

Techniques pour la recherche de similarités musicales - Un état de l'art -

Lucille Tanquerel¹ Luigi Lancieri¹

¹ France Télécom R&D
42, Rue des coutures
14 000 CAEN

{lucille.tanquerel, luigi.lancieri}@francetelecom.com

Mots clés : MIR (Music Information Retrieval), similarités, signal audio, métadonnées,
Concours jeune chercheur : Non

Résumé

Nous proposons ici une étude des différentes techniques actuelles permettant de déterminer des similarités musicales entre titres, artistes ou même fichiers audio. Nous nous appuyons pour cela sur quelques travaux effectués dans le domaine MIR (Music Information Retrieval). Les méthodes, très variées, vont du traitement du signal à l'analyse de métadonnées.

1 Introduction

La musique au format numérique est aujourd'hui présente partout. Son importance s'est accrue au cours de ces dernières années grâce à la compression des données et à l'essor des réseaux informatiques. Cette grande disponibilité est un atout car tout individu apprécie la musique sous une forme ou sous une autre. Le problème reste de faire des choix parmi ces quantités importantes de données. Ce problème est encore actuel pour l'information textuelle même si les moteurs de recherche ont fait des progrès importants. La caractérisation et la classification de documents sonores, qu'ils soient structurés ou non, pose des problèmes nouveaux par rapport au texte, notamment celui de la sémantique de la classification. Les genres, qui sont une déclinaison particulière de cette segmentation, ne font pas forcément l'objet d'un consensus car ils n'expriment pas toujours de manière objective la sensibilité et les nuances retransmises par le son. En effet, la perception de la similarité musicale peut varier en fonction des personnes mais aussi à travers le temps. La similarité est, il est vrai, une notion complexe, tout particulièrement en musique où elle peut s'exprimer sous plusieurs formes : il peut s'agir du timbre, de l'harmonie, du genre, de l'identité de l'artiste, de la similarité culturelle (les Beatles et les Rolling Stones), de la similarité symbolique (hymnes nationaux). La similarité peut aussi être due au contexte social (musique des années soixante) ou complètement subjective (ces deux morceaux m'évoquent mes souvenirs d'enfance), etc.

Par ailleurs, même dans l'hypothèse d'un consensus sur une forme particulière de ressemblance, les caractéristiques qui permettraient une classification ne sont pas toujours faciles à extraire des documents sonores non structurés. Ces limites font que, aujourd'hui encore, la description et l'organisation de la musique se font souvent manuellement, ce qui constitue une opération coûteuse.

Les travaux sur l'extraction automatique de similarités musicales se divisent en trois principales approches : celles basées sur le traitement du signal, la fouille de données contextuelles, et le filtrage collaboratif. La première catégorie est basée sur les propriétés acoustiques du morceau et peut être appliquée à n'importe quel fichier audio. Les caractéristiques extraites de cette analyse peuvent alimenter des systèmes de classification de manière à identifier des similarités. L'approche basée sur la fouille de données exploite des informations textuelles distribuées souvent non structurées décrivant un morceau de musique ou un artiste. Avec le développement du Web, d'autres techniques basées sur le filtrage collaboratif ont émergé. Ces techniques combinent des données provenant de nombreux individus pour déterminer des similarités sur la base de consensus. Le filtrage collaboratif représente la technique principale utilisée aujourd'hui dans les systèmes de recommandations (Amazon, AllMusicGuide, etc).

Nous proposons dans cet article une analyse de la littérature sur ces différentes approches.

2 Caractérisation du genre musical par l'extraction de propriétés physiques

Le genre a une importance particulière dans la recherche d'informations musicales. Il est utilisé par les marchands, les libraires, les musicologues et les auditeurs. C'est en effet le meilleur moyen d'organiser la musique. Ce type de segmentation est classiquement réalisé de manière

manuelle en référence à un consensus de classification. La référence à l'humain est reprise dans de nombreux travaux. Des chercheurs [28] ont noté par exemple que l'écoute de 3 sec de musique suffisait à un humain pour reconnaître un genre dans 70 % des cas et dans 50% des cas pour des échantillons de 250 ms. Pour automatiser et rendre moins subjective la classification, la recherche s'est orientée vers l'extraction de caractéristiques physiques du signal sonore. En effet, certains critères formels tels que la texture (superposition des différents instruments ou voix qui jouent en même temps créant ainsi un son clair ou dense), l'instrumentation mais aussi la structure rythmique peuvent être utilisés pour attribuer un genre particulier à un morceau de musique.

L'extraction de caractéristiques spectrales a par exemple été utilisée pour classifier la musique ou la parole [5], les sons isolés [6, 7] et les instruments [8]. D'autres travaux se basent sur la reconnaissance du timbre [9, 10] ou sur la détection automatique de rythme [11, 12]. Dans les deux approches présentées ci-dessous, un certain nombre de caractéristiques physiques sont extraites du morceau puis stockées dans un vecteur numérique qui est analysé par un ou plusieurs classifieurs afin de pouvoir déterminer le genre du morceau. Ces classifieurs sont en général paramétrés par apprentissage sur la base d'exemples musicaux correspondant aux genres à reconnaître. La différence la plus importante entre ces deux approches est le choix des données à extraire. Dans un cas, il s'agit de caractéristiques de bas niveau obtenues à partir d'un morceau sans structure musicale préalable (format audio - WAV pour l'exemple -). Les descripteurs de bas niveau expriment des propriétés sonores significatives directement observables ou calculables à partir du signal audio, mais qui n'ont pas forcément de signification évidente au sens musical. Dans l'autre cas, les morceaux sont structurés dans un format symbolique comme MIDI. Dans ces formats, il est possible d'exploiter des caractéristiques de haut niveau difficilement disponibles pour des morceaux au format Wave. Les descripteurs de haut niveau expriment des propriétés qui ont une signification musicale plus évidente et compréhensible par les utilisateurs (e.g. hauteur et la durée d'une note). Une autre différence notable est le choix de la hiérarchie de genres utilisée.

2.1 Analyse du signal

G. Tzanetakis, G. Essl et P. Cook [1], par exemple, proposent un ensemble de caractéristiques pour représenter la surface musicale et la structure rythmique basé sur la reconnaissance de motifs (Statistical Pattern Recognition-SPR). La "surface musicale" est décrite comme l'ensemble des caractéristiques relatives à la texture, le timbre et l'instrumentation correspondant à des propriétés d'échantillons de distribution spectrale sur plusieurs périodes de temps (entre autres : luminance

spectrale, forme du spectre). Le calcul des caractéristiques se fait en utilisant la transformée de Fourier à court terme (Short Term Fourier Transform-STFT) qui peut être calculée efficacement en utilisant la transformée de Fourier rapide (Fast Fourier Transform-FFT) [13]. La structure rythmique est basée sur la détection des périodicités du signal les plus saillantes. Cette analyse est réalisée par la transformation en ondelettes (Wavelet Transform-WT) [14] dont une variante est la WT discrète (Discrete Wavelet Transform) qui fournit une représentation compacte du signal en temps et en fréquence.

Pour évaluer la performance de ces caractéristiques, les classifieurs SPR ont été utilisés avec des exemples audio de référence correspondant à un ensemble de 50 échantillons de 30 secondes chacun (fichiers au format WAV) pour chaque genre. Pour la musique "classique" les auteurs ont utilisé un niveau de description plus important par rapport aux autres genres (country, disco, hip-hop, jazz, rock). Ces caractéristiques supplémentaires ont été apportées par les MFCC (Mel-Frequency Cepstral Coefficients) [15] qui fournissent des caractéristiques spectrales à très court terme couramment utilisées dans la reconnaissance vocale.

Dans le tableau suivant qui donne quelques éléments de résultats, les colonnes correspondent aux genres réels et les lignes aux résultats obtenus. Par exemple, la cellule [li:2, col:1] signifie que 1% des morceaux de musique classique ont été classifiés en tant que musique country. La diagonale fournit le taux de reconnaissance d'un genre donné. Les meilleurs résultats sont obtenus pour le classique et le hip hop. Par contre le jazz et le rock qui sont des catégories très larges aux frontières moins nettes donnent de moins bons résultats.

	<i>Classique</i>	<i>Country</i>	<i>Disco</i>	<i>Hip Hop</i>	<i>Jazz</i>	<i>Rock</i>
<i>Classique</i>	86	2	0	4	18	1
<i>Country</i>	1	57	5	1	12	13
<i>Disco</i>	0	6	55	4	0	5
<i>Hip Hop</i>	0	15	28	90	4	18
<i>Jazz</i>	7	1	0	0	37	12
<i>Rock</i>	6	19	11	0	27	48

Figure 1 – Matrice du taux de reconnaissance des genres

L'étiquetage du genre pour les échantillons a été fait en fonction de l'artiste ou du CD contenant le morceau. Dans certains cas, cela constitue une source d'erreurs de classification. En effet, certains morceaux de Sting par exemple sont beaucoup plus proches du jazz que du rock, même pour un auditeur humain. De même, certains morceaux de jazz contiennent des passages avec accompagnement de cordes qui peuvent ressembler fortement au classique. Il est fort probable que les résultats seraient meilleurs en refaisant les tests avec une classification de référence fiable.

2.2 Extraction de caractéristiques physiques de haut niveau

Les caractéristiques de haut niveau comme la hauteur et la durée d'une note peuvent être extraites de manière fiable et systématique à partir des partitions musicales, qu'elles soient sous forme papier ou sous forme numérique (MIDI, MusicXML, Humdrum, GUIDO). Cette technique est fort utile pour reconstruire un morceau de musique quand son enregistrement sonore n'est pas disponible.

Même si ces données semblent plus précises que celles obtenues par l'analyse du signal, elles ont néanmoins besoin d'être assemblées pour être associées à un genre. La composition de cet assemblage va être déterminante dans l'efficacité des classifieurs car certains assemblages peuvent être incohérents et contenir des chevauchements. Une option peut consister en l'utilisation d'un large panel de caractéristiques mais les classifieurs peuvent être surchargés si elles sont trop nombreuses. Dans la pratique, la performance d'un classifieur peut être accrue par la combinaison d'une taxonomie hiérarchique et d'un système de sélection de caractéristiques efficace. Par exemple, C. McKay et I. Fuginaga [2] ont exploité un catalogue de 160 caractéristiques [16] pouvant servir à décrire et à classer des morceaux de musique. Les auteurs ont implémenté 109 d'entre elles correspondant à 6 catégories (instrumentation, texture musicale, rythme, dynamiques, hauteur des notes, mélodie, accords). Ces caractéristiques peuvent être considérées individuellement ou être assemblées en une caractéristique multidimensionnelle.

Comme dans le cas de l'analyse du signal, le choix des caractéristiques et de leur pondération peut être basé sur l'apprentissage à partir d'exemples. La classification peut être basée sur les réseaux de neurones (feedforward neural networks) [31] efficaces et adaptables mais coûteux en ressources, ou sur des algorithmes de classification de type plus proches voisins (k-nearest neighbours) [32] plus rapides mais moins adaptables. L'approche utilisée par les auteurs est la combinaison de ces deux techniques ainsi que l'utilisation d'algorithmes génétiques (GA) [17] [18] pour la sélection des caractéristiques.

Un assemblage de classifieurs est composé comme suit. Un simple classifieur (k plus proches voisins) exploitant un sous-ensemble pondéré de toutes les caractéristiques unidimensionnelles possibles est associé à un ensemble de classifieurs (réseau de neurones) exploitant un sous-ensemble des caractéristiques multidimensionnelles. Un tel ensemble peut être considéré comme une boîte noire qui prend en entrée toutes les caractéristiques d'un morceau, ignore les caractéristiques qui n'ont pas été sélectionnées, et donne en sortie un score de classification pour chaque catégorie candidate qu'il a été entraîné à reconnaître. La classification s'effectue de manière

hiérarchique. Un morceau est d'abord classé par genre "racine" (jazz, classique, populaire, etc.). Puis la classification procède au niveau inférieur de la hiérarchie. Le choix d'une ou plusieurs branches se fait en fonction des résultats obtenus au niveau supérieur (c'est-à-dire l'ensemble des candidats possibles). On itère ce processus jusqu'à arriver à une feuille.

Pour tester la constitution des classifieurs, les auteurs ont utilisé 950 fichiers MIDI au travers de deux taxonomies différentes. La première consiste en trois genres racines (classique, jazz, populaire) et trois genres feuilles par genre racine (baroque, bebop, country, etc). Les résultats de cette approche sont très bons. En moyenne, le genre racine a été correctement identifié à 98% et les genres feuilles ont été correctement identifiés à 90%. Cette première taxonomie étant un peu limitée, une autre plus réaliste a été mise en place. Elle est constituée de 9 genres racines et de 38 catégories feuilles. Les genres racines et feuilles ont été respectivement identifiés à 81% et 57%.

3 Exploitation de données contextuelles

Les techniques que nous venons d'évoquer permettent d'obtenir des éléments de description objectifs mais elles ont des limites. La sensibilité humaine intègre aussi des facteurs intangibles et subjectifs comme des influences culturelles ou l'émotion qu'il est impossible d'obtenir uniquement à partir de l'analyse quantitative d'un contenu audio. Des recherches récentes en MIR ont abordé l'exploitation de métadonnées culturelles pour évaluer la similarité entre artistes musicaux [19, 20]. Les métadonnées culturelles sont des informations décrivant l'opinion publique et les tendances culturelles provenant de divers textes non structurés produits par le public. D'autres chercheurs ont utilisé des systèmes de filtrage collaboratif permettant la classification sur la base de consensus de groupes [21, 22, 23]. Ces systèmes font des recommandations à un utilisateur en fonction de l'opinion d'autres utilisateurs ayant exprimé des goûts similaires. L'inconvénient de ces techniques est qu'elles ne sont applicables qu'à de la musique pour laquelle une quantité importante de données est disponible, ce qui est difficilement envisageable, par exemple, pour un nouvel artiste.

3.1 Fouille de données musicales

Cette approche repose sur une technique bien connue en linguistique qui est l'analyse de co-occurrence. L'idée de base est que plus deux items apparaissent ensemble dans un contexte donné, plus ils présentent une certaine similarité. En linguistique, l'analyse de co-occurrence de mots clés a été utilisée pour classer des documents sémantiquement proches ou pour la recherche d'information [29, 30].

Dans la classification automatique de titres musicaux, l'objectif est d'exploiter des informations contextuelles sémantiques ou épisodiques. Par exemple, certaines radios sont spécialisées dans certains types de musique comme la musique des années soixante (Nostalgie), ou comme la musique classique (Radio Classique), et d'autres ont des répertoires plus variés (comme FIP/Radio France). Toutefois, dans tous les cas, les titres et leur ordonnancement sont soigneusement sélectionnés de manière à ne pas fausser l'identité du programme. C'est cette référence (choix des titres et choix de leur ordre de passage) qui est utilisée pour la fouille de données.

F. Pachet, G. Westermann et D. Laigre [4] ont mis en œuvre cette technique en s'appuyant sur des sources musicales variées. Dans leur expérience, c'est la radio française FIP qui a été choisie car elle a l'avantage de ne pas être spécialisée dans un genre musical particulier. Les albums CD, et en particulier les compilations, constituent une autre source importante d'informations. En effet, les compilations, qu'elles soient officielles ou personnelles, ont une réelle consistance. Par exemple, les titres de compilations comme "Musique baroque française" ou "Hits de 1984" possèdent des similarités explicites de différentes sortes. La principale hypothèse est que la similarité entre deux titres est d'autant plus forte s'ils sont co-occurents dans différentes compilations. Des robots ont été spécialement développés pour récupérer des informations contextuelles sous forme de fichier texte sur le Web. Chacun de ces fichiers représente soit un album, soit un programme radio, ou encore un document contenant au moins deux titres musicaux considérés comme co-occurents. L'autre étape concerne l'analyse de co-occurrence elle-même qui permet d'identifier des similarités directes. Celle-ci consiste à construire une matrice avec tous les titres sur les lignes et les colonnes. La valeur en (i, j) correspond au nombre de fois où les titres i et j apparaissent ensemble, soit dans une même compilation, soit parce qu'ils sont voisins dans un programme radio. La similarité entre deux titres ou deux artistes peut aussi être corrélée à partir des similarités indirectes (si a est similaire à b et si b est similaire à c, alors a et c ont de grandes chances d'être similaires).

Une évaluation de cette méthode consiste à calculer le niveau des intersections entre l'opinion humaine sur la similarité, celle extraite des programmes radio et celle extraite des compilations. Dans l'évaluation réalisée par les auteurs, ces trois sous-ensembles sont composés respectivement par 100 artistes sélectionnés à la main de manière à représenter différents genres, 80 titres fréquemment joués sur la radio FIP, et 12 titres issus d'un album (CDDB, une large base de données disponible sur le Web). La technique de clustering utilisée est la classification hiérarchique ascendante [27].

Les résultats sont intéressants car les genres spécifiques sont assez bien distingués aussi bien pour les similarités directes que indirectes. Par exemple, les deux titres de guitare jazz (Jim Hall et Wes Montgomery) sont dans le même cluster. De plus, les titres d'un même artiste ont tendance à être regroupés ensemble. Les distances ont une valeur plus petite (i.e. les clusters sont plus resserrés) dans le cas de la similarité indirecte. Pour les genres plus généraux (comme le classique), la compilation est trop petite (12 titres) ici pour établir une conclusion générale.

Feuilles des arbres de similarité Niveau 1 clusters avec 2 artistes	Bons clusters	Mauvais clusters	Inconnus
FIP similarités directes	70%	25%	5%
CDDB similarités directes	76%	15%	8%
FIP similarités indirecte	53%	43%	4%
CDDB similarités indirecte	59%	30%	11%
Niveau 2 clusters avec 3, 4 ou 5 artistes			
FIP similarités directes	28%	72%	0%
CDDB similarités directes	54%	23%	23%
FIP similarités indirecte	47%	38%	17%
CDDB similarités indirecte	74%	19%	7%

Figure 2 - Jugement humain des similarités trouvées sur CDDB et FIP

On différencie le niveau 1 du niveau 2 selon leur signification : les simples clusters avec 2 artistes sont surtout expressifs de similarités directes, alors que les clusters avec 3, 4 ou 5 artistes expriment mieux les similarités indirectes. Globalement, ces résultats montrent que les artistes ou titres d'une même compilation ou apparaissant consécutivement dans un programme radio présentent de fortes similarités. Les similarités de base telles que les artistes communs et le genre basique sont reconnues, ce qui valide la technique en elle-même. En plus des similarités basiques on peut trouver des ressemblances de genres et de périodes (venant probablement de l'abondance des compilations "best of de telle année"). De plus, les similarités sont parfois plus métaphoriques et peuvent être liées à différents facteurs tels que l'orchestration (par exemple Eleanor Rigby et un quartet de Haydn) ou basées sur les noms des titres ou la véritable signification des paroles (par ex "Kiss" de Prince est proche de "Le baiser" d'Alain Souchon.).

3.2 Utilisation des Web Services

Les Web Services conçus pour simplifier l'échange d'informations entre différents systèmes s'inscrivent dans le contexte du Web sémantique qui vise en particulier une plus grande interopérabilité des systèmes informatiques en

réseaux. La récente popularité de ces technologies a incité beaucoup d'entreprises à ajouter des interfaces normalisées et des composants Web Services à leurs applications et bases de données.

Même si le Web peut être considéré comme un reflet de l'opinion publique, sa taille et son manque de structure rendent difficile la récupération et l'exploitation des données. Une partie de ces problèmes a été résolue par les moteurs de recherches qui facilitent l'accès à une grande quantité d'information. Auparavant, ces systèmes étaient accessibles uniquement via des interfaces de navigation manuelles. Les Web Services permettent aujourd'hui de rendre accessibles ces données par des applications. Ceci est particulièrement utile dans les processus de recherche d'information musicale qui peuvent utiliser, par exemple, les API Web de Google et d'Amazon [25, 26]. L'idée est de générer une liste d'artistes proches d'un artiste de référence à partir de métadonnées culturelles disponibles sur le web. Le principal avantage de cette approche pour déterminer des similarités est le fait qu'elle repose sur des tendances actuelles de l'opinion publique et se base sur une grande quantité de données. Ces listes d'artistes similaires pourraient être utilisées pour des recommandations ou pour étudier l'évolution des genres à travers le temps.

Le programme fonctionne comme suit : partant d'un artiste cible, on génère une liste d'artistes potentiellement similaires à partir des données d'Amazon. On génère ensuite une liste pour chaque artiste de cette liste jusqu'à ce que l'arbre formé atteigne une certaine densité. Le nombre total de résultats de Google est utilisé pour évaluer la réelle similarité entre deux artistes. La similarité est considérée comme étant le pourcentage de pages retournées pour un artiste donné dans lesquelles apparaît le nom de l'autre artiste. Cette métrique, bien que très simple, donne des résultats raisonnables. Ainsi, on génère une liste d'artistes semblables et on élimine ceux dont la similarité est inférieure à un seuil de similitude de 0.05 % établi empiriquement.

Cette approche a été testée par M. Zadel et I. Fujinaga [3] avec 5 artistes plus ou moins populaires et ayant des styles différents (musique électronique, metal, dance, pop, jazz). Dans chaque cas, le programme retourne avec succès une liste d'artistes qui semblent raisonnablement similaires. Pour chacun des 5 artistes, le processus dure environ deux minutes. Les résultats sont en général ordonnés par genre. On observe cependant un certain recouvrement probablement dû au fait que les métadonnées culturelles renferment plus que le simple genre. Les résultats sont globalement très positifs (exemple : chanteuses solo pop pour Christina Aguilera et artistes et groupes de jazz pour John Coltrane, avec des niveaux de similarités allant de 72% à 25 %).

La similarité est doublement implémentée ici à travers l'utilisation des bases de données d'Amazon et de Google. Il n'est pas évident qu'une méthode produise de bons résultats à elle seule. Les *Listmania!* d'Amazon regroupent des artistes selon un certain critère mais ce n'est pas forcément ce qu'un individu lambda attend (exemple : "Musique terrible des années 80", "Artistes nés un mardi"). De même, la méthode de calcul via Google fonctionne bien quand elle est utilisée avec des artistes fournis par la recherche avec Amazon. La performance serait certainement différente si la recherche s'effectuait avec des artistes pris au hasard.

Dans cette étude, seules deux sources sont utilisées mais en théorie on pourrait automatiser l'accès à des centaines de bases de données à travers Internet. On constate que même avec cet exemple utilisant seulement deux bases de données, une grande quantité de relations entre artistes peut être récupérée. Le point important dans cette expérience est que les Web services permettent d'accéder facilement à ces ressources, et qu'elles contiennent une quantité importante d'information.

4 Conclusion

La recherche de similarités recouvre potentiellement une grande diversité de techniques. La segmentation par genre est certainement le meilleur moyen d'organiser la musique mais elle se heurte au manque de norme qui pourrait fournir une hiérarchie "universelle" de genres. Les travaux actuels dans ce domaine utilisent différentes taxonomies, ce qui à long terme ne peut aboutir à un système standard de classification.

Les méthodes liées aux propriétés acoustiques sont relativement complexes et demandent des compétences en traitement du signal et en reconnaissance de formes. Les résultats fournis par ces méthodes sont en général prometteurs car la granularité de caractéristiques qu'il est possible d'extraire et d'assembler est potentiellement très importante. La difficulté se trouve dans l'identification d'un modèle liant ces éléments de bas niveau et la vision symbolique associée au genre. Par ailleurs, les modèles complexes impliquent des ressources et des temps de calcul parfois lourds. Les mêmes difficultés existent aussi à un degré plus important dans la caractérisation et l'indexation d'image ou de vidéo.

L'exploitation des données disponibles sur le Web reste un domaine ouvert. La difficulté est l'exploitation de données non structurées qui ne donne pas toujours des résultats fiables. Une plus grande discipline dans la structuration des données multimédia et notamment une généralisation de format de description (XML, RDF, etc.) seraient utiles. La solution se trouve sans doute dans l'exploitation complémentaire des métadonnées et des données audio.

Références

- [1] George Tzanetakis, George Essl, Perry Cook. Automatic musical genre classification of audio signals. *ISMIR*, 2001
- [2] Cory McKay, Ichiro Fujinaga. Automatic genre classification using large high-level musical. *ISMIR*, 2004
- [3] Mark Zadel et Ichiro Fujinaga. Web services for music information retrieval *ISMIR* 2004
- [4] François Pachet, Gert Westermann, Damien Laigre. Musical Data Mining for Electronic Music Distribution *ISMIR*, 2004
- [5] Scheirer, E.D. et Slaney, M. Construction and evaluation of a robust multifeature speech/music discriminator. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, 1331-1334, 1997
- [6] Wold, E., Blum, T., Keislar, D., et Wheaton, J. Content-based classification, search and retrieval of audio. *IEEE Multimedia*, 1996 3(2)
- [7] Foote, J., Content-based retrieval of music and audio. *Multimedia Storage and Archiving Systems II*, 138-147, 1997
- [8] Martin, K. Sound-Source Recognition : A theory and computational model. *PhD thesis, MIT Media Lab* <http://sound.media.mit.edu/~kdm>
- [9] Rossignol, S et al. Feature extraction and temporal segmentation of acoustic signals. *Proceedings of International Computer Music Conference (ICMC)*, 1998
- [10] Dubnov, S., Tishby, N., et Cohen, D. Polyspectra as measures of sound and texture. *Journal of New Music Research*, vol. 26, 1997
- [11] Scheirer, E. Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, 103(1):588-601
- [12] Goto, M. and Muraoka, Y. Music understanding at the beat level : real time beat tracking of audio signals. D.F. rosenthal and H. Okuno (ed.), Readings in *Computational Auditory Scene Analysis*, 156-176
- [13] Oppenheim, A. and Schaffer, R. Discrete-Time Signal Processing. Prentice Hall. Edgewood Cliffs, NJ. 1989.
- [14] Mallat, S, G. A wavelet tour of signal processing. *Academic Press*, 1999
- [15] Logan, B. Mel Frequency Cepstral Coefficients for music modeling. Read at the first *International Symposium on Music Information Retrieval*. <http://ciir.cs.umass.edu/music2000>
- [16] McKay, C. Automatic genre classification of MIDI recording. *Master's thesis*. McGill University, Canada
- [17] Fujinaga, I. Exemplar-based learning in adaptive optical music recognition system. Proceedings of the *International Computer Music conference*, 55-6, 1996
- [18] Hussein, F., R. Ward, and N. Kharma. Genetic algorithms for feature selection and weighting, a review and study. *International Conference on Document Analysis and Recognition*, 1240-4, 2001
- [19] Baumann, S., and O. Hummel. Using cultural metadata for artist recommendations. Proceedings of the *International Conference on Web Delivering of Music*, 138-41, 2003
- [20] Whitman, B., and S. Lawrence. Inferring descriptions and similarity for music from community metadata. Proceedings of the *International Computer Music Conference*, 591-8, 2002
- [21] Hill, W., L. Stead, M. Rosenstein, and G. Furnas. Recommending and evaluating choices in a virtual community of use. Proceedings of the *SIGCHI Conference on Human Factors in Computing Systems*, 194-201, 1995
- [22] Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens : an open architecture for collaborative filtering of netnews. Proceedings of the *ACM Conference on Computer Supported Cooperative Work*, 175-86, 1994
- [23] Shardanand, U., and P. Maes. Social information filtering : algorithms for automating "word of mouth". Proceedings of the *SIGCHI Conference on Human Factors in Computing Systems*, 210-7, 1995
- [24] Pilgrim, M. 2004. PyAmazon. Available online : <http://josephson.org/projects/pyamazon>
- [25] Pilgrim, M. 2004. PyGoogle. Available online : <http://pygoogle.sourceforge.net>
- [26] Calishain, T. and R. Dornfest. Google hacks. O'Reilly and Associates. 2003
- [27] E. Diday, G. Govaert, Y. Lechevallier, J. Sidi, Clustering in pattern recognition. In *Digital Image Processing*, page 19-58, J.C. Simon, R Haralick, eds, Kluwer edition, 1981
- [28] Perrot, D. and R. O. Gjerdingen. Scanning the dial: An exploration of factors in identification of musical style. *Research notes*. Department of Music, Northwestern University, Illinois, USA. 1999
- [29] Schütze, H. "Dimensions of Meaning", Proceedings of *Supercomputing*, pp. 787-796, Minneapolis, MN, 1992
- [30] Lowe, W. and MacDonald, S. "The direct route : Mediated priming in semantic space", Proceedings of the *22nd Annual conference of the Cognitive Science Society*, 2000
- [31] CARLING A. Introducing neural networks, *Sigma Press*, New York, 1992
- [32] N. Roussopoulos, S. Kelley, F. Vincent, Nearest Neighbor Queries. *SIGMOD Conference*, 1995:71-79