Catégorisation libre d'extraits musicaux et analyse automatique

Nicolas Dauban, Paul Albenge, Ludovic Florin, Julien Pinquier, Christine Sénac, Pascal Gaillard, Patrice Guyot

Catégorisation libre d'extraits musicaux et analyse automatique

Nicolas Dauban [1], **Paul Albenge** [2], **Ludovic Florin** [2], **Julien Pinquier** [1], **Christine Sénac** [1], **Pascal Gaillard** [2], **Patrice Guyot** [1]

1. Institut de Recherche en Informatique de Toulouse
118, route de Narbonne
31062 Toulouse, France
{nicolas.dauban,julien.pinquier,christine.senac,patrice.guyot}@irit.fr

2. Université Toulouse II Jean-Jaurès
5, allées Antonio Machado
31058 Toulouse, France
{pascal.gaillard,ludovic.florin,paul.albenge}@univ-tlse2.fr

RÉSUMÉ. *Cet article décrit le protocole expérimental et les résultats obtenus lors d'une expérience de catégorisation. Cette expérience s'inscrit dans le cadre de travaux de recherche sur la recommandation musicale personnalisée et basée sur le contenu. Durant cette expérience, les volontaires ont dû catégoriser librement des extraits musicaux sélectionnés selon des critères musicologiques. Cette catégorisation est analysée via un dendrogramme représentant la « classification moyenne des participants ». Une analyse automatique des résultats menée a posteriori vise à identifier les paramètres acoustiques déterminants dans cette classification moyenne.*

ABSTRACT. *This paper describes the experimental protocol and the results of a sorting experiment. This experiment was conducted as part of research works in content based and personalised music recommendation. During this experiment, volunteers had to freely categorize musical excerpts selected on musicological criteria. This categorization leads to a dendrogram which represents « the volunteer's average categorization ». An automatic analysis afterward aims at identifying relevant acoustic parameters based on the obtained categories.*

MOTS-CLÉS : *Catégorisation, Recommandation, Musique*

KEYWORDS: *Categorization, Recommendation, Music*

DOI:10.3166/RIA.28.1-16

1. Contexte et objectif de l'étude

Le rôle des algorithmes de recommandation musicale est de proposer de nouveaux morceaux aux utilisateurs des sites d'écoute de musique en ligne. Les travaux de recherche en Recommandation Musicale (RM) sont très récents et peu développés dans le monde académique, peut-être en raison de la limitation à l'accès - dûe à des problèmes de licence - au signal de musique à large échelle. Parce que baser une recommandation sur de simples métadonnées issues de filtres collaboratifs n'est pas vraiment efficace, de plus en plus de travaux s'appuient sur l'expertise acquise en Recherche d'Information Musicale (MIR) qui vise à extraire du signal des informations à différentes échelles (notes, accords, séquence de notes...) afin de caractériser par exemple un instrument ou à calculer des descripteurs tels que le tempo ou la mélodie principale. Voir (Schedl *et al*., 2014) pour un état de l'art récent sur la MIR.

Ainsi, certains auteurs ont essayé de s'appuyer sur la mesure de similarité entre morceaux de musique (McFee *et al*., 2012). Si cette approche est pertinente pour la classification en genres (tâche la plus proche de la recommandation musicale), elle est assez décevante pour la RM. Aussi, était-il naturel d'introduire, en parallèle d'approches basées contenu, des informations sur les préférences des utilisateurs (Oord *et al*., 2013), (Humphrey *et al*., 2012), ou le comportement de l'utilisateur (Schedl, Hauger, 2015). Cependant, quelle que soit la méthode, les morceaux peu connus (situés dans la 'long tail') ne sont jamais (ou rarement) proposés : certains travaux visent à remédier à ce problème (Celma Herrada, 2009), (Domingues *et al*., 2013) .

L'une des principales difficultés soulevée par les méthodes basées contenu est la sélection des paramètres. En effet, parmi tous les paramètres que nous pouvons extraire d'un signal audio, lesquels permettent de décrire et d'expliquer les affinités des auditeurs? Comment pouvons-nous lier ces paramètres acoustiques à des paramètres musicologiques? Ces problématiques sont les enjeux majeurs du projet dans lequel s'inscrit cette expérience de catégorisation. Le but de cette expérience est d'identifier à la fois les paramètres acoustiques et les critères musicologiques ou « non-experts » selon lesquels les sujets classent les morceaux, en partant du postulat que ce sont les mêmes combinaisons de paramètres ou de critères qui déterminent les goûts d'un auditeur.

2. Paramètres acoustiques

La musique désigne communément une suite d'événements sonores (notes, sons percussifs, sons voisés ou non voisés) définis par leur rythme, leur timbre, leur dynamique et éventuellement leur hauteur.

2.1. Rythme

Le rythme décrit la localisation temporelle des événements sonores ainsi que leur durée. Dans la musique occidentale conventionnelle, une pulsation régulière déter-

mine les temps, une mesure étant composée de plusieurs temps. Dans une partition, le rythme est décrit par les différentes figures de notes (croche, noire, blanche...) et de silences (pause, soupir...) ainsi que par le chiffrage.

2.1.1. Détection et densité des événements

Tous les paramètres rythmiques s'appuient dans un premier temps sur la localisation temporelle de chaque événement. Pour cela, nous utilisons un algorithme de détection de pics sur l'enveloppe du signal. Une fois que les pics sont détectés, nous pouvons ensuite calculer le nombre d'événements par seconde.

2.1.2. Tempo

Le calcul du tempo s'appuie sur une détection de la périodicité des événements, et sélectionne le pic le plus élevé. La détection de périodicité s'effectue à l'aide de la fonction d'auto-corrélation (Lartillot, 2014).

2.1.3. Clarté de la pulsation

La clarté de la pulsation (*pulse clarity*) peut être calculée selon la méthode détaillée dans (Lartillot *et al*., 2008). Ce paramètre décrit à quel point la pulsation est dominante dans le rythme, ou autrement dit, à quel point l'accent est mis sur les temps (par exemple, la clarté de la pulsation est forte pour des rythmes disco, et est souvent faible pour des rythmes complexes, comme ceux du jazz).

2.2. Timbre

Le timbre décrit la composition spectrale d'une note, c'est-à-dire l'amplitude des harmoniques et la variation dans le temps de ces harmoniques. C'est ce qui distingue par exemple deux notes jouées à la même hauteur par un piano et une guitare.

2.2.1. Attaque

L'attaque d'une note décrit la variation d'amplitude à l'instant où elle est jouée. Elle se mesure par sa durée, son amplitude ou bien par sa pente (Grey, 1975). Par exemple, les cordes frappées du piano ont une attaque plus forte que les cordes frottées du violon. Quand une note est détectée, il faut repérer le début et la fin de l'attaque, puis calculer la différence d'amplitude ou la durée entre les deux points (la pente est obtenue à l'aide de ces deux informations).

2.2.2. Taux de passage par zéro (ou ZCR : Zero Crossing Rate)

Le taux de passage par zéro est calculé sur le signal original en faisant la multiplication de toutes les paires d'échantillons successifs, et en itérant une variable lorsque le produit est négatif (changement de signal). Nous divisons ensuite par la durée pour obtenir le taux (Gouyon *et al*., 2000).

2.2.3. Fréquence de roulement (ou Rolloff frequency)

La fréquence de roulement nous informe sur la quantité d'énergie présente dans les basses fréquences. Sur un spectre, nous calculons la fréquence en dessous de laquelle 85% de l'énergie est contenue. Plus cette fréquence est basse, plus l'énergie est concentrée dans les basses fréquences.

2.2.4. Brillance (ou Brightness)

La brillance nous informe sur la quantité d'énergie présente dans les hautes fréquences (Laukka *et al*., 2005). Sur un spectre, nous calculons la quantité d'énergie présente au delà d'une fréquence fixée (en général $1500Hz$).

2.2.5. Paramètres statistiques de la distribution spectrale

Il est possible de calculer des statistiques ainsi que des moments de différents ordres sur le spectre : centroïde (barycentre), spread (étalement), skewness (assymétrie), kurtosis (courbure), flatness (platitude...), ainsi que l'entropie.

2.2.6. MFCC (Mel Frequency Cepstral Coefficients)

Les MFCC sont des coefficients cepstraux calculés par une transformée en cosinus discrète appliquée au spectre de puissance d'un signal (Logan *et al*., 2000). Les différentes bandes de fréquences sont déterminées selon l'échelle logarithmique perceptive Mel qui est calquée sur le système d'audition humain.

2.2.7. Dissonance sensorielle (ou Roughness)

La dissonance sensorielle décrit le phénomène de « battement » audible en présence de deux fréquences proches (Sethares, 2005). Deux notes espacées d'un demi ton (ou moins) généreront une forte dissonance sensorielle, qui diminue à mesure que l'espacement augmente. La dissonance sensorielle est quasi-nulle à partir de 5 demitons.

2.2.8. Irrégularité du spectre

L'irrégularité d'un spectre est le degré de variation d'amplitude de deux pics (harmoniques ou non) successifs du spectre (Lartillot, 2014).

2.3. Dynamique

La dynamique décrit l'amplitude relative des différents sons, ce qui se traduit par des nuances d'intensité. Sur une partition, la dynamique est indiquée par des termes comme « pianissimo » ou « forte » qui indiquent au musicien de jouer plus ou moins fort. En traitement du signal et de manière générale, la dynamique décrit la plage de variation des différentes valeurs prises par un signal. En musique, la dynamique décrit le rapport des sons d'amplitudes fortes et faibles.

2.3.1. Niveau RMS

La valeur efficace d'un signal aléatoire ergodique sur un intervalle temporel est la racine carrée de la moyenne de ce signal au carré, ou autrement dit, la racine carrée de sa puissance moyenne. Dans la pratique, pour un signal à temps discret, le niveau RMS est calculé sur un nombre fini d'échantillons.

2.3.2. Taux de faible énergie (ou Low Energy Rate)

Le taux de faible énergie correspond au nombre de points dont la valeur est inférieure à la valeur RMS du signal. Pour un signal comportant des pics au niveau RMS élevé, ce taux sera élevé alors que pour un signal au niveau RMS plutôt constant, ce taux sera faible.

2.4. Hauteur

La hauteur décrit la fréquence fondamentale d'un son joué par un instrument, qui définit la note jouée.

2.4.1. Détection de notes

La méthode utilisée, par défaut, pour détecter les notes (ou « pitch ») est de décomposer le signal en plusieurs bandes de fréquences, de calculer ensuite l'autocorrélation et enfin de détecter les pics afin d'obtenir une estimation des notes.

2.4.2. Détection d'harmonies

À partir de la détection de notes, il est ensuite possible de détecter des harmonies, c'est-à-dire des combinaisons de différentes notes. Il est également possible de calculer la tonalité d'un extrait, ainsi que l'évolution temporelle de tous ces paramètres.

3. Protocole expérimental

3.1. Constitution du corpus

L'une des premières étapes du projet a été de constituer un corpus, qui devait répondre à plusieurs exigences :

- présenter un large panel de genres musicaux,
- disposer d'extraits de bonne qualité : CD Audio (stéréo, 16 bits, 44.1 kHz),
- posséder des extraits suffisamment longs ($\simeq$ 20 s) et suffisamment nombreux,
- utiliser, de préférence, une base de données en accès libre de droits.

Le corpus a été constitué d'un point de vue musicologique. Nous avons, dans un premier temps, listé un ensemble de quinze critères définissant de manière la plus exhaustive possible la musique.

Qualité de l'enregistrement : Perception du support et médium d'enregistrement (bruits, spectre sonore, intensité...).

Prédominance d'un instrument : Saillance d'un timbre particulier.

Voix : Présence ou absence de voix, type de voix (parlée, chantée, déclarative, répétitive, sonore...).

Espace : Sensation et représentation d'un espace de diffusion de la musique, profondeur de l'espace musical.

Travail mémoire : Présence d'un ou plusieurs éléments mémorisables, répétition d'un élément (stricte ou analogue), perception claire d'une forme ou d'une logique.

Dynamique : Changement de quantité/densité d'événements, contraste dans le développement musical.

Développement narratif : Évolution d'éléments musicaux, présence de différentes parties relativement distinctes.

Lisse/strié : Présence ou absence de pulsation, diversité des éléments, variation dans la quantité d'éléments (dans un temps court).

Sensori-motrice : Musiques d'instrumentistes, animées en grande partie par un désir du geste et par une recherche de l'effet sensoriel du son (on pourra donner comme exemple des musiques africaines, des improvisations de percussion, des solos de jazz ou des musiques concrètes de Pierre Henry).

Représentation : Qui représente de façon visible ou cachée le réel, sur le plan plastique, par des trajectoires, des vitesses, des chocs ou même des bruits réalistes (grégorien, romantisme occidental, beaucoup de musiques contemporaines).

Règle : Toutes musiques d'écriture qu'elles soient réellement écrites, comme le contrepoint classique, ou transmises de mémoire comme les polyphonies pygmées ou les jeux de trompes M'baka.

Énergie : Intensité, implication corporelle, implication du musicien.

Niveau technicité : Deux pôles : instrumental et composition. Perception de l'assurance des intentions du musicien et/ou présence de concepts structurels.

Éléments culturels : Référence à une classe socio-culturelle.

Situation chronologique : Perception d'éléments propres à une époque : type d'enregistrement, type de jeu, référence à une esthétique particulière.

Pour chacun de ces critères, nous avons sélectionné trois morceaux significatifs. Les extraits choisis contiennent différentes caractéristiques musicales qui permettent de proposer un ensemble éclectique. Bien que ceux-ci eussent été sélectionnés pour correspondre initialement à un type précis, il est possible de retrouver des caractéristiques d'autres types. Au final, un corpus de 45 extraits de 20 secondes a été défini (voir Figure 1).

Extrait	Début	Fin	Titre
1	00:00	00:20	Les Doubles Six - Au Bout du Fil (Meet Benny Bailey)
2	00:02	00:22	Bill Bruford - One Of A Kind (Part 1)
3	00:00	00:20	Harry Burleigh - Go Down Moses
4	03:20	03:40	Rautavaara - Cantus Arcticus 2e mvt
5	00:50	01:10	Hector Berlioz - Symphonie Fantastique, Op. 14, Songe d'une nuit de sabbat
6	00:00	00:20	Corette - Concerto pour musette de cour 2 Adagio
7	00:00	00:20	Namibie Chant De Guerison - Nom Tzisi
8	00:00	00:20	Han Bennink & Willem Breuker - Mr. M.A. de R. in A.
9	00:00	00:20	Death Grips - Thru The Walls
10	00:00	00:20	Jazzoo - le pic et le moineau
11	00:24	00:44	Big Satan - Geez
12	00:02	00:22	Lords Of The Underground - Here Come The Lords
13	00:02	00:22	Pygmées Aka
14	00:00	00:20	Deux Chants De Jeu Et De Danse - Polynésie Occ.
15	00:10	00:30	James Brown - Mother Popcorn
16	03:00	03:20	Suisse Yodel - Zauerli
17	00:13	00:33	Horace Silver - Capverdian Blues
18	03:10	03:30	Awa Poulo - Dimo Yaou Tata
19	00:00	00:20	Pharoah Sander - Love Will Find a Way
20	00:00	00:20	David Fiuczynski - Moonring Bacchanal
21	00:00	00:20	André Minvielle - L'Alambic
22	00:25	00:45	Edgard Varèse - Un Grand Someil Noir
23	00:30	00:50	The Residents - This Is Man's World
24	00:28	00:48	Theo Bleckmann & Ben Monder Late Green
25	03:40	04:00	Aphex Twin - Circlont14 [Shrymoming Mix]
26	00:00	00:20	Naked City - Une Correspondance
27	05:00	05:20	Bugge Wesseltoft - Dreaming
28	00:40	01:00	Ali Farka Touré - Sabu Yerkoy
29	00:46	01:06	A Ram Sam Sam
30	00:05	00:25	Sleepytime Gorilla Musuem - The Putrid Refrain
31	08:20	08:40	John Zorn - Through The Night
32	00:20	00:40	Liadov : Baba Yaga
33	00:28	00:48	Don Ellis - Strawberry Soup
34	00:30	00:50	Ligeti - Quatuor à cordes n°2 - come un meccanismo di precisione
35	00:27	00:47	Arvo Pärt "Ludus" du Tabula Rasa
36	00:20	00:40	John Zorn, Filmworks - Cynical Hysterie Hour Through the
37	01:50	02:10	Jaco Pastorius - Come On, Come Over
38	00:00	00:20	Bach/ Glenn Gould - The Art of the Fugue, BWV 1080- Contrapunctus III
39	02:30	02:50	Julien Loureau - Conrod
40	00:25	00:45	Dayton - Krackity Krack
41	00:43	01:03	Aka Moon - For Drummers Only
42	00:15	00:35	Sec - Run Away
43	00:23	00:53	Tool - Lateralus
44	00:00	00:20	Bruckner - scherzo 9e symphonie
45	00:30	00:50	Naked City - Speedball

Figure 1. Liste des morceaux de musique et minutage des extraits.

3.2. Conditions expérimentales

Afin de limiter l'impact de l'âge des participants sur les résultats, nous avons fait appel à des volontaires de 20 à 25 ans (au nombre de 30). Pour l'expérience, nous avons utilisé l'outil TCL-labX [1] (Gaillard, 2009). L'interface (voir Figure 2) était présentée de manière identique à tous les volontaires à qui il était demandé de trier librement des extraits et de former ainsi autant de catégories qu'ils le souhaitaient, en se basant sur les similarités entre les morceaux. Pour cela, les utilisateurs pouvaient écouter autant de fois que nécessaire les extraits et pouvaient les déplacer et les regrouper librement sur l'interface.

1. http://petra.univ-tlse2.fr/spip.php?article255

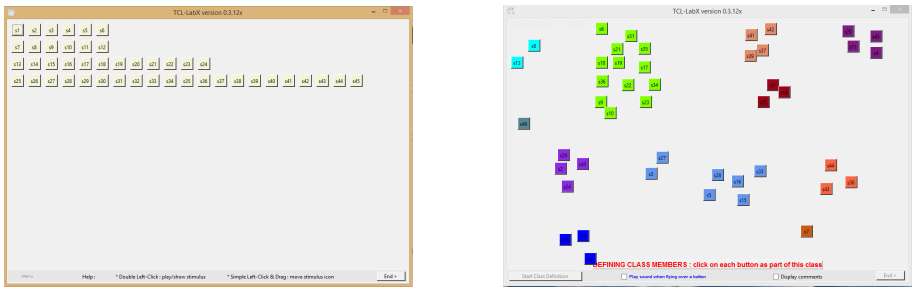

Figure 2. Interface présentée au départ de la passation, et après le tri.

4. Catégorisation libre et interprétation

4.1. Données

Pour chaque volontaire, le programme génère un fichier dans lequel est indiquée la répartition des extraits dans les différentes classes. Le logiciel génère également un fichier « mouchard » contenant l'historique des opérations effectuées par l'utilisateur: déplacement des icônes et écoute des extraits. Nous pouvons ainsi rejouer toutes les actions effectuées par le volontaire. De plus, le logiciel permet d'effectuer directement une analyse de ces fichiers et ainsi d'extraire plusieurs statistiques sur les participants. La durée moyenne de l'expérience a été de 37 minutes, la durée maximale a dépassé une heure (1h 2min) et la durée minimale fut de 15 minutes. L'écart-type sur la durée de l'expérience est de 10 minutes. En moyenne, les participants ont formé 15,5 classes, le minimum est de 8 classes et le maximum 20. L'écart-type sur le nombre de classes est de 3,2.

4.2. Analyse des résultats

4.2.1. Matrices de co-occurrence et de dissimilarité

Pour réaliser cette tâche, nous nous sommes basés sur les travaux de (Barthélémy, Guénoche, 1988). Dans un premier temps, nous avons construit une matrice de co-occurence C^i pour chaque participant i. Une matrice de co-occurence est carrée et symétrique, de dimension $N \times N$ avec N égal au nombre d'extraits triés. Dans chaque cellule, nous indiquons la distance entre deux extraits : si ces deux extraits sont dans la même catégorie, la distance est considérée comme nulle, sinon nous assignons une distance unité.

Nous calculons ensuite une matrice de co-occurence moyenne de tous les participants, appelée matrice de dissimilarité D. Cette matrice de dissimilarité nous donne une mesure de distance pour chaque paire d'extraits musicaux, cette distance étant basée sur la classification par les n participants.

Nous avons également calculé une matrice de variance M_{var} à partir des matrices de co-occurence C^i et de la matrice de dissimilarité D (voir équation 1). Pour une paire d'extraits, cette variance est nulle si l'ensemble des n participants ont trié d'une manière identique ces extraits. À l'inverse, cette variance est maximale ($var_{max} = 1$) si la moitié des participants ont placés ces extraits ensemble, et l'autre moitié séparément.

$$M_{var}^{j,k} = \frac{1}{n} \sum_{i=1}^{n} (C^{i,j,k} - D^{j,k})^2 \tag{1}$$

avec $M_{var}^{j,k}$ la cellule de la ligne j et colonne k de la matrice M_{var}, $C^{i,j,k}$ la cellule (j, k) de la matrice C^i et $D^{j,k}$ la cellule (j, k) de la matrice D.

Du fait que les matrices C^i ne contiennent que des valeurs binaires, pour une paire d'extraits, une variance nulle correspond forcément à une dissimilarité de 1 ou 0, et une variance unité correspond forcément à une dissimilarité de $0,5$. Plus la dissimilarité est proche d'une valeur extrême (1 ou 0), plus ces deux extraits auront fait « l'unanimité » dans leur classification. Plus la dissimilarité est proche de $0,5$, plus ces deux extraits auront « divisé l'opinion » des volontaires.

4.2.2. Dendrogramme

À partir de la matrice de dissimilarité, nous avons pu effectuer une classification hiérarchique ascendante : nous avons choisi d'utiliser la *méthode de Ward* (Ward Jr, 1963) qui consiste à regrouper les classes afin que l'augmentation de l'inertie interclasse (cf. équation 2) soit maximale. ou ce qui revient au même d'après le *théorème de Huygens*, afin que l'augmentation de l'inertie intraclasse (cf. équation 3) soit minimale (Saporta, 2006).

$$I_e = \frac{1}{n} \sum_{i=1}^{k} n_i \times d^2(g_i, g) \tag{2}$$

$$I_a = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} d^2(e_j, g_i) \tag{3}$$

Nous avons obtenu le dendrogramme de la Figure 3. Il nous renseigne sur les liens établis entre les morceaux par les utilisateurs. L'axe vertical représente la distance entre les morceaux ou groupes de morceaux. Ainsi, deux morceaux ayant été très souvent placés ensemble par les volontaires ont une liaison basse, comme par exemple le 11 et le 17 ou le 15 et le 37.

4.3. Interprétation musicologique

Le dendrogramme obtenu précédemment a été interprété d'un point de vue musicologique afin de comprendre comment les volontaires avaient effectué cette classification. Le dendrogramme a été annoté avec les critères communs aux morceaux appartenant à la même branche, voir Figure 3.

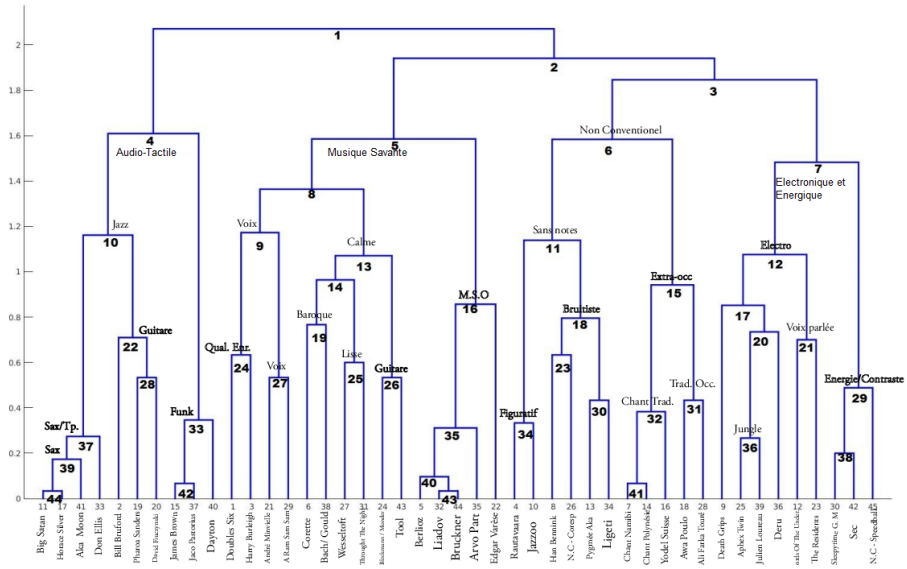

Figure 3. Dendrogramme annoté par les musicologues.

Le nom indiqué sous chaque nœud correspond à un critère musicologique commun à tous les morceaux qui sont situés en dessous. Les quatre grandes catégories sont décrites ci-dessous.

1. **Audio-tactile** : les musiques présentes dans cette catégorie sont toutes très rythmées et appartiennent au genre jazz ou funk et plus globalement à tendance afro-américaine. L'audiotactilité désigne un rapport particulier avec le corps. À l'intérieur de cette catégorie, les morceaux ont étés distingués par l'instrument prédominant.

2. **Musique savante** : cette catégorie est un hybride regroupant d'une part la Musique Savante Occidentale, d'autre part la musique évoquant un aspect sacré ou spirituel et enfin les morceaux où la voix est prédominante.

3. **Non Conventionnel** : cette catégorie regroupe la musique construite en dehors des règles conventionnelles régissant la musique occidentale notamment basé sur la mélodie et l'harmonie. Nous y retrouvons, par conséquent, les musiques sans hauteur de notes précises et les musiques extra-occidentales. Toutes musiques ne suivant pas les hiérarchies présentent dans les codes occidentaux sont forcément perçue comme un groupe à part.

4. **Électronique et Énergique** : la plupart des extraits de cette catégorie ont étés produits à partir d'instruments et de traitements électroniques. Cette catégorie regroupe également des morceaux présentant des contrastes en terme d'énergie. Cependant, la distance reste particulièrement conséquente entre ces deux sous-catégories.

L'objectif n'était pas de retrouver dans les résultats de la passation la classification créée mais de voir quelles dimensions sont les plus prégnantes à l'écoute. Nous pouvons remarquer que les critères musicologiques qui ont servi à établir le corpus ne se retrouvent pas à travers la classification des participants non-experts. Cela illustre différents types d'analyses et d'appréciations musicales. Les critères musicologiques nous ont permis d'obtenir un corpus très varié et les catégories identifiées par les participants révèlent d'autres critères plus accessibles pour des non-experts. C'est sur ces derniers que nous allons nous appuyer pour effectuer une classification automatique.

5. Vers une classification automatique

5.1. Sélection des paramètres acoustiques

Dans un premier temps, nous avons calculé 31 paramètres audio sur chaque morceau du corpus à l'aide de MIR Toolbox (Lartillot, 2014). L'extraction de ces paramètres est décrite en détail dans la partie 2.

Nous avons moyenné chaque paramètre afin d'obtenir une matrice de la forme $N \times P$ avec $N = 45$ (morceaux) et $P = 31$ (paramètres). Notons qu'en ne conservant qu'une moyenne, nous perdons l'évolution temporelle, mais cela nous permet de n'avoir qu'un seul scalaire par morceau et par paramètre. Pour chaque paramètre, nous avons pu calculer la distance pour chaque paire de morceaux et ainsi former une matrice de dissimilarité P^i pour chaque paramètre i. Ensuite, nous avons voulu établir un modèle de la matrice de dissimilarité de la passation à partir d'une combinaison linéaire des matrices des paramètres :

$$M_{modele} = \sum_{i=1}^{31} a_i P^i \tag{4}$$

Les valeurs contenues dans chaque matrice de dissimilarité ont été normalisées entre 0 et 1 afin de rester cohérentes avec les valeurs de la matrice de la passation. Plutôt que d'utiliser toutes les matrices P^i, nous avons sélectionné les matrices les plus pertinentes en calculant le coefficient de corrélation de chaque matrice de paramètres avec la matrice des volontaires et nous avons sélectionné les N plus corrélées. En effet, les matrices les plus corrélées avec la matrice de dissimilarité établie lors de la passation sont par définition les plus « ressemblantes ».

5.2. Régression

5.2.1. Matrice complète

Avec ces N premières matrices, nous avons utilisé un algorithme de descente de gradient afin de trouver la meilleure combinaison linéaire, le critère à optimiser étant l'erreur quadratique entre cette combinaison linéaire et la matrice de dissimilarité de la passation. Cet algorithme renvoie donc les coefficients a_i par lesquels sont multipliées les matrices de dissimilarité afin d'obtenir la matrice la plus ressemblante à la matrice de dissimilarité formée par l'ensemble des résultats des volontaires. Ces coefficients nous informent de l'importance de chaque paramètre : si un coefficient est faible alors il est peu influent pour les volontaires pour trier les morceaux, et inversement. Afin de simplifier les calculs, les matrices de dissimilarité ont été transformées en vecteurs V_p et V_d de longueur $L = 45 \times 45 = 2025$.

Pour N matrices de dissimilarité de paramètres utilisées, l'équation de l'erreur quadratique est définie par :

$$J = \sum_{j=1}^{L} \left[\left(\sum_{i=1}^{N} a_i V_p^{i,j} \right) - V_d^j \right]^2 \tag{5}$$

Le gradient de cette erreur est :

$$\overrightarrow{\text{grad}}\, J = \begin{bmatrix} \frac{\partial J}{\partial a_1} \\ \dots \\ \frac{\partial J}{\partial a_k} \\ \dots \\ \frac{\partial J}{\partial a_N} \end{bmatrix} \tag{6}$$

avec :

$$\begin{array}{ll} \frac{\partial J}{\partial a_k} & = \sum_{j=1}^{L} \frac{\partial \left[\left(\sum_{i=1}^{N} a_i V_p^{i,j} \right) - V_d^j \right]^2}{\partial a_k} \\ & = 2 \sum_{j=1}^{L} \left[\left[\left(\sum_{i=1}^{N} a_i V_p^{i,j} \right) - V_d^j \right] V_p^{k,j} \right] \end{array} \tag{7}$$

Nous avons testé successivement l'algorithme avec les N « meilleurs » paramètres, au sens de la corrélation. En augmentant successivement le nombre de paramètres, l'erreur quadratique totale diminue jusqu'à 220. À partir de 7 paramètres, l'erreur augmente de nouveau. Les 6 premiers paramètres sont : *Irrégularité, Brillance, Rolloff, Détection de changement d'harmonie, Entropie du spectre, Attaque.*

À partir de la matrice estimée avec 6 paramètres, nous avons généré un nouveau dendrogramme (cf. Figure 4) afin de pouvoir comparer visuellement le résultat de cette estimation avec le dendrogramme obtenu à l'issue de la passation (cf. Figure 3): nous pouvons constater une ressemblance assez éloignée entre ces deux dendogrammes. Ceci peut s'expliquer par le fait que les participants n'ont pas utilisé la même « règle » pour classer tous les extraits. Par exemple, certains extraits ont été groupés par rapport à des similarités rythmiques, et d'autres vis-à-vis de leur mélodie. Ainsi il est

difficile d'établir une règle générale sur les paramètres pour estimer avec une bonne précision la classification globale. Cela est sûrement dû au fait que les extraits présentaient une grande variabilité. Néanmoins, si nous considérons des sous-parties du dendrogramme, les extraits appartenant à chacune de ces sous-parties sont plus similaires entre eux, et nous pouvons donc supposer qu'il sera plus facile d'isoler des paramètres discriminants pour chacune de ces sous-parties.

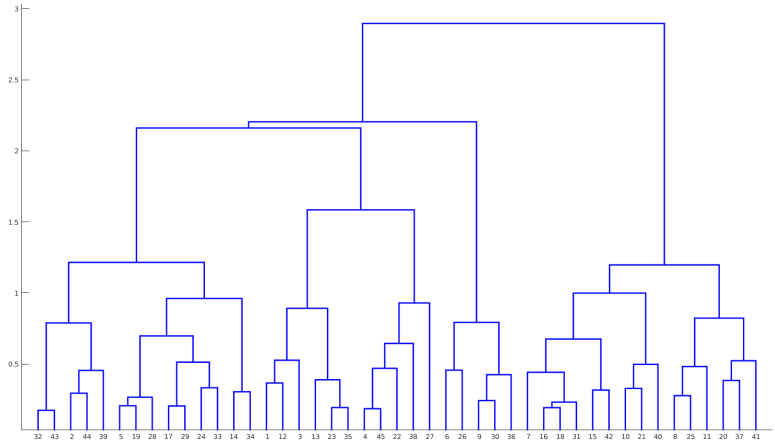

Figure 4. Dendrogramme estimé à partir de 6 paramètres.

5.2.2. Sous matrices/parties

Nous avons utilisé la même méthode que précédemment sur chacune des 4 sous-parties, nommées par les musicologues : Audio-Tactile, Musique Savante, Non conventionnel et Électronique/Énergique. Pour chacune de ces parties, nous avons recalculé les critères les plus corrélés, et les avons utilisés à nouveau pour former des combinaisons linéaires.

1. **Audio-Tactile** : Les paramètres les plus corrélés sont : *Brillance, Irrégularité, MFCC(10), Attaque, MFCC(4), Entropie du spectre, Clarté de la pulsation, Taux de passage par zéro, Low energy, Kurtosis du spectre, MFCC(3), Rolloff, Tempo, MFCC(8).*

À l'issue de la descente de gradient, l'erreur quadratique totale est de 5,5. Nous remarquons que plusieurs MFCC interviennent dans la classification. Nous pouvons expliquer cela par le fait que pour cette catégorie, les volontaires ont beaucoup distingué les extraits selon les instruments prédominants. Le dendrogramme a été bien reconstitué, mis à part pour les extraits 5 et 6 qui ont été échangés.

2. **Musique Savante** : pour cette catégorie, les résultats étaient plutôt mitigés, même en utilisant tous les paramètres. En effet, à l'issue de la descente de gradient, l'erreur quadratique totale est de 9,5. Ces résultats plus faibles peuvent s'expliquer par

le fait que cette catégorie contient des extraits disparates, il est donc plus difficile de généraliser une règle de catégorisation.

3. **Non conventionnel** : ici, la méthode a fourni de bons résultats avec 18 paramètres. À l'issue de la descente de gradient, l'erreur quadratique totale est de 3,8. Les paramètres les plus corrélés sont : *Dissonance sensorielle, Attaque, Nombre d'événements par seconde, Taux de passage par zero, Kurtosis du spectre, Clarté de la clé, Entropie, MFCC(8).*

4. **Électronique/Énergique** : c'est pour ce groupe que la méthode a été la plus performante (cf. Figure 5). Nous avons obtenu une erreur quadratique totale de 2,8. Nous avons utilisé les 10 paramètres suivants (du plus corrélé au moins corrélé) : *Attaque, MFCC(3), MFCC(8), MFCC(11), Rolloff, Brillance, MFCC(4), Taux de passage par zéro, MFCC(0) (i.e. énergie).*

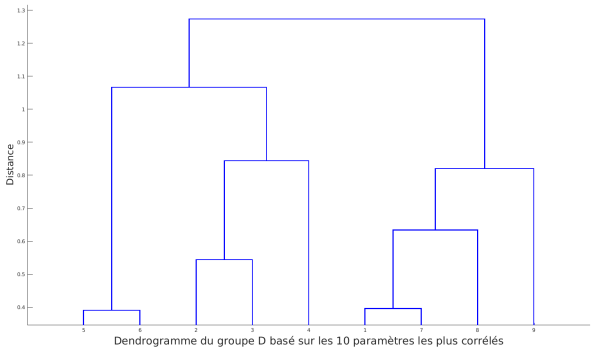

Figure 5. Dendrogramme obtenu pour la catégorie « Electronique/Energique » avec 10 paramètres. Excepté pour le premier extrait, le dendrogramme de ce groupe a été bien reconstitué. Pour chaque sous-groupe i de taille N_i, les indices initiaux des morceaux ont été remplacés par des indices allant de 1 à N_i. Ainsi, si le dendrogramme d'un sous-groupe a bien été reconstitué, les individus sont placés dans l'ordre croissant.

Globalement, les résultats sont plutôt satisfaisants car nous avons ainsi pu reconstituer les dendrogrammes de chacune des catégories avec un nombre d'erreurs limité.

5.3. Classification de nouveaux extraits

L'objectif de cette partie était de trouver une méthode permettant d'attribuer à un « nouvel » extrait la bonne catégorie. Dans toutes les méthodes qui suivent, nous avons considéré successivement chaque extrait comme un nouvel individu, en prenant soin de le retirer de la base d'apprentissage. Le score de chaque méthode est donc compris entre 0 et 45 (cas où tous les extraits ont étés attribués aux bonnes catégories). De plus, les paramètres ont été centrés réduits afin de supprimer l'influence de l'unité de mesure utilisée pour chacun d'eux.

La première méthode a consisté à calculer le barycentre de chaque catégorie selon les 31 paramètres, et à attribuer ensuite le nouvel extrait à la classe ayant son barycentre le plus proche: nous avons ainsi obtenu 30 attributions correctes (soit environ 67%).

Dans la seconde méthode, nous avons retenu un nombre restreint de paramètres: nous avons observé quels paramètres étaient pertinents pour la classification dans les sous catégories (section 5.2.2) et nous avons conservé seulement ceux qui étaient les plus corrélés dans les 4 classifications. Il s'agit de *Détection de changement d'harmonie, Attaque, Entropie du spectre, Rolloff MFCC(0) et Irrégularité*. Nous avons ainsi obtenu 22 attributions correctes (33%): les paramètres sélectionnés n'étaient donc pas particulièrement pertinents.

Dans la troisième méthode, nous avons utilisé les N paramètres les plus corrélés en établissant le classement de la même manière que dans la section 5.2.1. Sur la Figure 6, nous voyons que le score augmente globalement avec le nombre de paramètres mais qu'il diminue parfois lorsque nous en utilisons un nouveau. Le score maximal (32 attributions correctes, soit 71%) est atteint pour 25 paramètres. Cette méthode s'est donc avérée être la meilleure.

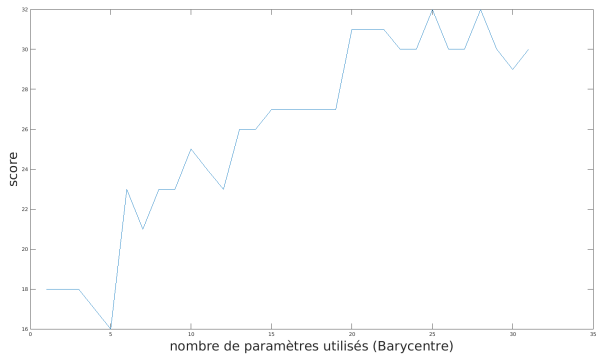

Figure 6. Score d'attribution en fonction du nombre de paramètres utilisés.

6. Conclusions et perspectives

L'analyse des résultats de la passation nous a permis d'établir une classification moyenne des morceaux par les volontaires qui a été représentée sous la forme d'un dendrogramme dans lequel apparaissent quatre groupes ainsi que des sous-groupes.
Afin de reconstruire de manière automatique cette classification humaine, nous avons établi une hiérarchie dans la pertinence des paramètres selon leur corrélation avec la classification des volontaires. Nous avons vu que cette reconstruction automatique est plus efficace pour distinguer les sous-groupes à l'intérieur d'un groupe, plutôt que les groupes entre eux. Au final, les paramètres identifiés pourront être privilégiés pour une application en recommandation musicale.

Bibliographie

Barthélémy J. P., Guénoche A. (1988). Les arbres et les représentations de proximité. *Paris. Dunod (english translation: Trees and Proximity Representations, New York, Wiley, 1991)*.

Celma Herrada Ò. (2009). *Music recommendation and discovery in the long tail*. Thèse de doctorat non publiée.

Domingues M. A., Gouyon F., Jorge A. M., Leal J. P., Vinagre J., Lemos L. *et al*. (2013). Combining usage and content in an online recommendation system for music in the long tail. *International Journal of Multimedia Information Retrieval*, vol. 2, n° 1, p. 3–13.

Gaillard P. (2009). Laissez-nous trier ! tcl-labx et les tâches de catégorisation libre de sons. , p. 189-210.

Gouyon F., Pachet F., Delerue O. (2000). On the use of zero-crossing rate for an application of classification of percussive sounds. In *Proceedings of the cost g-6 conference on digital audio effects (dafx-00.*

Grey J. M. (1975). *An exploration of musical timbre using computer-based techniques*. Department of Psychology, Stanford University.

Humphrey E. J., Bello J. P., LeCun Y. (2012). Moving beyond feature design: Deep architectures and automatic feature learning in music informatics. In *Ismir*, p. 403–408.

Lartillot O. (2014). *Mirtoolbox 1.6.1 user's manual.*

Lartillot O., Eerola T., Toiviainen P., Fornari J. (2008). Multi-feature modeling of pulse clarity: Design, validation and optimization. In *Ismir*, p. 521–526.

Laukka P., Juslin P., Bresin R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, vol. 19, n° 5, p. 633–653.

Logan B. *et al*. (2000). Mel frequency cepstral coefficients for music modeling. In *Ismir*, vol. 270, p. 1–11.

McFee B., Barrington L., Lanckriet G. (2012). Learning content similarity for music recommendation. *IEEE transactions on audio, speech, and language processing*, vol. 20, n° 8, p. 2207–2218.

Oord A. Van den, Dieleman S., Schrauwen B. (2013). Deep content-based music recommendation. In *Advances in neural information processing systems*, p. 2643–2651.

Saporta G. (2006). *Probabilités, analyse des données et statistique*. Editions Technip.

Schedl M., Gómez E., Urbano J. *et al*. (2014). Music information retrieval: Recent developments and applications. *Foundations and Trends® in Information Retrieval*, vol. 8, n° 2-3, p. 127–261.

Schedl M., Hauger D. (2015). Tailoring music recommendations to users by considering diversity, mainstreaminess, and novelty. In *Proceedings of the 38th international acm sigir conference on research and development in information retrieval*, p. 947–950.

Sethares W. A. (2005). *Tuning, timbre, spectrum, scale*. Springer Science & Business Media.

Ward Jr J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, vol. 58, n° 301, p. 236–244.