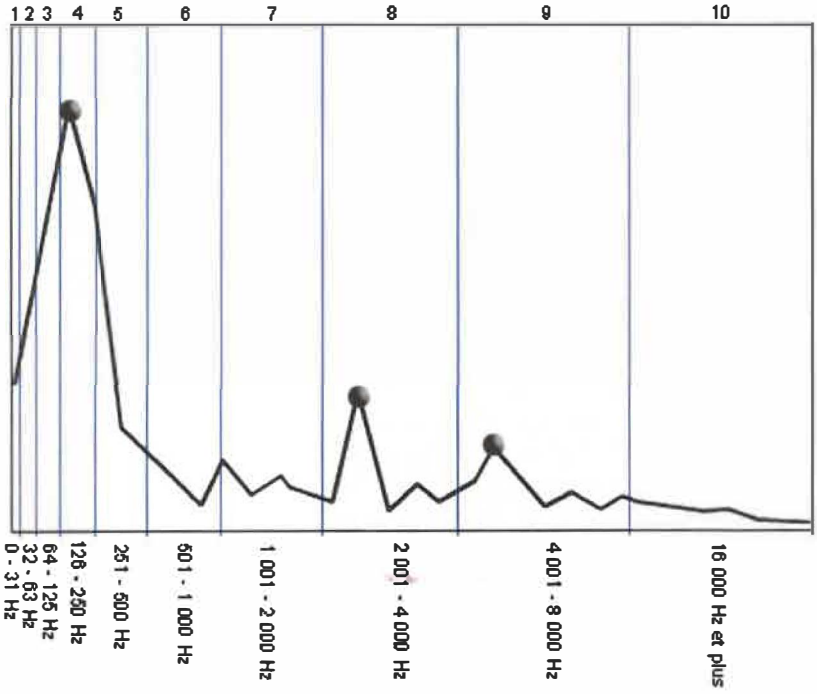


Figure 4.19 : Les bandes de fréquences des 3 fréquences dominantes du segment 2.

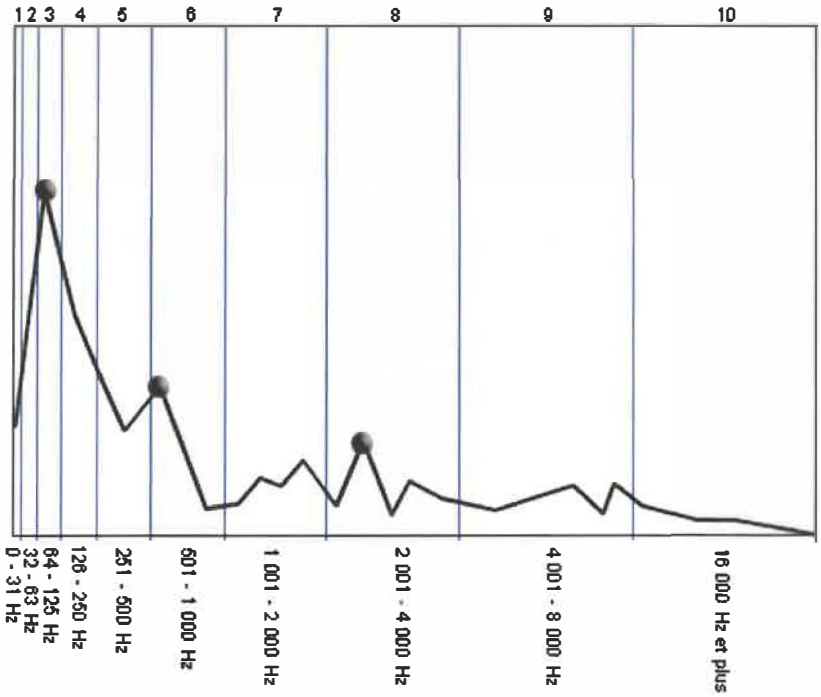


Figure 4.20 : Les bandes de fréquences des 3 fréquences dominantes du segment 3.

Les *n* bandes de fréquences dominantes de chacun des segments sont stockées en mémoire dans une liste par ordre d'apparition. Cette liste est par la suite découpée en n-grammes. Le tableau 4.7 donne un exemple de résultat obtenu.

Bi-Gramme
35, 59, 94, 48, 89, 93, 36, 68, ...

Tableau 4.7 : N-grammes composés de bandes de fréquences.

Les chiffres du tableau 4.7 représentent les bandes de fréquence des plus hauts sommets illustrés aux figures 4.18, 4.19 et 4.20.

Création de la représentation vectorielle

Cette opération consiste à apprêter les données selon le format attendu par le classifieur. Une matrice 3 par *n* est créée pour représenter l'ensemble des fichiers où *n* est le nombre total de n-grammes différents comptabilisés. Une rangée peut être interprétée comme le nombre d'apparitions du n-gramme *x* dans le fichier *y*.

Identifiant du fichier	Identifiant du n-gramme	Nombre d'apparition
1	1	2
1	2	54
1	3	1
1	4	1
2	2	2
2	3	5
2	5	1

Tableau 4.8 : Représentation vectorielle.

Le tableau 4.8 illustre la représentation vectorielle définie pour deux fichiers. Un fichier est représenté par plusieurs rangées successives. Un n-gramme doit avoir le même identifiant pour l'ensemble des fichiers. La matrice est conservée dans un fichier texte afin d'être exploitée par le classifieur.

4.2.4 Présentation des données au classifieur

La représentation vectorielle des fichiers est présentée au classifeur. Dans le cadre de nos expérimentations, ART a été utilisé. Le choix de ART n'est pas dicté par des raisons de performances particulières. L'objectif n'est pas d'optimiser la classification mais plutôt d'explorer les avenues possibles. Le choix de ART été effectué en continuité avec les travaux déjà effectués [Biskri et al, 2002].

a. ART (*Adaptative Resonnance Therory*)

ART est un réseau de neurones formels à apprentissage par compétition [Gail A. Carpenter et Stephen Grossberg, 2003]. Le neurone formel est une modélisation mathématique qui calque le fonctionnement du neurone biologique. ART est un réseau de neurones à apprentissage non supervisé. Le nombre total de classes obtenues dépend d'un paramètre de vigilance ρ dont la valeur varie entre 0 et 1. Théoriquement, le classifieur tend à être plus sélectif lorsque ρ tend vers 1.

La structure de ART est subdivisée en deux couches de neurones : une couche entrée/sortie et une couche cachée. Tous les neurones de la couche d'entrée/sortie sont reliés à tous les neurones de la couche cachée et vice versa. Il n'existe aucune connexion entre les neurones de la couche d'entrée/sortie. Un poids est fixé à chacune des connexions. Lorsqu'un vecteur $\vec{v}$ est soumit au réseau, la couche d'entrée/sortie active les neurones de la couche cachée. Les neurones sollicités sont en compétition de manière à sélectionner un seul candidat considéré comme le plus représentatif du vecteur $\vec{v}$. Le neurone gagnant sollicite les neurones de la couche d'entrée/sortie afin d'être comparé à l'input. Lorsque celui-ci ne répond pas à un seuil de similarité déterminé en fonction du paramètre de vigilance ρ, la couche d'entrée/sortie existe de nouveau la couche cachée en excluant le neurone jugé comme mauvais candidat. Si tous les neurones de la couche cachée ne correspondent pas à $\vec{v}$, un nouveau neurone caché est créé afin de représenter le vecteur $\vec{v}$. Lorsqu'un neurone caché est considéré comme semblable, le vecteur $\vec{v}$ est associé à ce neurone. La pondération entre les différentes connexions est révisée afin de consolider l'association effectuée.

4.2.5 Évaluation des classes obtenues

L'évaluation de la pertinence des classes obtenues est réalisée en relation avec la nature du caractère descriptif. Les classes obtenues sont analysées afin de déterminer s'il est possible de faire ressortir une caractéristique commune à tous les membres. Une classe est dite neutre lorsqu'elle contient un seul élément.

Une classification est jugée satisfaisante si le nombre de classes cohérentes surpasse le nombre de classes erronées ou neutres.

4.3. Sommaire du chapitre 4

Nous avons développé une application qui prépare les données de manière à générer une liste de caractères descriptifs. Deux propriétés de l'onde sont considérées : la forme et le contenu fréquentiel. La liste créée est découpée en n-grammes. Les n-grammes répertoriés sont structurés de façon à créer une représentation vectorielle des fichiers. Cette opération consiste à transposer la distribution des n-grammes dans le format attendu par le classifieur. L'étape de la classification est réalisée par l'entremise de réseaux de neurones ART.

Les résultats obtenus sont présentés au chapitre 5.

5. Expérimentation et discussions

Plusieurs expérimentations ont été effectuées afin de mesurer l'efficacité du système proposé. Le présent chapitre dresse les résultats obtenus. La complexité du domaine d'information a été augmentée d'une expérimentation à l'autre.

Lors de la première expérimentation, des notes de musique ont été classées afin de vérifier la capacité du système à les différencier.

Au cours de la deuxième expérimentation, des séries de notes ont été classées. Nous avons tenté de déterminer si les phénomènes observés lors de la première expérimentation persistent lorsque plusieurs notes sont contenues dans le même fichier.

Finalement, la classification de documents polyphoniques a été le thème étudié lors de la troisième classification.

5.1. Classification de notes de musique

La première expérimentation a été réalisée à partir de fichiers sonores dont le contenu correspond à une note de musique. La notation utilisée pour représenter les notes est donnée au tableau 5.1. Cet ordonnancement équivaut aux touches d'un piano où les notes altérées sont représentées par les touches noires et les notes non altérées par les touches blanches.

do	**ré b**	**ré**	**mi b**	**mi**	**fa**	**fa #**	**sol**	**sol #**	**la**	**si b**	**si**
C	Db	D	Eb	E	F	F#	G	G#	A	Bb	B

Tableau 5.1 : Notes contenues dans les fichiers sonores.

Le dièse, noté « # », altère une note d'un demi-ton plus aigu tandis que le bémol, noté « b », altère une note d'un demi-ton plus grave. L'ordre définie (C, Db, D, Eb, E, F, F#, G, G#, A, Bb et B) est en fonction du rapport de fréquence entre les notes et par conséquent leur rapport de forme. Le tableau 5.2 donne le rapport de fréquences entre les sept notes non altérées d'une octave. La fréquence du *do* est utilisée comme référence.

do	**ré**	**mi**	**fa**	**sol**	**la**	**si**
1	9/8	5/4	4/3	3/2	5/3	15/8

Tableau 5.2 : Rapport de fréquence entre les notes non altérées.

Le rapport de fréquence donné par le tableau 5.2 permet de définir la fréquence de chacune des notes. Cette opération est réalisée à partir de la fréquence du diapason qui

équivaut à la note « la » de la troisième octave. La charte de fréquences des notes des octaves 0 à 9 est illustrée à la figure 5.1.

Notes		Octaves									
		0	1	2	3	4	5	6	7	8	9
Do	C	32.7 Hz	65 Hz	131 Hz	262 Hz	523 Hz	1 046.5 Hz	2 093 Hz	4 186 Hz	8 372 Hz	16 744 Hz
Reb	Db	34.6 Hz	69 Hz	139 Hz	277 Hz	554 Hz	1 109 Hz	2 217 Hz	4 435 Hz	8 870 Hz	17 740 Hz
Re	D	36.7 Hz	74 Hz	147 Hz	294 Hz	587 Hz	1 175 Hz	2 349 Hz	4 698 Hz	9 396 Hz	18 792 Hz
Mib	Eb	38.9 Hz	78 Hz	156 Hz	311 Hz	622 Hz	1 244.5 Hz	2 489 Hz	4 978 Hz	9 956 Hz	19 912 Hz
Mi	E	41.2 Hz	83 Hz	165 Hz	330 Hz	659 Hz	1 318.5 Hz	2 637 Hz	5 274 Hz	10 548 Hz	21 098 Hz
Fa	F	43.6 Hz	87 Hz	175 Hz	349 Hz	698.5 Hz	1 397 Hz	2 794 Hz	5 588 Hz	11 176 Hz	
Fa #	F#	46.2 Hz	92.5 Hz	185 Hz	370 Hz	740 Hz	1 480 Hz	2 960 Hz	5 920 Hz	11 840 Hz	
Sol	G	49.0 Hz	98 Hz	196 Hz	392 Hz	784 Hz	1 568 Hz	3 136 Hz	6 272 Hz	12 544 Hz	
Sol #	G#	51.9 Hz	104 Hz	208 Hz	415 Hz	831 Hz	1 661 Hz	3 322 Hz	6 645 Hz	13 290 Hz	
La	A	55.0 Hz	110 Hz	220 Hz	440 Hz	880 Hz	1 760 Hz	3 520 Hz	7 040 Hz	14 080 Hz	
Sib	Bb	58.0 Hz	117 Hz	233 Hz	466 Hz	932 Hz	1 865 Hz	3 729 Hz	7 458 Hz	14 918 Hz	
Si	B	62.0 Hz	123 Hz	247 Hz	494 Hz	988 Hz	1 975 Hz	3 951 Hz	7 902 Hz	15 804 Hz	

Figure 5.1 : Fréquence des notes des octaves 0 à 9.

Il existe une relation directe entre la fréquence d'une note et la forme de l'onde de laquelle elle résulte. Une basse fréquence génère un signal dont les oscillations sont peu rapprochées et vice versa.

a. Hypothèse

Considérant un domaine d'information constitué de notes de musique distinctes, le système proposé est potentiellement en mesure de grouper les notes d'une même octave ou les notes d'octaves adjacentes puisque ces notes possèdent une fréquence et une forme similaires.

b. Banque de données

La banque de données est constituée de 72 fichiers monophoniques représentant chacun une note distincte. Afin de réduire les traitements nécessaires à la normalisation des données, la durée des fichiers a été fixée à trois secondes. Le tableau 5.3 liste les fichiers où le chiffre suivant la note représente l'octave.

do	ré b	ré	mi b	mi	fa	fa #	sol	sol #	la	si b	si
C2	Db2	D2	Eb2	E2	F2	F#2	G2	G#2	A2	Bb2	B2
C3	Db3	D3	Eb3	E3	F3	F#3	G3	G#3	A3	Bb3	B3
C4	Db4	D4	Eb4	E4	F4	F#4	G4	G#4	A4	Bb4	B4
C5	Db5	D5	Eb5	E5	F5	F#5	G5	G#5	A5	Bb5	B5
C6	Db6	D6	Eb6	E6	F6	F#6	G6	G#6	A6	Bb6	B6
C7	Db7	D7	Eb7	E7	F7	F#7	G7	G#7	A7	Bb7	B7

Tableau 5.3 : Fichiers sources de l'expérimentation 1.

La banque de données a été créée à l'aide d'un système MIDI de manière à réduire le bruit contenu dans le signal. Les fichiers résultants ont ensuite été exportés en format WAVE à l'aide d'un utilitaire de conversion.

5.1.1 L'analyse spatiale à des fins de classification de notes

Dans le premier volet de l'expérimentation, les paires formées de l'amplitude et de la demi-période (chapitre 4.1.2) ont été utilisées comme caractère descriptif.

Les variables σ, β, n et ρ représentent respectivement le niveau du lissage, l'intervalle de projection, la taille des n-grammes et le seuil de tolérance du classifieur.

Prétraitements et paramètres fixes :

- L'onde sonore a été lissée à l'aide d'un filtre gaussien de paramètre $\sigma = 5$;
- L'intervalle de projection β a été défini par [0,16];
- La taille des n-grammes a été fixée à 2;
- Les hapax ont été supprimés.

Le choix de ces valeurs a été guidé par l'expérience acquise lors de travaux antérieurs [Biskri et al. 2006][Meunier et al 2002] et par les connaissances tirées de la littérature [Grefenstette, 1995] [Damashek, 1995] [Gonzalez, Woods, 1992].

Un lissage trop important engendre l'élimination de données susceptibles d'être porteuses d'information. Un filtre gaussien de paramètre σ égal à 5 permet donc de réduire ce risque tout en supprimant le bruit résultant du passage du format MIDI au format WAVE.

L'intervalle de projection β a été défini de manière à réduire l'alphabet tout en permettant une représentation jugée acceptable.

La taille des n-grammes a été fixée en continuité avec les travaux effectués dans le passé [Biskri].

Le paramètre de vigilance ρ du classifieur :

La qualité de la classification est largement affectée par le classifieur utilisé. Le paramètre de vigilance ρ influence la tolérance d'ART. Les résultats obtenus varient considérablement selon sa valeur. Il demeure difficile de définir précisément la valeur à considérer pour obtenir une classification de qualité. Nous avons donc fait varier ce paramètre de vigilance.

a. Résultats de la classification I

Pour la classification I, le paramètre de vigilance ρ à été fixé arbitrairement à 0.1. Le tableau 5.4 donne la répartition des fichiers au sein des classes générées suite au traitement.

Classes	Fichiers
1	Eb2, E2, F2, F#2, G2, G#2, Bb2, B2, C3, Db3, D3, Eb3, E3, F3, F#3, G3, G#3, A3, Bb3, B3, C4, Db4, D4, Eb4, F4, F#4, G4, A4, Bb4, B4, C5, Db5, E5, F#5, G5, G#5, A5, B5, Bb5, C6, Db6, D6, Eb6, E6, F6, F#6, G6, G#6, A6, Bb6, Db7, D7, Eb7, F#7, G7, G#7, A7, B7, Bb7
2	C2
3	A2, E4, G#4, D5, Eb5, F5, B6, C7, E7, F7
4	Db2
5	D2

Tableau 5.4 : Répartition des fichiers pour la classification I.

Un aspect de la classification paraît intéressant : les classes 2, 4 et 5 sont formées d'un seul fichier contenant respectivement les notes *do*, *ré* bémol et *ré*. Ces notes sont toutes jouées à l'octave 2 et représentent les notes les plus graves contenues dans la banque de données.

59 fichiers appartiennent à la classe 1. Ce nombre équivaut à près de 82 % de la totalité des fichiers. Cette particularité donne un caractère peu attrayant à la classification.

b. Résultats de la classification II

Afin de mesurer l'impact du paramètre ρ, la classification II a été réalisée avec un paramètre de vigilance ρ = 0.01. Les classes résultantes sont données au tableau 5.5.

Classes	Fichiers
1	Bb2
2	C2, D2, Eb2, E2, F2, F#2, G2, G#2 A2, Bb2, C3, Db3, Bb4, C5
3	B7
4	B4, D5, Bb5
5	G#5
6	G7, A7, Bb7
7	D4, Eb4, E4, F4, G5, A5, Bb5
8	F#4, F#5
9	E3, F3, G#3, A4
10	E5
11	Bb3, B3, C4, Db4, G#7
12	C6, G#6, A6, Bb6
13	G4, G#4, E7, F7, F#7
14	B5, Db6
15	D6, Eb6, E6, F6, G6, B6, C7, D7, Db7
16	F#6, Eb7

Tableau 5.5 : Répartition des fichiers pour la classification II.

Même si théoriquement ART est plus sélectif lorsque le paramètre de vigilance ρ se rapproche de 1 [Juan-Manuel et al, 1999], le nombre de classes produites est plus grand que lors de la classification I.

Les classes 6 et 12 sont les plus intéressantes car les notes qu'elles contiennent appartiennent à la même octave.

Les classes 4, 7, 8, 9, 14, 15 et 16 sont également satisfaisantes. Elles sont composées de notes de deux octaves voisines.

Les classes 2 et 13 peuvent difficilement être interprétées car des octaves distantes sont représentées.

c. Observations

La qualité des classifications varie considérablement selon la valeur du paramètre de vigilance introduit. La classification I souligne l'impact du choix des paramètres ce qui peut être considéré comme une certaine faiblesse.

La classification II valide l'hypothèse posée. Plusieurs classes produites sont composées de notes d'une même octave ou de deux octaves adjacentes.

5.1.2 L'analyse spectrale à des fins de classification de notes

Le second volet de l'expérimentation introduit l'analyse fréquentielle. Les caractères descriptifs équivalent aux bandes de fréquences des fréquences dominantes du signal (chapitre 4.1.3).

Deux classifications ont été effectuées afin de comparer les résultats obtenus à ceux de l'analyse spatiale.

Prétraitements et paramètres :

- Les 3 fréquences les plus importantes ont été considérées;
- Le spectre de fréquence a été subdivisé en 31 bandes;
- La taille de la fenêtre de la transformée de Fourier a été fixée à 32 768;
- La taille des n-grammes a été fixée à 2;
- Les hapax ont été supprimés.

Dû à la simplicité des fichiers utilisés, le choix de considérer une seule fréquence aurait sans doute produit des résultats convenables. Les fichiers sources sont constitués d'une seule note définie par une fréquence particulière (tableau 5.3). Cet environnement simpliste n'est cependant pas représentatif de la réalité. Les banques de données sont généralement composées de fichiers dont le contenu est un amalgame de notes concurrentes. Dans la majorité des cas, il est préférable de tenir compte de plus d'une fréquence [Cavicchi, 2000]. Pour cette raison, trois fréquences dominantes ont été considérées.

Le spectre du signal traité est donné par la fréquence d'échantillonnage utilisée pour numériser le son. 44 100 fréquences composent le spectre des fichiers sources. Celles-ci ont été subdivisées en 31 bandes afin de maintenir le lexique à un seuil raisonnable.

La compression générée lors de l'opération de la transformée de Fourier a été limitée par l'utilisation d'une fenêtre de 32 octets et par le découpage du spectre en 31 bandes de fréquences.

La taille des n-grammes a été déterminée en relation avec les résultats obtenus lors d'expérimentations antérieures [Biskri et al, 2006].

Le paramètre de vigilance ρ du classifieur :

Afin d'être fidèle aux classifications I et II, la valeur du paramètre de vigilance a été variée. Le paramètre de vigilance ρ a été fixé à 0.1 pour la classification III et à 0.01 pour la classification IV.

a. Résultats de la classification III

Le tableau 5.6 donne les classes produites lors de la classification III où $\rho = 0.01$.

37 classes ont été créées dont la majorité forme un ensemble cohérent. Les classes 1, 3, 4, 5, 6, 12, 14, 16, 19, 24, 26, 28, 30, 33, 34 et 36 sont composées de notes d'une même octave tandis que les classes 7, 8, 15 et 20 sont formées de notes réparties sur deux octaves voisines. Ces classes valident l'hypothèse selon laquelle le système permet de grouper les notes d'une même octave ou d'octaves voisines.

Lorsque l'on étudie le contenu de la classe 8, on constate que les notes sont successives et constituent une suite logique. En fait, la note *do* précède la note *si* et par conséquent les ondes associées aux notes Bb_i et C_{i+1} sont plus ressemblantes que celles associées aux notes situées aux extrémités d'une même octave. La présence de notes successives est un bon critère d'évaluation.

Classes	Fichiers
1	Eb7, Bb7
2	G#2
3	A3, Bb3
4	A4, Bb4, B4
5	G#5, A5, Bb5, B5
6	A6, Bb6, B6
7	B2, D3
8	B3, C4
9	C2
10	C3
11	F3, Db5, Eb5
12	C7, Db7, D7
13	G7
14	D4, Eb4
15	D5, C6
16	Db6, D6, Eb6
17	Db3
18	Db4
19	E2, F2, F#2, G2
20	E3, F#3, F#4
21	G5
22	E5
23	E6
24	E7, F#7
25	Eb3
26	F6, F#6, G6
27	F7
28	F5, F#5
29	G3
30	E4, G4
31	G#3
32	G#4
33	G#7, A7, B7
34	A2, Bb2
35	C5
36	Db2, D2, Eb2
37	F4

Tableau 5.6 : Répartition des classes pour la classification III.

Les classes 3, 4, 5, 8, 12, 14, 16, 19, 26, 28, 30, 33, 34 et 36 contiennent des notes successives. Cet ensemble représente plus du tiers des classes produites. Parmi celles-ci, les classes 5 et 19 sont les plus intéressantes dues à leur nombre de membres.

16 classes (2, 9, 10, 13, 17, 18, 21, 22, 23, 25, 27, 29, 31, 32, 35 et 37) sont formées d'un seul ficher. Au premier regard, ce nombre peut sembler élevé. Cependant, considérant qu'aucune note n'est reproduite, ce nombre est représentatif de la diversité du domaine d'information.

La classe 11 est composée de fichiers contenant des notes situées aux octaves 3 et 5. Aucune signification ne peut être instinctivement attribuée à ce regroupement. Par conséquent, cette classe est considérée comme bruitée.

Ces résultats suggèrent que le modèle d'analyse fondé sur le contenu fréquentiel de l'onde sonore est moins sensible à la variation du paramètre de vigilance.

b. Résultats de la classification IV

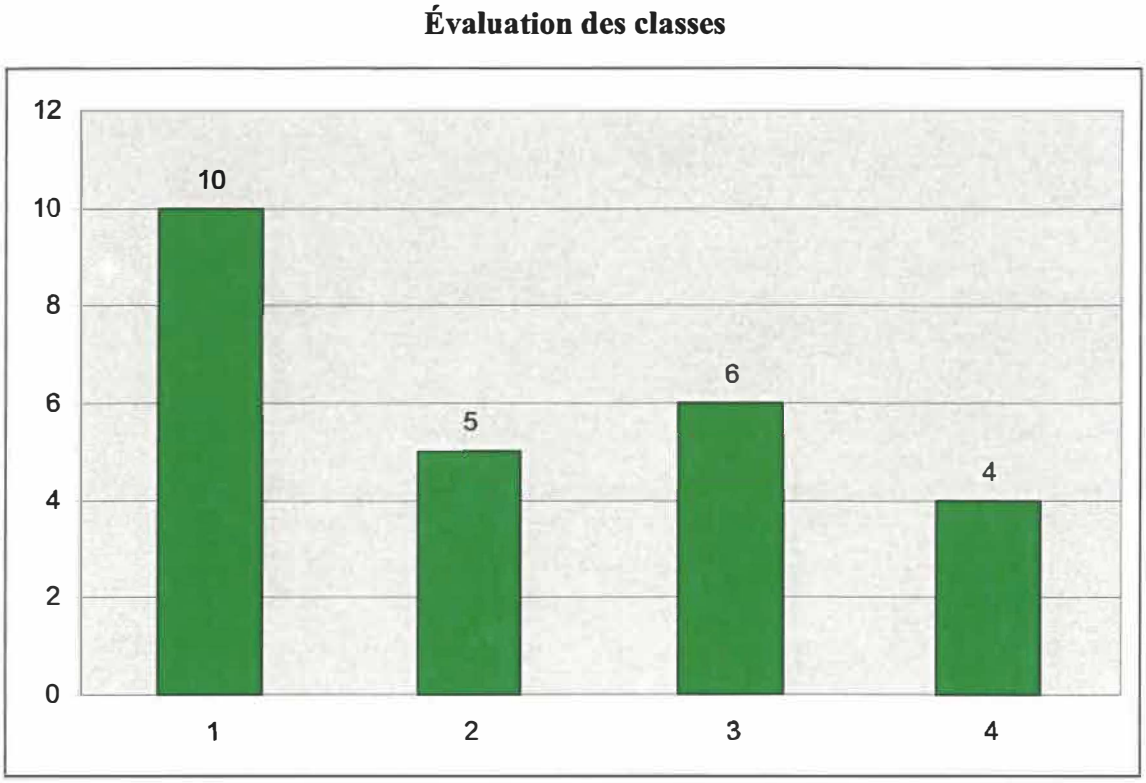
La valeur de ρ a été fixée à 0.01. Le nombre de classes produites a diminué à 25. Ces classes sont présentées au tableau 5.7.

Classes	Fichiers
1	Eb7, Bb7
2	G#2, A2, Bb2, A3
3	G#4, A4, Bb4, B4
4	G#5, A5, Bb5, B5,
5	G#6, A6, Bb6, B6
6	B2, Db3, D3
7	Eb3, B3
8	G#3, Bb3
9	C2
10	C3, F3, C4
11	C5, D5, Eb5 C6, Eb6
12	C7, Db7, D7
13	Db2, D2, Eb2, G7
14	Db4, D4, Eb4
15	Db6, D6
16	Db5
17	E2, F2, F#2, G2
18	E3, F#3, F#4, G4
19	E4, F4, G5
20	E5, F5, F#5
21	E6, F6, F#6, G6
22	E7, F#7
23	F7
24	G3
25	G#7, A7, B7

Tableau 5.7 : Répartition des fichiers lors de la quatrième classification.

Tout comme lorsque la valeur du paramètre de vigilance ρ a été fixée à 0.1, la classification IV donne des résultats intéressants.

Les classes 1, 3, 4, 5, 7, 8, 12, 14, 15, 17, 20, 21, 22 et 25 sont constituées de notes d'une même octave et les classes 2, 6, 10, 11, 18, et 19 contiennent des notes de deux octaves voisines. Parmi ces classes, 10 sont composées de notes successives et 5 de plus de 50% de notes successives. La figure 5.2 illustre cette répartition.



Série	Description	Classes
1	100% de notes successives.	3, 4, 5, 6, 12, 14, 15, 17, 20, 21
2	50 % et plus de notes successives.	2, 13, 18, 19, 25
3	50 % et plus de notes distantes.	1, 7, 8, 10, 11, 22
4	Une seule note.	9, 16, 23, 24

Figure 5.2 : Histogramme de l'évaluation des résultats.

Parmi les classes de la série 1, les classes 3, 4 et 5 sont particulièrement étonnantes. Leurs membres équivalent aux mêmes séries de notes mais à des octaves différentes. La série 3

ne peut être interprétée comme étant le regroupement de classes bruitées puisque les notes des classes 8 et 22 sont peu distantes.

c. Observations

L'analyse fréquentielle permet de regrouper un nombre considérable de notes avoisinantes. L'effet de compression généré par le groupement de fréquences a un lien direct avec ce phénomène. La figure 5.3 illustre une portion du spectre des notes *do* (262 hertz), *mi* (330 hertz) et *la* (440 hertz) de la troisième octave. La fréquence dominante de chacune de ces notes est clairement définie. Une compression supérieure à 176 hertz se traduirait par le regroupement de ces notes même si elles sont distantes l'une de l'autre.

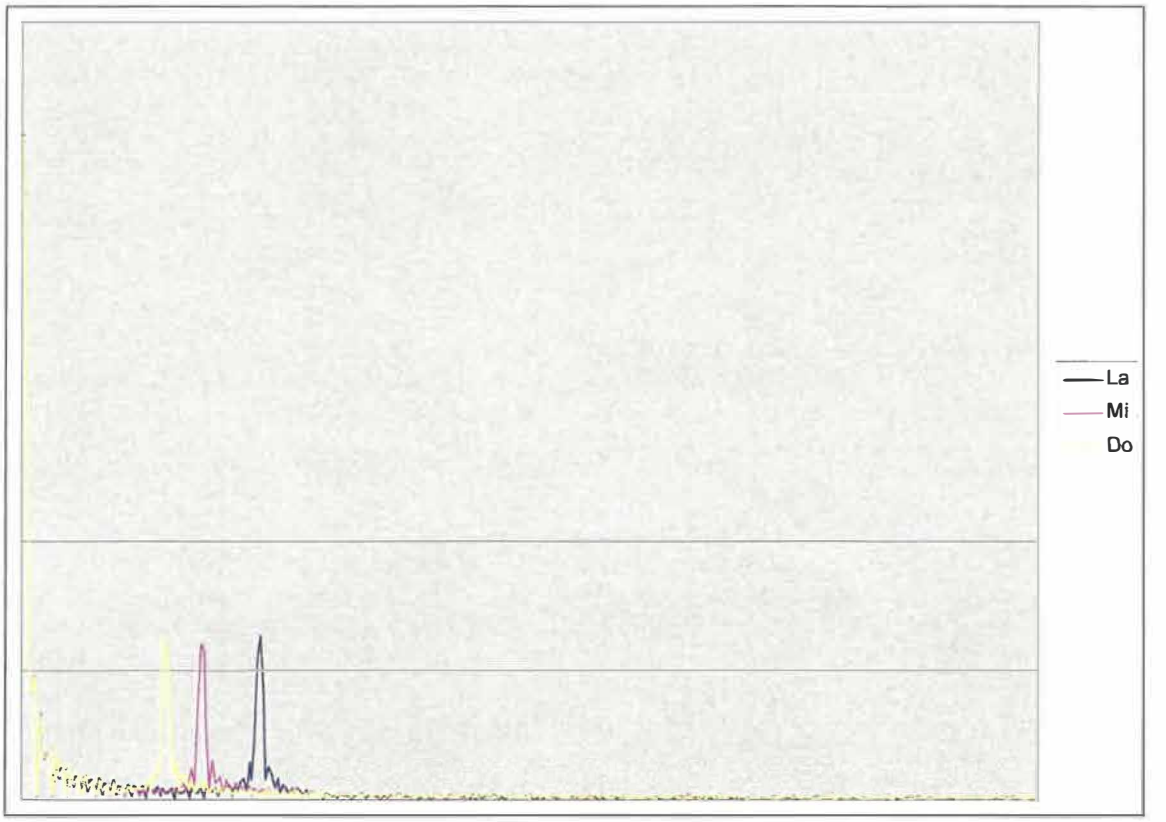


Figure 5.3 : Visualisation de l'onde sonore de la note *la* de la troisième octave.

Les harmoniques peuvent créer des associations plus abstraites. La figure 5.4 illustre la forme de l'onde sonore de la note A3 donnée par un piano. Les harmoniques contenues dans ce signal sont également présentes dans l'ensemble des notes jouées par le même instrument. Cet effet de catégorisation peut être favorable à la reconnaissance de la source d'un son.

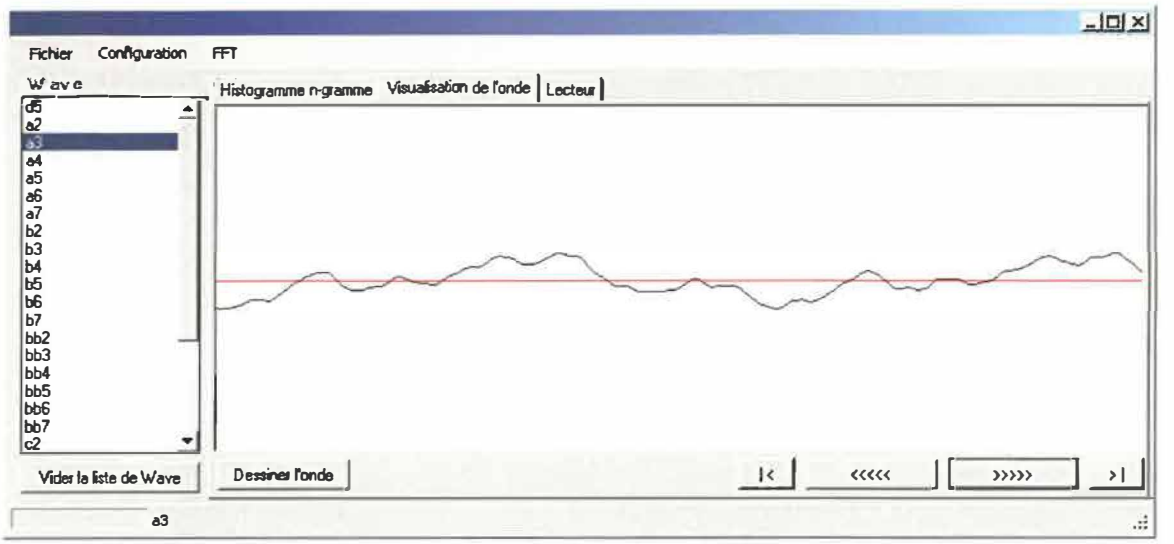


Figure 5.4 : Visualisation de l'onde sonore de la note *la* de la troisième octave.

5.1.3 Sommaire de l'expérimentation 1

L'analyse basée sur la forme de l'onde est particulièrement sensible à la variation des paramètres. La qualité des classifications produites diffère considérablement selon la valeur du paramètre de vigilance introduit.

Lors de cette expérimentation, il a été démontré que le système proposé était en mesure de créer des associations pertinentes, particulièrement l'analyse fréquentielle qui génère un nombre intéressant de classes contenant uniquement des notes avoisinantes. Ces observations suggèrent que l'analyse fréquentielle est plus performante que l'analyse spatiale.

5.2. Classification de séries de notes de musique d'une même octave

Suite à la première expérimentation, le comportement du système lorsque les fichiers sources sont composés de plusieurs notes a été testé. Des fichiers contenant une série de notes d'une même octave ont été utilisés.

a. Hypothèse

Nous avons posé l'hypothèse que les fichiers contenant des mélodies jouées aux mêmes octaves seraient regroupés.

b. Banque de données

La banque de données utilisée pour la deuxième expérimentation est constituée de 27 fichiers monophoniques contenant une suite de 16 notes.

Neuf mélodies différentes sont jouées à trois octaves. Afin de réduire les traitements nécessaires à la normalisation des données, la durée des fichiers a été fixée à dix secondes.

Tout comme à la première expérimentation, la banque de données a été créée à l'aide d'un système MIDI et exportée en format WAVE. Le tableau 5.8 liste les fichiers utilisés.

Mélodie	Fichier	Notes
1	1	A2, F#2, E2, F#2, A2, F#2, E2, F#2, A2, F#2, E2, F#2, A2, F#2, E2, F#2
	2	A3, F#3, E3, F#3, A3, F#3, E3, F#3, A3, F#3, E3, F#3, A3, F#3, E3, F#3
	3	A6, F#6, E6, F#6, A6, F#6, E6, F#6, A6, F#6, E6, F#6, A6, F#6, E6, F#6
2	4	D2, D2, D2, C2, E2, D2, D2, D2, C2, E2, D2, D2, D2, C2, E2, D2
	5	D3, D3, D3, C3, E3, D3, D3, D3, C3, E3, D3, D3, D3, C3, E3, D3
	6	D6, D6, D6, C6, E6, D6, D6, D6, C6, E6, D6, D6, D6, C6, E6, D6
3	7	G2, G2, G2, F2, A2, G2, G2, G2, F2, A2, G2, G2, G2, F2, A2, G2
	8	G3, G3, G3, F3, A3, G3, G3, G3, F3, A3, G3, G3, G3, F3, A3, G3
	9	G6, G6, G6, F6, A6, G6, G6, G6, F6, A6, G6, G6, G6, F6, A6, G6
4	10	G2, B2, D2, C2, G2, B2, D2, C2, G2, B2, D2, C2, G2, B2, D2, C2
	11	G3, B3, D3, C3, G3, B3, D3, C3, G3, B3, D3, C3, G3, B3, D3, C3
	12	G6, B6, D6, C6, G6, B6, D6, C6, G6, B6, D6, C6, G6, B6, D6, C6
5	13	E2, F2, G2, A2, G2, E2, F2, G2, A2, G2, E2, F2, G2, A2, G2, E2
	14	E3, F3, G3, A3, G3, E3, F3, G3, A3, G3, E3, F3, G3, A3, G3, E3
	15	E6, F6, G6, A6, G6, E6, F6, G6, A6, G6, E6, F6, G6, A6, G6, E6
6	16	A2, F#2, F#2, F#2, A2, F#2, F#2, F#2, A2, F#2, F#2, F#2, A2, F#2, F#2, F#2
	17	A3, F#3, F#3, F#3, A3, F#3, F#3, F#3, A3, F#3, F#3, F#3, A3, F#3, F#3, F#3
	18	A6, F#6, F#6, F#6, A6, F#6, F#6, F#6, A6, F#6, F#6, F#6, A6, F#6, F#6, F#6
7	19	Bb5, Eb5, F#5, Bb5, G#5, F#5, Eb5, F#5, Bb5, G#5, F#5, Eb5, F#5, Bb5
	20	Bb7, Eb7, F#7, Bb7, G#7, F#7, Eb7, F#7, Bb7, G#7, F#7, Eb7, F#7, Bb7
	21	Bb8, Eb8, F#8, Bb8, G#8, F#8, Eb8, F#8, Bb,8 G#8, F#8, Eb8, F#8, Bb8
8	22	D1, G1, G1, F1, E1, D1, G1, G1, F1, E1, D1, G1, G1, F1, E1, D1
	23	D2, G2, G2, F2, E2, D2, G2, G2, F2, E2, D2, G2, G2, F2, E2, D2
	24	D4, G4, G4, F4, E4, D4, G4, G4, F4, E4, D4, G4, G4, F4, E4, D4
9	25	C7, C7, Db7, D7, E7, F7, E7, F7, C7, C7, Db7, D7, E7, F7, E7, F7
	26	C8, C8, Db8, D8, E8, F8, E8, F8, C8, C8, Db8, D8, E8, F8, E8, F8
	27	C9, C9, Db9, D9, E9, F9, E9, F9, C9, C9, Db9, D9, E9, F9, E9, F9

Tableau 5.8 : Fichiers sources de la deuxième expérimentation.

5.2.1 L'analyse spatiale à des fins de classification de série de notes d'une même octave

Le choix des paramètres a été effectué en relation avec les résultats observés lors de l'expérimentation précédente.

Prétraitements et paramètres fixes :

- L'onde sonore a été lissée à l'aide d'un filtre gaussien de paramètre $\sigma = 5$;
- L'intervalle de projection β est [0,16];
- La taille des n-grammes a été fixée à 2;
- Le paramètre de vigilance ρ a été fixé à 0.1 pour la classification V;
- Le paramètre de vigilance ρ a été fixé à 0.01 pour la classification VI;
- Les hapax ont été supprimés.

a. Résultats de la classification V

Seulement deux classes sont produites. La quasi-totalité des fichiers a été groupé dans la classe 2. Seule la mélodie la plus aiguë a été isolée dans la classe numéro 1. Le tableau 5.9 liste la répartition des fichiers de la classification V.

Classes	Fichiers
1	9-o9
2	1-o2, 1-o3, 1-o6, 2-o2, 2-o3, 2-o6, 3-o2, 3-o3, 3-o6, 4-o2, 4-o3, 4-o6, 5-o2, 5-o3, 5-o6, 6-o2, 6-o3, 6-o6, 7-o5, 7-o7, 7-o8, 8-o1, 8-o2, 8-o4, 9-o7, 9-o8

Tableau 5.9 : Répartition des fichiers lors de la classification V.

b. Résultats de la classification VI

Lors de la classification VI, cinq classes sont obtenues (tableau 5.10). Les six fichiers contenant des mélodies jouées à la sixième octave ont été regroupés dans la classe 2 avec deux fichiers contenant des mélodies jouées à la septième octave. Cette classe vérifie l'hypothèse selon laquelle notre système est en mesure de rassembler les fichiers contenant des mélodies interprétées à la même octave.

Classes	Fichiers
1	9-o9
2	1-o6, 2-o6, 3-o6, 4-o6, 5-o6, 6-o6, 7-o7, 9-o7,
3	1-o2, 1-o3, 2-o2, 2-o3, 3-o2, 3-o3, 4-o2, 4-o3, 5-o2, 5-o3, 6-o2, 6-o3, 7-o5, 8-o1, 8-o2, 8-o4
4	7-o8
5	9-o8

Tableau 5.10 : Répartition des fichiers lors de la classification VI.

Quoique les octaves 6 et 7 soient adjacentes, il aurait été souhaitable que la classe 2 soit fragmentée en deux de manière à ce que seuls les fichiers propres à une octave soient contenus à l'intérieur d'une classe. Cette observation s'applique également à la classe 3 où 81 % des fichiers représentent des mélodies des octaves 2 et 3.

Les classes 1, 4 et 5 sont formées d'un seul fichier. Ces singletons représentent les mélodies jouées aux octaves les plus aiguës.

c. Observations

Le comportement du système, lorsque l'input est une suite de notes, est analogue à celui observé lorsque l'input est constitué uniquement de notes distinctes. Lorsque le paramètre de vigilance égale 0.1, un nombre restreint de classes est généré.

5.2.2 L'analyse spectrale à des fins de classification de série de notes d'une même octave

Prétraitements et paramètres fixes :

- Les 3 fréquences les plus importantes ont été considérées;
- Le spectre de fréquences a été subdivisé en 31 bandes;
- La taille de la fenêtre de la transformée de Fourier a été figée à 32 768;
- La taille des n-grammes a été fixée à 2;
- Le paramètre de vigilance ρ a été fixé à 0.1 pour la classification VII;
- Le paramètre de vigilance ρ a été fixé à 0.01 pour la classification VIII;
- Les hapax ont été supprimés.

a. Résultats de la classification VII

18 classes sont produites. Ce nombre est largement supérieur à celui obtenu lors de l'analyse spatiale pour la même valeur de ρ.

Les classes 1, 5, 9, 11, 12, 13, 14, 15, 16, 17 et 18 contiennent un seul fichier. Ces singletons constituent plus de la majorité des classes produites. Les autres classes sont formées de deux fichiers à l'exception de la classe 2 qui en contient quatre. Toutes les classes comptant plus d'un membre confirment l'hypothèse posée.

La liste des classes de la classification VII est donnée par le tableau 5.11. Aucune classe produite n'est constituée de mélodies jouées à des octaves différentes.

Classes	Fichiers
1	9-o9
2	1-o2, 3-o2, 4-o2, 6-o2
3	1-o3, 5-o3
4	1-o6, 6-o6
5	2-o2
6	2-o3, 4-o3
7	2-o6, 4-o6
8	3-o6, 5-o6
9	5-o2
10	3-o3, 6-o3
11	7-o5
12	7-o7
13	7-o8
14	8-o1
15	8-o2
16	8-o4
17	9-o7
18	9-o8

Tableau 5.11 : Répartition des fichiers lors de la classification VII.

b. Résultats de la classification VIII

Trois classes parmi les neuf générées contiennent des mélodies jouées à la même octave et deux à des octaves voisines. La répartition des fichiers est donnée au tableau 5.12.

Classes	Fichiers
1	9-o9
2	1-o2, 1-o3, 3-o2, 3-o3, 4-o2, 5-o2, 5-o3, 6-o2, 6-o3
3	1-o6, 2-o6, 3-o6, 5-o6, 6-o6
4	2-o2, 2-o3, 4-o3
5	4-o6
6	7-o5, 8-o1, 8-o4
7	7-o7, 9-o7
8	7-o8, 9-o8
9	8-o2

Tableau 5.12 : Répartition des fichiers lors de la classification VIII.

Les classes 2, 4 et 6 sont composées de mélodies jouées à deux octaves adjacentes. Parmi celles-ci, seule la classe 6 ne peut être interprétée de manière cohérente.

Les fichiers des mélodies jouées à la sixième octave sont presque entièrement rassemblés dans la classe numéro 3.

Contrairement à la classification VII, seules deux classes contiennent un seul membre.

c. Observations

Les résultats de la classification VIII démontrent la capacité de notre système à différencier les notes jouées à des octaves différentes.

5.2.3 Sommaire de l'expérimentation 2

Il semble exister une relation constante entre les mélodies jouées aux octaves 2 et 3. Les deux modes d'analyse ont créé des associations entre les fichiers contenant ces mélodies.

Dans le même ordre d'idée, les deux modes d'analyse ont été en mesure de rassembler la quasi-totalité des fichiers contenant des mélodies jouées à la sixième octave.

Les fichiers utilisés étaient tous créés à partir d'un système MIDI. Cette particularité a favorisé la reconnaissance de la forme de l'onde.

Les modèles basés sur les caractéristiques spatiales et spectrales sont en mesure d'effectuer une distinction plus ou moins précise des octaves. Bien que cette propriété puisse être désirable à des fins d'analyses, elle peut engendrer des lacunes lorsqu'il s'agit de reconnaissance de mélodies. En effet, la même mélodie jouée à deux octaves différentes ne pourrait être jugée comme similaire. Cette distinction résulte de l'altération de la forme et du contenu fréquentiel de l'onde attribuable au changement de l'octave. Ce facteur pourrait néanmoins être contourné par une méthode de factorisation permettant de rapporter l'ensemble des notes sur une octave référence.

5.3. Classification de segments de musiques polyphoniques

La musique polyphonique constitue la majorité des documents recherchés sur le web. Par conséquent, il devient primordial qu'un système de classification de fichiers sonores soit en mesure de traiter ce type de document. Nous avons donc testé le comportement du système proposé lorsque le domaine d'information est formé de documents polyphoniques.

a. Hypothèse

Nous avons posé l'hypothèse que des segments d'une même musique seraient regroupés.

b. Banque de données

La banque de données a été formée à partir de quatre pièces d'artistes et de genres différents. Le tableau 5.13 présente ces pièces.

Artiste	Pièce	Genre
Hot Snake	Retrofit	Rock
Elvis	Houng Dog	Rock'n Roll
Calexico	Tres Avisos	Folk
Red Hot Chili Pepper	Higher Ground	Funk

Tableau 5.13 : Pièces de musiques polyphoniques

Trois segments d'une durée de cinq secondes ont été extraits à chacune des pièces. Le tableau 5.14 donne la liste des segments utilisés comme source de données.

Fichiers
1 - hot snake - retrofit
2 - Elvis - houng dog
3 - hot snake - retrofit
4 - Calexico - Tres Aviso
5 - hot snake - retrofit
6 - Elvis – houng dog
7 - Calexico - Tres Avisos
8 - Elvis houng dog
9 - Red hot - Higher Ground
10 - Red hot - Higher Ground
11 - Calexico - Tres Avisos
12 - Red hot - Higher Ground

Tableau 5.14 : Pièces de musiques polyphoniques de l'expérimentation 3.

Un chiffre a été ajouté au début du nom du fichier afin d'exercer un contrôle sur l'ordonnancement des fichiers.

5.3.1 L'analyse spatiale à des fins de classification de segments de musiques polyphoniques

Prétraitements et paramètres fixes :

- L'onde sonore a été lissée à l'aide d'un filtre gaussien de paramètre $\sigma = 5$;
- L'intervalle de projection β est [0,16];
- La taille des n-grammes a été fixée à 2;
- Le paramètre de vigilance ρ a été fixé à 0.1 pour la classification IX;
- Le paramètre de vigilance ρ a été fixé à 0.01 pour la classification X;
- Les hapax ont été supprimés.

a. Résultats de la classification IX

La complexité de l'onde sonore augmente le nombre de paires formées de l'amplitude et de la demi-période. Ce facteur génère une signature quasi distincte pour chacun des fichiers. De cette façon plusieurs segments ont été isolés. Le tableau 5.15 donne la répartition des fichiers lorsque la valeur de ρ a été fixée à 0.1.

Classes	Fichiers
1	1 - hot snake - retrofit
2	3 - hot snake - retrofit
3	4 - Calexico - Tres Avisos, 7 - Calexico - Tres Avisos
4	5 - hot snake - retrofit
5	6 - Elvis - houng dog
6	11 - Calexico - Tres Avisos
7	10 - Red hot - Higher Ground, 12 - Red hot - Higher Ground
8	2 - Elvis - houng dog
9	8 - Elvis - houng dog
10	9 - Red hot - Higher Ground

Tableau 5.15 : Répartition des fichiers lors de la classification IX.

Un total de 10 classes est obtenu. Malgré la complexité de la forme de l'onde sonore de chacun des extraits, le système a été en mesure de regrouper certains segments.

Les deux classes contenant plus d'un fichier regroupent deux segments d'une même pièce. Les classes 3 et 7 contiennent respectivement des segments de musique folk et funk.

Toutes les autres classes produites contiennent un seul membre.

b. Résultats de la classification X

Les résultats de la classification X diffèrent de ceux obtenus précédemment lorsque le paramètre de vigilance a été fixé à 0.01. Le classifieur a regroupé un plus grand nombre de fichiers à l'intérieur des classes de similarité. La répartition des fichiers est donnée au tableau 5.16

Classes	Fichiers
1	11 - Calexico - Tres Avisos
2	1 - hot snake - retrofit, 3 - hot snake - retrofit, 5 - hot snake - retrofit, 4 - Calexico - Tres Avisos, 7 - Calexico - Tres Avisos
3	2 - Elvis - houng dog, 6 - Elvis - houng dog, 8 - Elvis - houng dog, 9 - Red hot - Higher Ground, 10 - Red hot - Higher Ground, 12 - Red hot - Higher Ground

Tableau 5.16 : Répartition des fichiers lors de la classification X.

Seulement trois classes sont produites. La classe 1 contient un seul segment tandis que les classes 2 et 3 sont formées de segments de deux pièces différentes.

À l'exception de la pièce folk où un segment a été isolé, on remarque que l'ensemble des segments d'une même pièce est contenu dans une seule classe.

Lorsque l'on compare le segment isolé avec les autres segments de la pièce qu'il représente, on constate qu'il est caractérisé par l'absence de violon.

c. Observations

La figure 5.4 illustre l'évaluation des classifications considérant :

i. Les classes formées uniquement de segments d'une même pièce comme pertinentes;

ii. Les classes contenant majoritairement les segments d'une même pièce comme peu bruitées ;

iii. Les classes contenant moins de 50 % de segments d'une même pièce comme bruitées.

iv. Les classes contenant un seul membre comme neutres.

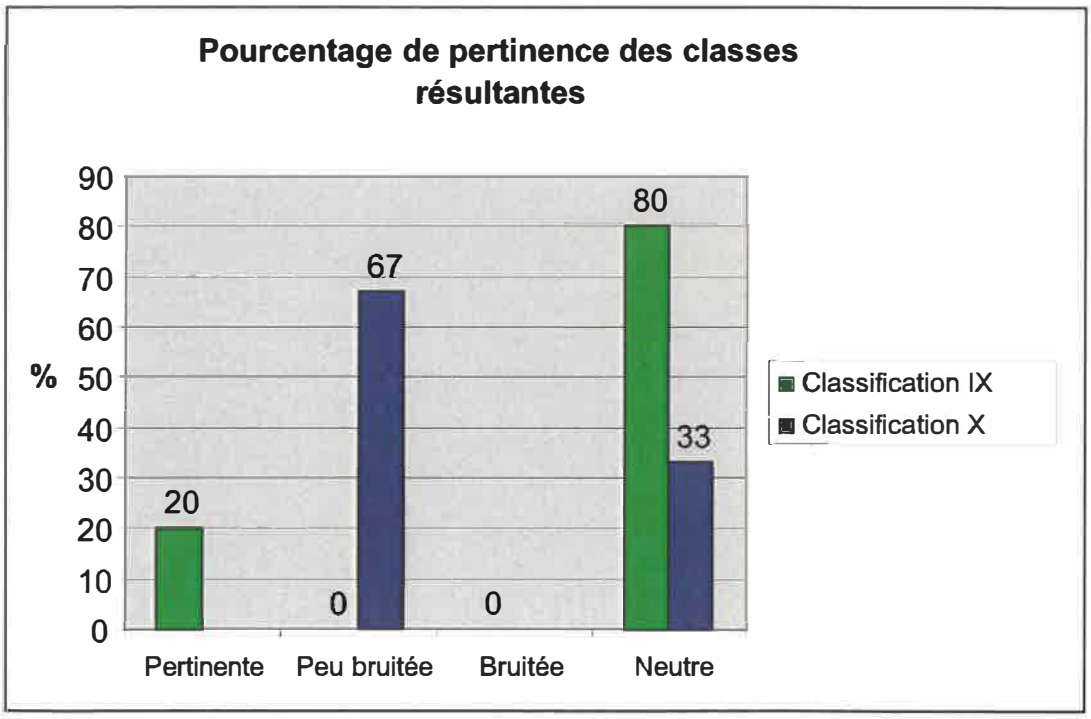


Figure 5.4 : Histogramme de l'évaluation des classes obtenues.

Un pourcentage considérable des classes produites représente des classes neutres donc peu significatives. La complexité de l'onde sonore polyphonique explique en partie ce phénomène. Néanmoins, comme l'illustre la figure 5.4, aucune classe obtenue ne peut être considérée comme bruitée.

Pour certains fichiers, un lissage considérable améliore la classification. Cependant, cette opération peut également déformer l'onde sonore et par conséquent détruire des caractéristiques importantes. Les résultats ainsi générés peuvent devenir incohérents. Pour cette raison le lissage doit être utilisé avec précaution.

5.3.2 L'analyse spectrale à des fins de classification de segments de musiques polyphoniques

Prétraitements et paramètres fixes :

- Les 3 fréquences les plus importantes ont été considérées;
- Le spectre de fréquence a été subdivisé en 31 bandes;
- La taille de la fenêtre de la transformée de Fourier a été figée à 32 768;
- La taille des n-grammes a été fixée à 2;
- Le paramètre de vigilance ρ a été fixé à 0.1 pour la classification XI;
- Le paramètre de vigilance ρ a été fixé à 0.01 pour la classification XII;
- Les hapax ont été supprimés.

a. Résultats de la classification XI

Les résultats sont similaires à ceux obtenus lors de l'analyse spatiale comme le démontre le tableau 5.17. Cependant, les segments groupés ne correspondent pas à ceux groupés lors des classifications IX et X.

Classes	Fichiers
1	1 - hot snake – retrofit 5 - hot snake - retrofit
2	2 - Elvis - houng dog
3	3 - hot snake - retrofit
4	9 - Red hot - Higher Ground
5	6 - Elvis - houng dog 8 - Elvis - houng dog
6	11 - Calexico - Tres Avisos
7	10 - Red hot - Higher Ground
8	12 - Red hot - Higher Ground
9	4 - Calexico - Tres Avisos
10	7 - Calexico - Tres Avisos

Tableau 5.17 : Répartition des fichiers lors de la classification XI.

Sur les dix classes produites lorsque le paramètre de vigilance a été fixé à 0.1, huit contiennent un segment. Seules les classes 1 et 5 ne sont pas composées de segments orphelins. Fait notable, ces classes contiennent des pièces rythmées.

b. Résultats de la classification XII

Le nombre de singletons est considérablement réduit lorsque le paramètre de vigilance est fixé à 0.01 comme l'indique le tableau 5.18.

Classes	Fichiers
1	1 - hot snake - retrofit 2 - Elvis - houng dog
2	3 - hot snake – retrofit 5 - hot snake - retrofit
3	4 - Calexico - Tres Avisos, 7 - Calexico - Tres Avisos
4	6 - Elvis - houng dog 8 - Elvis - houng dog
5	9 - Red hot - Higher Ground
6	10 - Red hot - Higher Ground 11 - Calexico - Tres Avisos
7	12 - Red hot - Higher Ground

Tableau 5.18 : Répartition des fichiers lors de la classification XII.

Un total de sept classes est généré. Contrairement à la classification XI où plusieurs classes contiennent un seul membre, la majorité des classes produites sont formées de plus d'un segment.

Les classes 2, 3 et 4 permettent de valider l'hypothèse selon laquelle le système est en mesure de rassembler les segments d'une pièce. Chacune de ces classes est constituée de deux segments et représente une musique particulière.

La classe 6 est formée de deux segments de musique à sonorité différente. Cette classe est donc considérée comme bruitée.

Les classes 2 et 4 de la classification XII sont équivalentes respectivement aux classes 1 et 5 de la classification XI. Cette constante indique la présence de caractéristiques fondamentales pour ce genre de musique. Ce caractère peut expliquer le contenu de la classe 1 qui regroupe un segment de musique rock'n roll et un segment de musique rock. Par conséquent, tenant compte de ce facteur, cette classe peut être jugée comme étant pertinente.

c. Observations

La figure 5.5 illustre l'évaluation des classifications considérant :

i. Les classes formées uniquement de segments d'une même pièce comme pertinentes;

ii. Les classes contenant majoritairement les segments d'une même pièce comme peu bruitées ;

iii. Les classes contenant moins de 50 % de segments d'une même pièce comme bruitées;

iv. Les classes contenant un seul membre comme neutres.

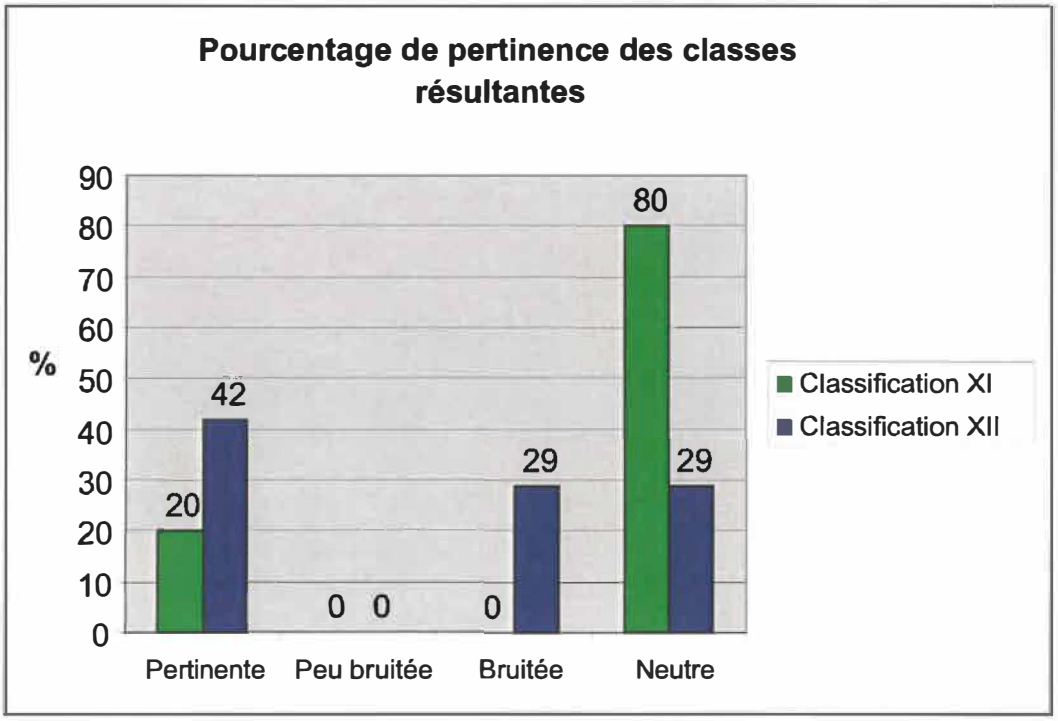


Figure 5.5 : Histogramme de l'évaluation des classes obtenues lors des classifications XI et XII.

Selon l'évaluation illustrée à la figure 5.5, 29 % des classes de la classification XII sont considérées comme bruitées. Cependant, tenant compte du facteur de ressemblance existant entre deux enregistrements, l'interprétation des résultats diffère considérablement.

La figure 5.6 illustre l'évaluation de la classification XII considérant :

i. Les classes formées de segments d'une même pièce ou de pièces similaires comme pertinentes;

ii. Les classes contenant majoritairement les segments d'une même pièce ou de pièces similaires comme peu bruitées;

iii. Les classes contenant moins de 50 % de segments d'une même pièce ou de pièces similaires comme bruitées;

iv. Les classes contenant un seul membre comme neutres.

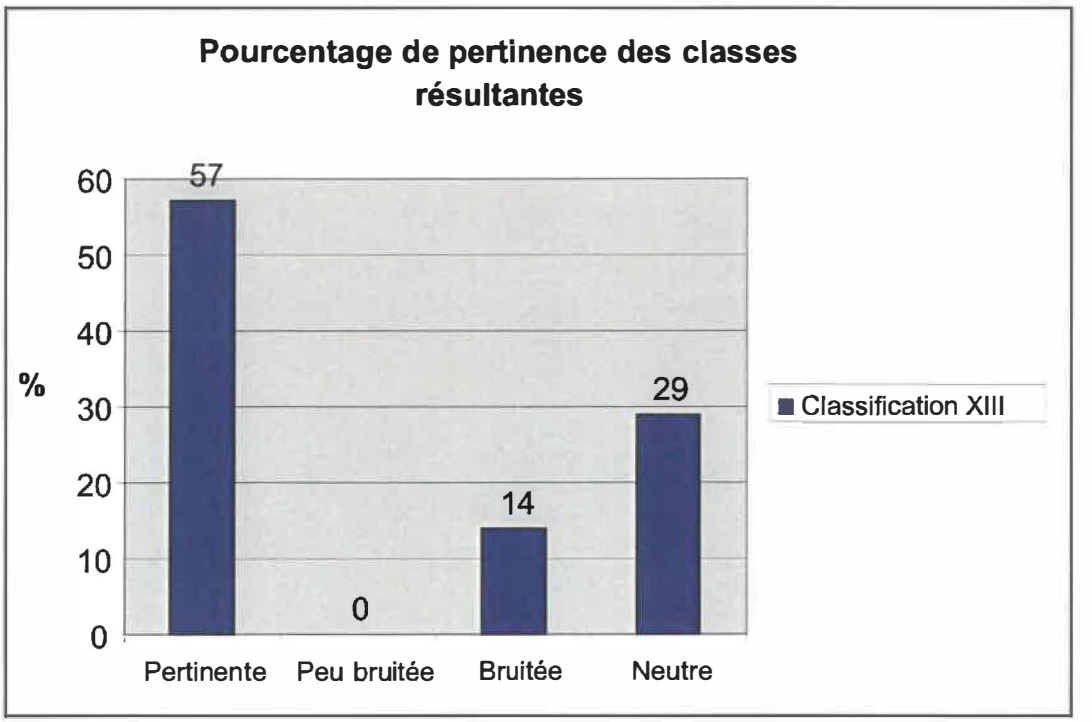


Figure 5.6 : Histogramme de l'évaluation des classes obtenues lors de la classification XII en tenant compte du style des pièces.

Lorsque l'on considère les caractéristiques des genres musicaux, 57% des classes de la classification XIII sont jugées pertinentes et le nombre de classes bruitées diminue à 14%.

L'interprétation multiple de cette classification démontre la subjectivité de l'évaluation des résultats obtenus. Notre culture musicale influence nos critères d'évaluation d'une classification.

5.3.3 L'impact des paramètres sur le résultat de la classification

L'analyse spectrale étudie la structure fondamentale de l'onde sonore. Les systèmes de reconnaissance vocale utilisent pour la plupart cette technique afin d'identifier la provenance d'un son. Les résultats obtenus lors de la classification de documents polyphoniques fondée sur le contenu fréquentiel sont inférieurs à ceux escomptés, ce qui laisse croire que certains facteurs ont détérioré le résultat de la classification. Afin de valider cette hypothèse, nous avons varié certains paramètres afin de définir leur impact.

a. Le nombre de fréquences considérées

Pour l'ensemble des classifications effectuées précédemment, nous avons fixé le nombre de fréquences considérées à 3. Le son polyphonique résulte de la sommation de plusieurs signaux de fréquences différentes. Nous avons donc incrémenté progressivement le nombre de fréquences considérées afin de vérifier si l'analyse des fréquences d'une moindre importance permettait d'améliorer la classification. Jusqu'à 10 fréquences ont été considérées. La nature des résultats obtenus ne permet pas d'affirmer que ce facteur améliore le processus.

b. La taille des n-grammes

Des travaux effectués dans le cadre d'expérimentations antérieures [Biskri et al, 2006] ont démontré que les bi-grammes étaient favorables à l'identification de caractéristiques de fichiers multimédias. Lorsque nous avons augmenté la taille des n-grammes la capacité du classifieur à distinguer les caractéristiques des documents a diminué. Le classifieur est dans l'incapacité de différencier les documents lorsque n est supérieur à 4 et ce indépendamment de la valeur du paramètre de vigilance introduit.

c. Les hapax

Le rôle des hapax demeure encore imprécis. Certaines associations on été créées lorsque ceux-ci ont été conservés. Cependant, ces associations ont été effectuées au détriment d'autres.

Aucune règle élémentaire n'a pu être établie.

d. L'ordonnancement des fichiers

Un fichier ne peut faire partie de plus d'une classe. La création d'une nouvelle classe dépend du niveau de ressemblance entre le fichier en traitement et ceux existants. L'ordonnancement des fichiers a donc un impact direct sur les résultats produits. Ce phénomène, attribuable à ART, nous force à tenir compte d'un taux d'erreurs dans l'évaluation des résultats.

e. La nature des fichiers

L'impact de l'ordonnancement des fichiers sur le processus de classification laisse croire que la nature des fichiers influence le classifieur. Afin de vérifier cette hypothèse, nous avons augmenté le nombre de candidats similaires afin de permettre au classifieur de mieux cibler les caractéristiques recherchées. Le tableau 5.19 dresse la liste des documents utilisés où les fichiers dont le nom commence par « Copy of » sont des clones.

Fichiers
1 - hot snake - retrofit
2 - Elvis - houng dog
3 - hot snake - retrofit
4 - Calexico - Tres Aviso
5 - hot snake - retrofit
6 - Elvis – houng dog
7 - Calexico - Tres Avisos
8- Elvis - houng dog
9 - Red hot - Higher Ground
10 - Red hot - Higher Ground
11 - Calexico - Tres Avisos
12 - Red hot - Higher Ground
Copy of 1 - hot snake - retrofit
Copy of 2 - Elvis - houng dog
Copy of 3 - hot snake - retrofit
Copy of 4 - Calexico - Tres Aviso
Copy of 5 - hot snake - retrofit
Copy of 6 - Elvis – houng dog
Copy of 7 - Calexico - Tres Aviso
Copy of 8 - Elvis - Tres Avisos
Copy of 9 - Red hot - Higher Ground
Copy of 10 - Red hot - Higher Ground
Copy of 11 - Calexico - Tres Avisos
Copy of 12 - Red hot - Higher Ground

Tableau 5.19 : Pièces de musiques polyphoniques et leurs clones.

La paramétrisation a été effectuée de manière analogue à la classification XII :

- Les 3 fréquences les plus importantes sont considérées;
- Le spectre de fréquence a été subdivisé en 31 bandes;
- La taille de la fenêtre de la transformée de Fourier a été fixée à 32 768;
- La taille des n-grammes a été fixée à 2;
- Les hapax ont été supprimés.

Classification XIII

Le tableau 5.20 donne la répartition des classes produites lorsque le classifieur a été conditionné à l'aide de candidats identiques.

Les résultats obtenus sont surprenants. De nouvelles associations ont été créées par rapport à la classification XII. De plus, un nombre considérable de classes pertinentes est produit. Des 7 classes obtenues, seule la classe numéro 6 semble bruitée.

Classes	Fichiers
1	Copy of 1 - hot snake - retrofit 1 - hot snake - retrofit 2 - Elvis - houng dog Copy of 2 - Elvis - houng dog
2	3 - hot snake - retrofit Copy of 3 - hot snake - retrofit 5 - hot snake - retrofit Copy of 5 - hot snake - retrofit
3	4 - Calexico - Tres Avisos Copy of 4 - Calexico - Tres Avisos 7 - Calexico - Tres Avisos Copy of 7 - Calexico - Tres Avisos
4	6 - Elvis - houng dog Copy of 6 - Elvis - houng dog 8 - Elvis - houng dog Copy of 8 - Elvis - houng dog
5	9 - Red hot - Higher Ground Copy of 9 - Red hot - Higher Ground
6	10 - Red hot - Higher Ground Copy of 10 - Red hot - Higher Ground 11 - Calexico - Tres Avisos Copy of 11 - Calexico - Tres Avisos
7	12 - Red hot - Higher Ground Copy of 12 - Red hot - Higher Ground

Tableau 5.20 : Répartition des fichiers lors de la classification XIII.

Les classes 5 et 7 sont composées d'un segment et de sa copie. Elles sont considérées comme neutres. On remarque que les deux classes neutres contiennent des extraits de la même pièce.

La relation ressortie lors de la classification XII entre les fichiers « 1 - hot snake – retrofit » et « 2 - Elvis - houng dog » est également observée. La classe 1 contient ces deux fichiers et leur copie. Cette association redondante réaffirme la ressemblance entre les deux genres musicaux représentés.

Les classes 2, 3 et 4 regroupent des segments de la même pièce et leur copie respective. Ces classes donnent les résultats attendus.

La classification XIII démontre que lorsque le classifieur est conditionné par l'ajout de candidats identiques, la qualité des classes produites est considérablement accrue. L'augmentation du nombre de candidats similaires permet au classifieur de mieux cibler les caractéristiques à rechercher.

Classification XIV

De manière à consolider l'observation effectuée lors de la classification XIII, nous avons répété l'expérimentation mais avec un nombre de fichier deux fois supérieur. La liste des fichiers utilisés est donnée au tableau 5.21.

Fichiers	**Fichiers clones**
1 - Elvis - houng dog.wav	Copy of 1 - Elvis - houng dog.wav
2 - hot snake - retrofit .wav	Copy of 2 - hot snake - retrofit .wav
3 - Calexico - Tres Avisos .wav	Copy of 3 - Calexico - Tres Avisos .wav
4 - hot snake - retrofit .wav	Copy of 4 - hot snake - retrofit .wav
5 - Sigur Ros - Untitled.wav	Copy of 5 - Sigur Ros - Untitled.wav
6 - Calexico - Tres Avisos .wav	Copy of 6 - Calexico - Tres Avisos .wav
7 - Elvis - houng dog.wav	Copy of 7 - Elvis - houng dog.wav
8 - Calexico - Tres Avisos .wav	Copy of 8 - Calexico - Tres Avisos .wav
9 - Sigur Ros - Untitled.wav	Copy of 9 - Sigur Ros - Untitled.wav
10 - Red hot - Higher Ground.wav	Copy of 10 - Red hot - Higher Ground.wav
11 - Sigur Ros - Untitled.wav	Copy of 11 - Sigur Ros - Untitled.wav
12 - Calexico - Tres Avisos .wav	Copy of 12 - Calexico - Tres Avisos .wav
13 - Elvis - houng dog.wav	Copy of 13 - Elvis - houng dog.wav
14 - Red hot - Higher Ground.wav	Copy of 14 - Red hot - Higher Ground.wav
15 - Red hot - Higher Ground.wav	Copy of 15 - Red hot - Higher Ground.wav
16 - Sigur Ros - Untitled.wav	Copy of 16 - Sigur Ros - Untitled.wav
17 - Calexico - Tres Avisos .wav	Copy of 17 - Calexico - Tres Avisos .wav

18 - Sigur Ros - Untitled.wav	Copy of 18 - Sigur Ros - Untitled.wav
19 - Elvis - houng dog.wav	Copy of 19 - Elvis - houng dog.wav
20 - Red hot - Higher Ground.wav	Copy of 20 - Red hot - Higher Ground.wav
21 - Elvis - houng dog.wav	Copy of 21 - Elvis - houng dog.wav
22 - Red hot - Higher Ground.wav	Copy of 22 - Red hot - Higher Ground.wav
23 - Calexico - Tres Avisos .wav	Copy of 23 - Calexico - Tres Avisos .wav
24 - Red hot - Higher Ground.wav	Copy of 24 - Red hot - Higher Ground.wav
25 - Elvis - houng dog.wav	Copy of 25 - Elvis - houng dog.wav
26 - Sigur Ros - Untitled.wav	Copy of 26 - Sigur Ros - Untitled.wav
27 - hot snake - retrofit .wav	Copy of 27 - hot snake - retrofit .wav

Tableau 5.21 : Pièces de musiques polyphoniques de l'expérimentation XIV.

Les résultats sont en relation avec ceux de la classification XIII. Parmi les 15 classes générées, 11 contiennent uniquement des segments d'une même pièce. Le tableau 5.22 donne la répartition des fichiers.

Classes	Fichiers
1	Copy of 27 - hot snake - retrofit .wav 1 - Elvis - houng dog.wav 14 - Red hot - Higher Ground.wav Copy of 14 - Red hot - Higher Ground.wav Copy of 1 - Elvis - houng dog.wav 27 - hot snake - retrofit .wav
2	2 - hot snake - retrofit .wav 4 - hot snake - retrofit .wav Copy of 2 - hot snake - retrofit .wav Copy of 4 - hot snake - retrofit .wav
3	3 - Calexico - Tres Avisos .wav 6 - Calexico - Tres Avisos .wav 12 - Calexico - Tres Avisos .wav Copy of 3 - Calexico - Tres Avisos .wav Copy of 6 - Calexico - Tres Avisos .wav Copy of 12 - Calexico - Tres Avisos .wav
4	5 - Sigur Ros - Untitled.wav 8 - Calexico - Tres Avisos .wav Copy of 5 - Sigur Ros - Untitled.wav Copy of 8 - Calexico - Tres Avisos .wav
5	7 - Elvis - houng dog.wav 13 - Elvis - houng dog.wav 19 - Elvis - houng dog.wav Copy of 7 - Elvis - houng dog.wav Copy of 13 - Elvis - houng dog.wav Copy of 19 - Elvis - houng dog.wav
6	9 - Sigur Ros - Untitled.wav

	Copy of 9 - Sigur Ros - Untitled.wav
7	10 - Red hot - Higher Ground.wav Copy of 10 - Red hot - Higher Ground.wav
8	11 - Sigur Ros - Untitled.wav 18 - Sigur Ros - Untitled.wav 23 - Calexico - Tres Avisos .wav Copy of 11 - Sigur Ros - Untitled.wav Copy of 18 - Sigur Ros - Untitled.wav Copy of 23 - Calexico - Tres Avisos .wav
9	15 - Red hot - Higher Ground.wav 16 - Sigur Ros - Untitled.wav Copy of 15 - Red hot - Higher Ground.wav Copy of 16 - Sigur Ros - Untitled.wav
10	17 - Calexico - Tres Avisos .wav Copy of 17 - Calexico - Tres Avisos .wav
11	20 - Red hot - Higher Ground.wav Copy of 20 - Red hot - Higher Ground.wav
12	21 - Elvis - houng dog.wav Copy of 21 - Elvis - houng dog.wav
13	22 - Red hot - Higher Ground.wav 24 - Red hot - Higher Ground.wav Copy of 22 - Red hot - Higher Ground.wav Copy of 24 - Red hot - Higher Ground.wav
14	25 - Elvis - houng dog.wav Copy of 25 - Elvis - houng dog.wav
15	26 - Sigur Ros - Untitled.wav Copy of 26 - Sigur Ros - Untitled.wav

Tableau 5.22 : Répartition des fichiers lors de la classification XVI.

Classification XV

Afin de maximiser le conditionnement du classifieur, nous avons ordonnancé les fichiers de manière à grouper les extraits de la même pièce. Le tableau 5.23 donne la répartition des classes produites dans ce contexte optimal.

Les résultats sont concluants. Cinq classes ont été créées. Les classes 1, 2 et 3 contiennent la totalité des segments d'une pièce. Seule la pièce funk a été subdivisée en deux classes. Lorsque l'on écoute les segments, on constate que le refrain a été isolé des couplets de la

chanson. Cette distinction reflète les variances perceptibles dans la musique. Différentes atmosphères résultant d'un changement de mélodie ou de rythme créent des distinctions au sein d'une même pièce. Ce caractère d'hétérogénéité donne lieu à la répartition des segments d'une pièce à l'intérieur de plus d'une classe.

Classes	Fichiers
1	Elvis - houng dog 1 Elvis - houng dog 2 Elvis - houng dog 3 Copy of Elvis - houng dog 1 Copy of Elvis - houng dog 2 Copy of Elvis - houng dog 3
2	hot snake - retrofit 1 hot snake - retrofit 2 hot snake - retrofit 3 Copy of hot snake - retrofit 1 Copy of hot snake - retrofit 2 Copy of hot snake - retrofit 3
3	Red hot - Higher Ground 1 Copy of Red hot - Higher Ground 1
4	Red hot - Higher Ground 2 Red hot - Higher Ground 3 Copy of Red hot - Higher Ground 2 Copy of Red hot - Higher Ground 3
5	Calexico - Tres Avisos 1 Calexico - Tres Avisos 2 Calexico - Tres Avisos 3 Copy of Calexico - Tres Avisos 1 Copy of Calexico - Tres Avisos 2 Copy of Calexico - Tres Avisos 3

Tableau 5.23 : Répartition des fichiers lors de la classification XV.

5.3.4 Sommaire de l'expérimentation 3

Le traitement de son polyphonique nécessite des ajustements. La complexité de l'onde sonore augmente la taille du lexique ce qui génère une signature variée pour chacun des fichiers. Malgré le niveau de complexité accru, le système proposé parvient tout de même à générer des classes cohérentes. On remarque des associations constantes, ce qui se traduit par la présence de caractéristiques importantes tel que des sonorités similaires.

Ces caractéristiques ne se limitent pas à une pièce mais peuvent être observées à travers différents styles musicaux. Ainsi, des extraits de musique rock'n roll et de musique rock sont communément regroupés. Lorsque l'on considère les ressemblances entre les genres musicaux, la qualité des résultats est haussée.
Un paramétrage inadéquat se traduit par la création d'un nombre considérable de classes neutres. Certains paramètres ont un impact énorme sur la classification.

Les effets suivants ont été observés :

- Le classifieur est dans l'incapacité de différencier les documents lorsque n est supérieur à 5 puisque la probabilité d'occurrence de ceux-ci devient très faible;

- L'ordonnancement des fichiers a un impact direct sur les résultats produits;

- La nature des fichiers traités influence le classifieur.

Ainsi, lorsque l'on conditionne le classifieur, une classification parfaite peut être générée. Dans un contexte réel de classification, la duplication des documents ne peut être envisagée. La rétroaction avec l'utilisateur devient donc une option à considérer. Pour produire les résultats escomptés, le classifieur doit être en mesure de bâtir un modèle valable des candidats recherchés. Une connaissance plus approfondie du fonctionnement de ART permettrait un meilleur conditionnement.

La musique est une source de données riche et diversifiée. Plusieurs ambiances sont perçues à même une pièce. Ce caractère hétérogène ajoute un aspect multi caractériel à la musique. Il devient par le fait même difficile d'associer une musique à une seule catégorie.

5.4. Sommaire du chapitre 5

Deux modes d'analyse ont été étudiés : spatiale et spectrale. En général, l'analyse spectrale donne de meilleurs résultats. Néanmoins, on ne peut exclure la forme de l'onde comme fondement d'une classification. Par contre, l'estimation de cette forme doit être plus précise que celle présentée.

Lors de la classification de fichiers monophoniques, une reconnaissance plus ou moins précise des octaves a été perçue. Dans ce contexte, l'analyse fréquentielle offre un niveau intéressant de reconnaissance. L'effet de compression généré par le groupement de fréquences similaires a permis de rassembler un nombre considérable de notes successives au sein d'une même classe.

De manière générale, la classification de fichiers monophoniques donne des résultats plus constants que la classification de documents polyphoniques. Le niveau de complexité de l'onde sonore est à l'origine de cette constatation. Néanmoins, le système parvient à générer des classes cohérentes lorsque l'input est formé de sons polyphoniques. La qualité de la classification est optimale lorsque le nombre de candidats similaires est augmenté de manière à ce que le classifieur puisse mieux cibler les caractéristiques recherchées.

La paramétrisation est un aspect critique, tout particulièrement en ce qui concerne le paramètre de vigilance du classifieur.

6. Conclusion

La demande de fichiers sonores est sans cesse croissante. Les données recherchées sont généralement désordonnées et réparties à travers divers entrepôts de données. L'obtention des fichiers désirés requiert l'utilisation de diverses techniques telle l'indexation. L'augmentation du volume de données contraint la gestion des index et nécessite l'automatisation de certains processus. La classification est un prétraitement qui peut améliorer l'indexation. L'appartenance d'un document à une classe de similarité est établie suite à une analyse statistique des données qui le composent. La recherche et l'identification des caractéristiques communes constituent l'essence de la classification. Les index peuvent être bâtis à partir de ces caractéristiques.

Le processus de classification peut être automatisé. Pour ce faire, l'attention doit être portée sur les mécanismes d'identification et d'analyse des caractéristiques. Nous nous sommes intéressé à ces mécanismes dans l'objectif d'automatiser la classification de documents sonores.

Nous avons approfondi une approche hybride qui fusionne des concepts tirés du traitement automatisé des langues naturelles et de l'analyse fréquentielle. La classification est réalisée à partir d'une comparaison statistique de séquences particulières traditionnellement associées au domaine du traitement automatisé des langues naturelles. Ces séquences nommées n-gramme sont constituées de n caractères descriptifs consécutifs. L'avantage de ce modèle de découpage est qu'il assure un niveau d'adaptabilité intéressant en plus de maintenir le lexique à un seuil raisonnable. Ce

dernier aspect est un facteur important considérant que le son est un phénomène continu défini par une infinité de valeurs.

Notre approche est innovatrice parce qu'elle utilise une chaîne de traitements flexible pouvant être utilisée pour différents formats de fichiers. Seule la lecture et la préparation des données diffèrent d'un format à l'autre. La recherche et l'optimisation du caractère descriptif sont les tâches les plus délicates. Dans le cadre de nos travaux, la forme de l'onde et le spectre du signal ont été considérés comme caractères descriptifs puisque ces attributs donnent plusieurs indications sur la nature et la composition de l'onde sonore. Un mode d'analyse a été dédié à chacune de ces propriétés.

Le processus défini est semi-automatique car le choix de certains paramètres est fait par l'utilisateur. Les étapes menant à la classification sont:

i. La lecture et la préparation des données;

ii. L'extraction des n-grammes et création des représentations vectorielles;

iii. Présentation des données au classifieur;

iv. L'évaluation des classes obtenues.

Plusieurs expérimentations ont été effectuées afin de valider le modèle étudié. La nature des résultats obtenus suggère que l'analyse du contenu fréquentiel génère de meilleurs résultats. Le contenu fréquentiel d'un signal sonore donne une multitude d'informations sur la provenance du son. Ce mode d'analyse permet également de distinguer efficacement les sons graves et aigus. Le développement futur doit donc accorder une importance plus grande à cet aspect. La direction a prendre pour augmenter la robustesse du système est de se servir de l'analyse fréquentielle pour détecter des indicateurs du rythme des pièces soumises à la classification.

La qualité de la classification est largement affectée par le classifieur utilisé. Dans le cadre de nos expérimentations, le réseau de neurones ART a été utilisé. Cet outil ne permet pas qu'un élément appartienne à plus d'une classe et par conséquent l'ordonnancement des fichiers influence grandement le comportement du classifieur et par conséquent la qualité des résultats générés. Les classes obtenues lors des expérimentations résultent d'un jeu d'essais en rapport avec les différents paramètres du système. Bien que plusieurs combinaisons puissent donner de bons résultats, la paramétrisation du classifieur demeure un aspect empirique. Le manque de régularité causé par cette défaillance peut être ciblée et résolue par l'utilisation d'algorithmes de clustering. L'adoption d'un nouveau classifieur, SVM (*Support Vector Machine*) par exemple ou tout autre classifieur permettant d'associer un élément à plus d'une classe, est une solution à envisager pour obtenir des résultats plus constants.

Une classification peut relever plusieurs aspects comme des sonorités similaires par exemple. Deux mélodies distinctes mais constituées des mêmes notes peuvent être jugées différente ou semblable selon que l'on s'intéresse uniquement à la structure de la mélodie ou aux notes qui la composent. Par conséquent, il est possible d'interpréter une classification de multiples façons. Cette réalité démontre la subjectivité de l'évaluation des résultats obtenus.

Un système de classification doit pouvoir s'adapter aux critères de recherche définis. Pour arriver à ce résultat, le système doit être conditionné. Dans un environnement en constante évolution, il est irréaliste de préconiser le contrôle de l'ordonnancement et de la nature des fichiers à classifier. Dès lors, l'intervention de l'utilisateur devient inévitable. Celui-ci doit être interrogé afin pondérer la pertinence des classes qui lui sont retournées. Suite à cette évaluation, les caractéristiques communes des membres des classes pertinentes doivent être recherchées de manière à optimiser le nombre de candidats susceptibles de répondre aux attentes.

Une rétroaction doit être ajoutée à la chaîne de traitements de manière à personnaliser la classification et ainsi en améliorer la qualité.

7. Références bibliographiques

[1] Ismaïl Biskri, Louis Rompré, Lamri Laouamer & François Meunier (2006). *Classification de documents Multimédias : vers une approche générale.* Acte du colloque JADT 2006, Besançon, France.

[2] Charles Bouveyron (2006): Modélisation *et classification des données de grande dimension, Application à l'analyse d'images*. Thèse. Université de Joseph Fourier, Grenoble, France.

[3] Nikunj Patel et Padma Mundur (2005): *An n-gram based approach to finding the repeating pattern in muical data.* EuroIMSA 2005, Grindelwald, Switzerland.

[4] Lucille Tanquerel, Luigi Lancien (2005): *Techniques pour la recherche de similarités musicales, un état de l'art*. CORESA 2005, Rennes, France.

[5] J.M Brück, S. Bres & D. Pellerin (2004): *Construction d'une signature audio pour l'indexation de documents audiovisuels*. CORESA 2004, Lille, France.

[6] Stephen Robertson (2004): *Understanding Inverse Document Frequency: On theoretical arguments for IDF*. Journal of Documentation, Volume 60, Issue 5, Page 503-520.

[7] J. Stephen Downie (2003): *Music Information Retrival.* Annual Review of Information Science and Technology. Volume 37, Pages 295-340.

[8] Hadi Harb (2003): *Classification du signal sonore en vue d'une indexation par le contenu des documents multimédias.* Thèse. École Centrale de Lyon, France.

[9] Melissa Ray Weimer (2003): *Waveform Analysis Using The Fourier Transform.* DATAQ Instrument Technical Sheet.

[10] J. K. Paulus & A. P. Klapuri (2003). *Conventional and periodic n-grams in the transcription of drum sequences.* Proceedings of the 2003 International Conference on Multimedia and Expo - Volume 1.

[11] Gail A. Carpenter & Stephen Grossberg (2003): *Adaptive Resonance Theory.* In M.A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, Second Edition, Cambridge, MA : MIT Press, Pages 87-90.

[12] Ismaïl Biskri & Jean-Guy Meunier (2002): *SATIM : Système d'Analyse et de traitement de l'information Multidimensionnelle.* Articles du Colloque international: L'édition électronique en littérature et dictionnairique: évaluation et bilan. Rouen, France, Juin 2002.

[13] Chih-Wei Hsu & Chih-Jen Lin (2002): *A comparison of methods for multiclass support vector machines.* IEEE Transactions on Neural Networks 13, 415-425.

[14] François Meunier, Jean Meunier, François Cavayas (2001): *Synthetic Aperture Radar Image of Agricultural Fields with Surface Network.* Simulation and Spatial Information Retrieval, Optical Engineering, Vol. 40, Number 10, October 2001, pp. 2319-2330.

[15] Thomas L. Floyd (2001): *Systèmes numériques : Concepts et application.* Les Éditions Reynald Goulet inc.

[16] Thomas J. Cavicchi (2000). *Digital Signal Processing.* John Wiley & Sons, Inc.

[17] J. Stephen Downie, Michael Nelson (2000): *Evaluation of a simple and effective music information retrieval method.* ACM SIGIR conference on Research and development in information retrieval, Athens, Greece.

[18] Juan-Manuel Torres-Moreno, Patricia Velaquez-Morales & Jean-Guy Meunier (2000) : *Classphères : un réseau incrémental pour l'apprentissage non supervisé appliqué à la classification de texte.* Acte du colloque JADT 2000, Lausanne, France.

[19] J. Stephen Downie (1999). *Evaluating a simple approach to music information retrieval: conceiving melodic n-grams as text.* PhD Thesis, Faculty of Graduate Studies of the University of Western Ontario.

[20] Ethan L. Miller, Dan Shen, Junli Liu, Charles Nicholas & Ting Chen (1999): *Techniques for Gigabyte-Scale N-gram Based Information Retrieval on Personal Computers*, Acte du colloque PDPTA 99, Las Vegas, États-Unis.

[21] Bernard Caillaud et Mireille Leriche (1999): *Analyse Sonographique et aspect de la phonétique appliquée.* Revue de l'EPI, Bulletin 93, Pages 57-70.

[22] Alain Lelu, Mohammed Halleb et Bruno Delprat (1998): *Recherche d'information et cartographie dans des corpus textuels à partir des fréquences de n-grammes.* Acte du colloque JADT 98, Nice, France.

[23] Gregory Grefenstette (1995). *Comparing Two Language Identification Schemes.* Acte du colloque JADT 1995, Rome, Italie.

[24] Marc Damashek, (1995). *Gauging Similarity with n-Grams: Language-Independent Categorization Of Text.* Science, Vol. 267, pages 843-848.

[25] Ted Dunning (1994). *Statistical Identification of Language.* Technical Report MCCS 94-273. Computing Research Laboratory, New Mexico State University.

[26] Rafael. C. Gonzalez & R. E. Woods (1992): *Digital Image Processing*, Addison-Wesley

[27] Carpenter, Grossberg & Rosen (1991): *A Fuzzy Art : Fast Stable Learning and Categorization of Analog Patterns by Adaptative Resonance System.* Neural Networks, Volume 4, Pages 759-771.

[28] Rafael C. Gonzalez & Paul Wintz (1987): *Digital Image Processing.* Addison Wesley.

[29] Salton, Gerard, and Michael J. McGill (1983): *Introduction to modern information retrieval.* New York: McGraw-Hill.

[30] G. Salton, A. Wong & C. S Yang (1975) *A Vector Space Model for Automatic Indexing.* Communication of the ACM. Volume 18, pages 613-620. New York, États-Unis.

[31] Oppenheim, A. V., Schafer, R. W. (1975): *Digital Signal Processing*, Prentice Hall.

[32] Karen Spärck Jones (1972): A *statistical interpretation of term specificity and its application in retrieval.* Journal of Documentation, Volume 28, Pages 11-21.

[33] Jerome Bruner, Jaqueline Jarret Goodnow & George A. Austin (1956): *A Study of Thinking*. John Wiley & Sons, New York.

[34] C. E. Shannon (1951): *Prediction and Entropy of Printed English.* Bell System Technical Journal, Volume 30, Pages 50-64.

[35] G.K. Zipf : *Human Behavior and the Principle of Least Effort. Addison-Wesley*, New York, 1949

[36] C. E. Shannon (1948): *A Mathematical Theory of Communication.* Bell System Technical Journal, Volume 27, Pages 379-423 and 623-656, July and October 1948.