

Auteur 1 *Affiliation*

Auteur 2 *Affiliation*

Auteur 3 *Affiliation*

Une approche info-communicationnelle des librairies musicales en ligne

Résumé

Les SIC étudient, depuis leur origine, les médias et les systèmes d'accès à ces derniers (Jeanneret, 2007). Dans un contexte d'accélération et de croissance exponentielle des sources d'informations mises à disposition, le rapport à ces sources devient de plus en plus crucial afin d'en garantir l'accès, la disponibilité ou encore le suivi. La présente étude porte sur les librairies musicales en ligne, des sites agencés par des moteurs d'indexation et de recherche qui mettent à disposition des professionnels de l'audiovisuel des médias musicaux spécialement réalisés pour être synchronisés à des images. De l'avis de ces utilisateurs, ces solutions semblent néanmoins lacunaires. Dans une nouvelle approche exploratoire remettant l'humain au centre du processus de qualification - à la base de toute indexation - nous essayons de résoudre le problème en mêlant diverses approches.

Introduction

Cette étude porte sur les catalogues de médias musicaux spécialisés pour une utilisation audiovisuelle. Les utilisateurs cibles sont ceux que l'on appelle music supervisors ; dans un contexte de production, leur rôle est de rechercher, en somme, la bonne musique pour les bonnes images. Les plateformes en ligne, proposant ces musiques, sont pléthores et peuvent contenir dans leur base plusieurs milliers de médias, ce qui est non préhensible à un agent humain, surtout dans le contexte d'une recherche en temps et moyens limités. Elles sont donc structurées et rendues accessibles par un moteur d'indexation et de recherche. Considérées comme non efficaces par leurs utilisateurs (bruitées, silencieuses, coûteuses en temps, ...) (Chai et Vercoe, 2000)¹, ces solutions semblent donc amendables (Yvart, 2013).

La question se pose donc de savoir d'où pourraient venir ces améliorations et plusieurs axes s'offrent en réponse. D'une part, une étude comparative des librairies fournit des éléments de diagnostic du ou des problèmes rencontrés (Yvart, 2013 ; Li, Ogihara, 2004)² en s'intéressant au « côté » interface. D'un autre, une étude des processus informationnels et décisionnels intervenant dans la recherche du média et dans son exposition au sujet peut aussi révéler des orientations inédites. De fait, en reprenant à la base le processus de qualification de la musique, on intervient avant même l'indexation et la recherche.

Nous proposons ici une approche originale qui remet l'auditeur au cœur du processus de qualification au sens de la méthode de conception relativisée (Mugur-Schächter, 2006). Toute la solution au problème réside dans le passage de ces construits de sens personnels à des cons-

¹ One problem of current music information retrieval systems is performance. Usually the computational complexity in these systems is huge, while real-time response is required in most cases.

² Traditionally musical information has been retrieved and/or classified based on standard reference information, such as the name of the composer, the title of the work, the album title, the style of music, and so on. These basic pieces of information will remain essential, but information retrieval based purely on these is far from satisfactory.

truits partageables, puis à des aspects qualifiants recueillant un degré de consensus intersubjectif suffisant sur une population cible afin de passer du sens à la signification (Rastier, 2001).

Une étude exploratoire, adaptée de la méthodologie MEDIA-REPERES (Labour, 2011), portant sur un public de 12 coparticipants écoutant 3 extraits musicaux visait à expliciter leurs construits de sens à l'écoute au travers de questionnaires. A la suite d'un dépouillement par proximité sémantique (Gaume, 2004), il a été possible de constituer des clusters de sens (Lamprier et al., 2010) convergents permettant de tendre vers ce consensus intersubjectif porteur de signification.

Contexte

Dans le web sémantique (3.0) (Corby et al., 2005), les moteurs de recherche se basent sur un processus *a priori* d'indexation (Ertzscheid, 2008 ; Tat, 2011) des fichiers afin de pouvoir construire les réseaux de correspondance entre fichiers (similitude, distance, ...) sans avoir à analyser toute la base de données à chaque requête.

Ces critères sémantiques d'indexation correspondent aux métadonnées des fichiers eux-mêmes supports de l'information. Il en existe 2 types³ :

1. les critères fondés sur le contenu (positivistes) : ex. le rythme exprimé en BPM ;
2. les critères fondés sur l'auditeur (anthropocentrés et constructivistes) : ex. l'émotion (Chai et Vercoe, 2000).⁴

Les premiers critères sont étudiés majoritairement par les sciences « dures », notamment le traitement du signal au travers de programmes internationaux comme l'ISMIR ou le MIREX qui regroupent et diffusent différents algorithmes permettant de déterminer des qualifiants à granularité très différentes (ex. rythme, tonalité, instrumentation). Certains de ces algorithmes donnent de très bons résultats sachant que d'autres sont encore en développement. Ainsi, par exemple, la reconnaissance d'une instrumentation complexe n'est pas garantie, à l'heure actuelle, passé une demi-douzaine d'instruments.

Dans la seconde famille de critères, certains des aspects pouvant entrer en jeu, comme l'émotion, ont déjà été grandement abordés par les sciences humaines, que ce soit au travers du prisme de la psychologie (Debaecker, 2012 ; Yvart, 2004), ou encore de l'économie et du marketing (Alpert et Alpert, 1990). Même les programmes fondés sur le contenu s'intéressent à cette question et essaient de faire ressortir du signal des émotions suscitées (Kim et al., 2010) en s'appuyant sur certaines heuristiques (ex. une tonalité mineure est triste).

Par opposition aux critères « objectifs » (positivistes) ceux-ci ne peuvent rencontrer qu'un certain degré de consensus intersubjectif. De plus, leur étude implique de remettre l'auditeur au cœur de tout processus d'étiquetage (Schedl et Flexer, 2012).

Matériau

D'un point de vue médiologique, un extrait musical est un document qui contient une information de type audiophonique musicale. C'est donc un sous-ensemble des documents audiophoniques (émission de radio, voix enregistrée, etc. ...) qui a cette caractéristique musicale,

³ Il existe en plus de ceux-ci un troisième type de métadonnées celles que nous nommerons légales qui portent sur l'auteur, le titre, les droits et les différentes affiliations. Dans le contexte de l'étude elles sont peu intéressantes.

⁴ Music features can be classified as quantitative and qualitative. Although the classification is not absolute, we use these in this paper for the sake of simplicity. - Quantitative features : background information (composer, artists, age, genre, etc), time signature, key signature, tonality, tempo, instruments, music structure, pitch, loudness, melody, rhythm, all the other parameters that could be derived from the music, etc. - Qualitative features : rating, functionality, emotional features, etc.

mais la question se pose de savoir si ce contenu particulier en fait un média à part, qu'il s'agirait donc de traiter, dans une perspective archivistique (par exemple) d'une manière qui lui soit propre.

Sans entrer dans un débat ontologique sur ce qu'est la musique et ce qu'est une information musicale, on peut déjà poser le fait - unanimement reconnu - que la musique est un concept purement humain qui, outre le fait qu'il soit totalement artificiel, ne connote ni ne dénote de rien (Johnson-Laird et Oatley, 2008)⁵. Sémantiquement parlant, un objet musical n'a pas de référent.

Pourtant, malgré cela, la musique est vectrice d'une information compréhensible et communicable suivant bien des supports. Bien qu'en elle-même la musique ne soit que l'information particulière transportée par une onde sonore dans un fluide, cette dernière peut être captée, encodée, stockée puis transportée et rendue par divers moyens technique.

Le fichier informatique convoyant la musique, bien que non musical en-lui-même, est qualifié comme étant musical au vu d'un auditeur. Cette qualification est subjective et soumise à un consensus intersubjectif valable dans un certain lieu à un certain moment.⁶

Au sein même de ce qui est objet musical, il existe un contenu informationnel, lui, à la base de la compréhension de ce dernier (de la construction de sens chez l'auditeur). C'est cette qualification de contenu qui est intéressante et même primordiale dans un contexte d'indexation et de recherche du média et qui est soumise, comme la qualification de ce qui est musique ou pas, à un nécessaire consensus suffisamment fort pour être signifiant. L'enjeu est d'étiqueter au mieux le média musical pour pouvoir, comme c'est le cas pour les textes, le chercher au moyen de l'information qu'il porte.

Dans l'approche constructiviste, on essaye, par un processus statistique, de discriminer les étiquettes socialement valides (consensuelles : ex. 95% des sujet trouvent que telle musique est triste) des étiquettes qui ne le sont pas (purements subjectives : ex. cette musique me fait penser à mes dernières vacances).

Les deux approches (positiviste et constructiviste) sont complémentaires. Ainsi, pour juger de paramètres relevant de la structure même du morceau : rythme, instrumentation, etc. ... une approche positiviste semble la meilleure⁷. En revanche, pour ce qui est du contenu émotionnel de la musique : est-elle triste ? joyeuse ?... Seul un humain semble apte à déterminer cette qualification, et cela amène donc une approche constructiviste, d'autant plus que la musique n'exprime rien de concret, n'a pas de référent.

L'émotion véhiculée, ou l'humeur de l'objet musical, semble, de toute première importance dans une configuration de recherche de musique (Bigand et al., 2005⁸, Kim et al. 2010⁹). Près d'un tiers des 427 sujets d'une étude dont la question était : « Lorsque vous recherchez de la musique ou de l'information musicale, lesquelles des options de recherche / navigation êtes

⁵ Pure music is without propositional content. By "pure" music, we mean music that is not a setting for words — that does not follow a program such as Richard Strauss's *Til Eulenspiegel's Merry Pranks*, or that does not depict a series of events such as Claude Debussy's *La Mer*, with its three movements entitled, "From dawn till noon on the ocean", "Play of the waves", and "Dialogue of wind and sea".

⁶ Cela se traduit par l'existence en des moments spatio-temporels différents de plusieurs concepts de qualification de ce qui est musique ou non. Ainsi on retrouve bien que la « définition » de la musique n'est pas la même dans l'Europe du XVIe ou du XXe ou encore entre l'Europe du XXIe et l'Asie du XXIe.

⁷ Que la détermination des qualificants soit faite par un humain ou non d'ailleurs.

⁸ [An] important feature of the current study is the provision of evidence that categorising musical excerpts on the basis of emotional experience transcends the usual criteria of musical categorisation, such as surface similarities, musical instrumentation, and musical style.

⁹ Music is oftentimes referred to as a "language of emotion" [...] and it is natural for us to categorize music in terms of its emotional associations. Myriad features, such as harmony, timbre, interpretation, and lyrics affect emotion, and the mood of a piece may also change over its duration.

vous le plus susceptible d'utiliser ?¹⁰ ont répondu par la teneur émotionnelle (Lee et Downie, 2004). Cette importance (cette quasi-prévalence dans le cas des music supervisors (Yvart, 2013)) se retrouve aussi dans le fait que 15% des utilisateurs de « lastfm » aiment découvrir de la musique en naviguant via le mood¹¹ (Barthet, Fazekas, et Sandler, 2012)¹².

Dans cet article nous traiterons justement de la qualification de cette « teneur » émotionnelle, de l'humeur de la musique et des possibilités, à partir d'une constellation de qualifiants exprimés de retrouver une convergence menant à un consensus intersubjectif et à une minimisation des termes en vue d'une computation possible (par attachement des qualifiants sous forme de métadonnées par exemple).

Dans cette nouvelle approche, plus spécifiquement info-communicationnelle, le mood est préféré à l'émotion (Meyer, 2008)¹³. Il s'agit donc de déterminer si un mood commun et consensuel se dégage et permet la qualification.

Le média musical, par extension, son information n'ont pas d'humeur *stricto sensu*. L'humeur de la musique, serait idéalement, un transfert de l'état émotionnel interne de l'auditeur sur l'objet musical. Il attribuerait au stimulus extérieur l'humeur que ce dernier lui induit. D'ailleurs c'est à ce niveau que l'on discerne l'utilisation de mood (humeur) de celle d'émotion. L'humeur étant vue comme un état long et stable accessible à la conscience et qualifiable (ex. je me sens triste), semble mieux convenir que l'émotion qui est de nature plus impulsionnelle et difficilement reportable en temps réel.

L'émotion induite par la musique est ainsi vue comme la variation d'une humeur H à un instant t vers une humeur H' à un instant t+n. La mesure par introspection de l'humeur par l'auditeur peut être retenue comme à la base de sa qualification. Le résultat de l'introspection n'étant disponible qu'à la conscience de l'auditeur (Barrett, 1996)¹⁴, le seul moyen d'y accéder est l'explicitation par ce dernier de son vécu par le biais du report personnel de l'humeur (self-report) et ceux, même si cela induit certaine biais inhérents à la méthode.

Méthode

Une expérimentation, à valeur d'étude exploratoire, a donc été menée sur 12 sujets représentatifs d'une population encline à prospecter de la musique pour la synchroniser à des documents audiovisuels.¹⁵ L'étude commençait par la complétion d'un questionnaire visant à caractériser le passif musical des auditeurs¹⁶.

¹⁰ "When you search for music or music information, how likely are you to use the following search/browse options?" – Traduction personnelle.

¹¹ Le terme anglais mood définit l'humeur suscitée. Il a été préféré au terme français par le fait qu'il est plus présent dans la communauté scientifique et technique, mais aussi dans les agences de publicité où l'on parle couramment de mood design. L'humeur étant plus partageable et longue en durée que l'émotion elle n'en est que plus facilement mesurable, explicitable et donc analysable.

¹² "The importance of musical mood metadata was further confirmed in the investigations by Lesaffre et al. which give high importance to affective/ emotive descriptors, and indicate that users enjoy discovering new music by entering mood-based queries, as well as those by Bischoff et al. which showed that 15% of the song queries on the web music service Last.fm were made using mood tags."

¹³ "Few psychologists dealing with music have been as accurate on this point as Weld, who notes that : "The emotional experiences which our observers reported are to be characterized rather as moods than as emotions in the ordinary sens of term [...] The emotion is temporary and evanescent ; the mood is relatively permanent and stable."

¹⁴ "For better or worse, self-report represents the most reliable and possibly only window that researchers have on conscious, subjective emotional experience."

¹⁵ Ces derniers étaient étudiants en dernière année d'un cursus en audiovisuel et multimédia et avaient suivi un cours spécifique à l'introduction au métier de music supervisor.

¹⁶ Il leur était demandé notamment : - leur éveil musical : questionnement sur leurs plus anciens souvenirs musicaux ; l'âge de début éventuel d'une pratique musicale ; les styles de musiques qu'ils ont écoutés. - leur éducation musicale « scolaire » : conservatoire et apprentissage du solfège ; apprentissage d'un instrument ; res-

Un deuxième questionnaire était ensuite distribué, répétant trois fois un set de questions ouvertes visant à qualifier par adjectifs les humeurs des extraits. Pour limiter le temps de l'expérience il a été choisi de diffuser des extraits de 31 secondes pris au début des œuvres suivantes¹⁷ :

- Zoé Keating - We insist (A)
- Yunus – Longing (B)
- Mindthings - Our lives, our destinies (C)¹⁸

Sur ces durées de temps, les morceaux avaient la particularité de laisser place à une courte introduction puis à un thème maintenu pouvant être pris comme une unité homogène¹⁹. Les trois morceaux étaient écoutés à la suite puis chaque extrait était repassé le temps de remplir la partie correspondante du questionnaire. Aucunement il n'a été indiqué ni le titre ni l'artiste pour ne pas influencer la qualification des « émotions ou impressions provoquées par l'extrait »^{20 21}.

Résultat

Diverses approches ont été utilisées afin de tenter de faire converger les lexèmes annotés par les sujets vers un nombre plus restreint de qualifiants, et moins de qualifiants subjectifs, permettant potentiellement une computation efficace. Une première approche, que nous appellerons par occurrence était un comptage simple du nombre de fois qu'un mot était utilisé. Une deuxième utilisait un algorithme de proxémie *ad hoc*²². Enfin, à partir de ces proximités nous avons déterminé des clusters de mots proches pointant vers un mot supposé le plus consensuel²³.

Approche par occurrence

Une première phase a donc consisté en le fait de mesurer le nombre d'occurrences de chaque terme et ainsi de déterminer le degré de consensus par la fréquence. De ce point de vue, l'hypothèse suivie pose que plus la fréquence d'un terme est élevée, plus ce terme est vu comme consensuel et donc signifiant. Il s'agit, de fait, d'une analyse purement quantitative des données.

senti éventuel. - leur éducation musicale sociale ; pratique d'un instrument dans une formation ; la musique dans leur entourage ; fréquence et type de concerts. - leur pratique éventuelle et leur formation à un instrument ou au chant : autodidacte ou scolaire. - leur capacité physiologique : éventuelles affections de l'oreille ; amusies partielles ou totales ; capacité à apprécier de la musique ; oreille absolue ou relative. - l'importance de la musique pour eux : importance émotive notamment ; - leur consommation (commerciale) musicale : situation ; fréquence ; genre ; achat ou piratage ; physique ou numérique. - leur consommation de musique d'illustration et sa motivation : usage éventuel de librairie musicale ; retour sur expérience. - leur ouverture (ou fermeture) aux différents mouvements musicaux : préjugés sur X ou Y genre ; leur préféré ; leur détesté ; justifié si possible. - leur culture générale musicale : leur éclectisme ; ...

¹⁷ Ces œuvres ont été prises sur le site jamendo.com car libres de droit et typiquement orientées pour la synchronisation audiovisuelle.

¹⁸ Notons que par souci d'écologie, la musique était diffusée par l'intermédiaire d'un système de qualité monitoring équivalent en qualité à ceux des music supervisors.

¹⁹ Cela est nécessaire à la qualification. Ainsi, si un morceau est composé de plusieurs unités hétérogènes les unes par rapport aux autres, il faudrait en toute théorie décomposer et qualifier chaque sous-partie homogène pour obtenir une qualification efficace.

²⁰ Formulation de la question.

²¹ Un troisième questionnaire tentait, quant-à-lui, de mettre en place une analyse de la construction de sens par une grille triadique suivant la méthodologie MEDIA-REPERE (Labour, 2011).²¹ Cependant cette phase a posé problème du fait que nous ne procédions pas à un entretien semi-dirigé mais à un questionnaire distribué en groupe et que, du fait, la complétion des grilles était ardue. Les résultats de cette partie ne seront donc pas traités dans ce papier.

²² Développé sur Matlab par Benjamin Picart (TCTS) et Willy Yvart.

²³ Notons dès à présent que la consigne posant l'utilisation des adjectifs a été quelque peu ignorée. Cela, dans le cadre d'un traitement informatisé, a réduit sensiblement l'efficacité du processus.

Pour l'extrait A, il a été donné en tout 50 adjectifs et substantifs²⁴ différents. Seuls quatre étaient répétés²⁵, et ce, avec des fréquences relativement faibles :

Mot	Occurrences	Fréquence
Stress	5	45.45%
Suspens(e)	3	27.27%
Peur(s)	2	18.18%
Tristesse	2	18.18%

Tableau 1 : Qualifiants de l'extrait A

Mot	Occurrence	Fréquence
Tristesse	4	33.33%
Mélancolie	3	25.00%
Mélancolique	3	25.00%
Amertume	2	16.67%
Amour	2	16.67%
Triste	2	16.67%

Tableau 2 : Qualifiants de l'extrait B²⁶

Mot	Occurrence	Fréquence
Dramatique	3	25.00%
Angoissant	2	16.67%
Angoisse	2	16.67%
Oppression	2	16.67%
Peur	2	16.67%

Tableau 3 : Qualifiants de l'extrait C²⁷

Tel quel, il semble impossible de conclure quoi que ce soit. Aucune fréquence n'est suffisante pour que le qualifiant correspondant fasse consensus. De plus on ne perçoit que très finement le rapprochement entre les extraits pourtant sélectionnés comme proches à la base.

Pour essayer de mettre en avant cette proximité ressentie lors de la sélection, il a été mis en place une comparaison des lexèmes deux à deux. Ce rapprochement se faisait par un processus informatisé sensible à la fois à l'orthographe mais aussi à la forme du mot. Il a donc fallu rapprocher à la main les mots équivalents de formes différentes (ex. triste/tristesse) :

Communs 1-2	Communs 1-3	Communs 2-3
Tristesse	Tristesse	Tristesse
Dynamique	Suspens	Mélancolie
Effort	Peur	Triste
Espoir	Dramatique	Colère
	Gravité	Incertain
	Oppression	Lent
	Puissant	Perdu
	Tourmenté	

Tableau 4 : Comparaison des qualifiants communs deux à deux²⁸

²⁴ Cela montre une mauvaise compréhension de l'énoncé de la question sans pour autant à notre sens remettre en cause le fait que l'on qualifie plus facilement par adjectifs qualificatifs la musique (Meyer, 2008).

²⁵ Précisons que l'on ne reprend que les mots ayant au moins deux occurrences

²⁶ Sur 50 adjectifs différents.

²⁷ Sur 40 adjectifs différents. On peut penser que le fait qu'il y ait eu moins d'adjectif pour ce dernier extrait soit dû à une fatigue croissante du panel.

²⁸ En orange les mots commun étant en occurrence multiples, en vert les mots communs uniques.

La base était le dictionnaire DicoSyn³³ créé à partir de sept dictionnaires classiques³⁴ et se présentant sous la forme d'une double matrice de 49133 mots dans l'une et d'une matrice carrée de cette dimension remplie de « 1 » et de « 0 » selon que les mots étaient synonymes ou non.

Il fallait donc dans un premier temps réattribuer à chaque lexème ses synonymes³⁵. Puis reprendre les qualificatifs mais cette fois en parcourant pour chaque mot ses synonymes et les synonymes du mot comparé. Dans une première approche on mettait « 1 » ou « 0 » selon qu'il y avait une relation d'intersynonymie ou non (non pondérée) et dans une seconde approche nous reportions le « poids » de cette intersynonymie représenté par le nombre de synonymes communs (intersynonymie pondérée).

On part ainsi de l'hypothèse que deux termes seraient d'autant plus proches (plus fortement liés l'un à l'autre) qu'ils auraient un grand nombre de synonymes communs³⁶. Cela donne des résultats très différents relativement à l'approche quantitative^{37,38} :

Non pondéré		Pondéré	
Mot	Intersyn	Mot	Intersyn
Mouvement	15	Etrange	44
Effort	12	Mouvement	41
Obsession	12	Curieux	39
Profond	12	Profond	35
Curieux	11	Magique	28

Tableau 5 : Intersynonymie extrait A³⁹

Comme on peut le constater ici, le fait de prendre en compte la pondération de l'intersynonymie change grandement le résultat. Le mot le plus représenté par fréquence « stress » n'est plus du tout présent alors qu'il avait la plus grande fréquence.

Non pondéré		Pondéré	
Mot	Intersyn	Mot	Intersyn
Difficultés	21	Tristesse	112
Emotion	21	Peine	107
Mélancolie	20	Mélancolie	93
Effort	20	Emotion	83
Amour	19	Amertume	79

Tableau 6 : Intersynonymie extrait B⁴⁰

Non pondéré		Pondéré	
Mot	Intersyn	Mot	Intersyn
Noir	21	tristesse	109
Triste	21	triste	85
Mal	16	mal	68
Tristesse	15	douleur	64
Dramatique	14	mélancolie	46

Tableau 7 : Intersynonymie extrait C⁴¹

³³ CRISCO, Université de Caen.

³⁴ Bailly, Benac, Du Chazaud, Guizot, Lafaye, Larousse et Robert

³⁵ Faisant par exemple passer la liste des 50 mots de l'extrait 1 à un tableau de 1704 mots

³⁶ Nous posons que le fait de regarder sous cet angle la synonymie peut permettre de réduire le bruit induit par la polysémie de certains termes.

³⁷ On ne représente que les cinq premiers dans l'ordre décroissant.

³⁸ Les mots jaunis sont en multiples occurrences dans les questionnaires.

³⁹ Moyenne / Ecart-type : 6.04/3.48 – Moyenne / Ecart-type (pondéré) : 12.8/10.92

⁴⁰ Moyenne / Ecart-type : 10.28/6.01 – Moyenne / Ecart-type (pondéré) : 31.68/27.92

⁴¹ Moyenne / Ecart-type : 7.65/5.54 – Moyenne / Ecart-type (pondéré) : 23.28/26.58

Ici, on voit l'une des limites de la méthode, le « stress » est certes fréquent mais du fait qu'il a peu de synonymes (2 dans dicosyn), il est discriminé par l'algorithme de proxémie. C'est un biais informatique dans cette approche qualitative.

Cela pose une question infocommunicationnelle pure, est-ce qu'un mot porte plus d'information s'il a beaucoup de synonyme ou au contraire ?⁴²

Puis il a été réintroduit, une seconde pondération consistant en le fait de multiplier le poids d'intersynonymie par la fréquence statistique donnant une approche mêlant le quantitatif et le qualitatif :

Mot	Intersyn	Occurrence	I*O
Etrange	44	1	44
Mouvement	41	1	41
Curieux	39	1	39
Peur	19	2	38
Profond	35	1	35

Tableau 8 : Extrait A - avec occurrence^{43,44}

On voit ici que le mot de multiple occurrence « peur » est remonté dans le classement grâce à cette pondération.

Mot	Intersyn	Occurrence	I*O
Tristesse	112	4	448
Mélancolie	93	3	279
Amertume	79	2	158
Amour	72	2	144
Mélancolique	47	3	141

Tableau 9 : Extrait B - avec occurrence⁴⁵

On trouve pour l'extrait B que cette pondération fait tendre cette liste de qualifiants vers la liste du dépouillement statistique, redonnant toute son importance à l'agent humain (en reprenant en compte les fréquences quantitatives de chaque terme). Il en va de même pour le C :

Mot	Intersyn	Occurrence	I*O
Angoisse	55	2	110
Tristesse	109	1	109
Triste	85	1	85
Dramatique	28	3	84
Mal	68	1	68

Tableau 10 : Extrait C - avec occurrence⁴⁶

En prenant en compte les moyennes et les écart-types on détermine globalement si les qualifiants sont consensuellement fiables :

⁴² Rappelons que l'information naît d'une différence qui crée une différence (Bateson et Drosso, 1995 ; Bateson, Lot, et Simion, 2008 ; Wittezaele, 2006). Même dans une approche Shannonnienne (Shannon, 1948 ; Shannon, 1964) ou néguentropique (Leleu-Merviel, 1997), la valeur d'une information est proportionnelle à sa rareté car proportionnelle au logarithme de l'inverse de la probabilité d'occurrence. Est-ce qu'un mot avec peu de synonymes est plus rare qu'un mot avec beaucoup de synonymes ?

⁴³ Une nouvelle fois par souci de présentation nous ne montrons que les cinq premiers de chaque tableau.

⁴⁴ Moyenne/Ecart-type : 13.62/11.64.

⁴⁵ Moyenne/Ecart-type : 48.26/77.30.

⁴⁶ Moyenne/Ecart-type : 27.2/31.09.

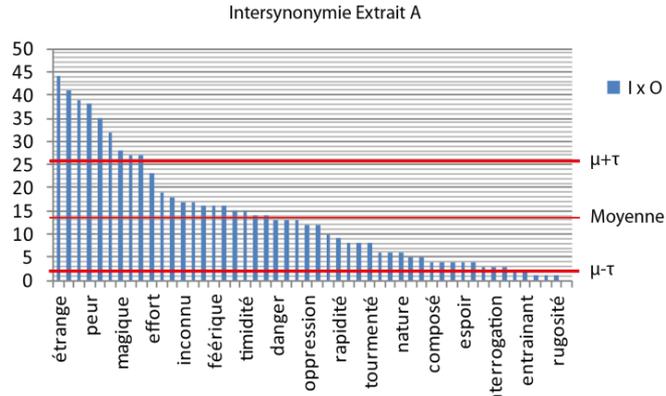


Figure 2 : Extrait A avec moyenne et écart-type

On constate que les quatre premiers qualificatifs sont au-delà de zone $(\mu+2\tau)$ ⁴⁷ ce qui les rend statistiquement significatifs pour qualifier l'extrait.

De ce point de vue, l'extrait se qualifie comme étrange, suscitant le mouvement, curieux, et provoquant la peur. Si on applique la même logique à l'extrait B, on obtient 2 qualificatifs qui dépassent la zone limite posée ci-avant⁴⁸, à savoir la tristesse et la mélancolie. Enfin pour le troisième on trouve encore deux qualificatifs qui se détachent⁴⁹, l'angoisse et la tristesse.

Approche par cluster

Pour aller plus loin, il est possible, en partant de la proxémie, de déterminer des clusters de termes pour qualifier l'humeur de l'extrait musical par le mot « cime ». On reprend les qualificatifs avec leur pondération puis on relie par calcul chacun d'eux au terme le plus proche mais ayant un « poids » supérieur. Ce faisant, nous créons des clusters chapeautés par les termes de poids les plus forts vers lesquels tous les autres termes du cluster pointent par synonymie directe ou par intersynonymie.

Pour l'extrait A on arrive à caractériser deux clusters représentés ci-après :

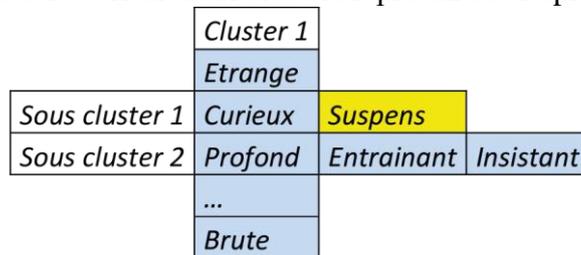


Figure 3 : Cluster "Etrange" Extrait A (partiel)⁵⁰

Ce premier cluster comporte deux branches, des sous-clusters qui permettent de relier un terme synonyme uniquement du terme de la colonne « tronc » mais pas de la cime du cluster. Cette présentation à l'avantage de faire remonter le terme « suspens » qui lui aussi avait été discriminé (comme « stress ») alors qu'il était de fréquence statistique élevée.

⁴⁷ Moyenne+2*Ecart-type = 36.91.

⁴⁸ Moyenne+2*Ecart-type = 202.85.

⁴⁹ Moyenne+2*Ecart-type = 89.38.

⁵⁰ Les mots en jaune sont ceux d'occurrence multiple dans l'approche statistique.

	Cluster 2	
	Mouvement	
Sous cluster 1	Changement	Interrogation
	Légèreté	
	Effort	
	Peur	
	...	
	Tristesse	
	...	
	Stress	
	Regard	
	...	

Figure 4 : Cluster "Mouvement" extrait A (partiel)

Encore une fois on nomme ce cluster par sa tête et on remarque la présence d'un sous-cluster montrant encore un synonyme propre à un terme du tronc mais pas à la cime.

On notera que le terme « inoffensif » n'a pu être intégré à aucun des deux clusters. Cela est dû techniquement au fait que « inoffensif » n'est synonyme d'aucun des autres termes. On peut aussi penser à une qualification aberrante⁵¹ par l'un des sondés. Si ce mot avait été pris pour chercher le fichier, ce dernier n'aurait pu être trouvé ; on aurait donc eu un silence du système.

Autre constat : on remarque que la différenciation entre les deux clusters est non-pertinente pour un humain. En effet, le cluster « étrange » devient le cluster des adjectifs et le cluster « mouvement » celui des substantifs. Certains des termes passés à leur substantif équivalent (ou *a contrario* leur adjectif) redeviennent membres d'un même cluster. Cet écart non calculable par ordinateur provient d'une sensibilité à la forme.⁵²

	Cluster 1	
	Tristesse	
	Peine	
Sous-cluster 1	Mélancolie	Calme
Sous cluster 2	Emotion	Incertain ...
	Amertume	
Sous cluster 3	Amour	Hésitant ...
	Triste	
	...	

Figure 5 : Cluster "Tristesse" Extrait B (partiel)

Pour l'extrait B, on arrive à un seul cluster dominant alors que de la même manière la consigne « adjectif » avait été ignorée. Cela vient d'une disposition de la langue française qui fait que certains termes peuvent être substantifs et adjectifs (comme dynamique par exemple). Néanmoins cette confusion informatique, bien que donnant en apparence un gain induit, en fait du bruit car « dynamique » (adjectif) n'a pas du tout le même sens que « dynamique » (substantif). De même, pour « déstructuré » qui est absent de tout cluster, son exclusion vient d'une difficulté supplémentaire du français qui fait que pour un ordinateur c'est un participe passé, alors qu'il prend valeur d'adjectif dans un contexte de qualification.

⁵¹ Au sens de ne rendant pas compte de l'émotion induite par la musique.

⁵² Cela doit être pris en compte dans la conception de tout moteur d'indexation et de recherche orientée vers l'humain, il faut traduire le langage naturel en quelque chose de plus formel et cela passe aussi par le fait de travailler sur des termes de même nature (après une traduction adéquate).

	Cluster 1	
	Tristesse	
Sous-cluster 1	Triste	Inconfortable ...
	Mal	
	Douleur	
	Mélancolie	
	Noir	
	Angoisse	
	...	

Figure 6 : Cluster "Tristesse" Extrait C (partiel)

Pour l'extrait C on se retrouve dans le même cas de figure et avec les mêmes limites et remarques que précédemment. Notons enfin, que comme dans l'approche proxémique simple les extraits B et C semblent plus proches l'un de l'autre que de l'extrait A.

Discussion

Gain de la méthode

On voit dans cette étude que bien qu'à la base les sujets aient été mis en condition de qualification libre (c'est-à-dire sans limitation par une liste à la Prévert) l'outil informatique permet, à l'aide de technique statistiques, proxémiques et par cluster, de ressortir (avec une efficacité certes amendable) un set restreint (et donc computable) de qualificants.

Sans cela, l'extrait A, par exemple, était qualifié par 50 idiomes différents dont le plus consensuel n'était répété que 5 fois. On avait donc une information « utile » faible noyée dans un bruit assez important. On remarque même que le mot de plus forte fréquence (ex. « stress ») n'est pas toujours celui vers lequel tendent le plus de termes.

Seconde chose intéressante, notons qu'il semble naître un lien rapprochant B et C donnant qu'ils sont plus proches l'un de l'autre que de A.⁵³

Biais expérimentaux

Rappelons ici que l'écoute des trois médias était enchaînée. Si l'on suit la théorie de la transmission de l'excitation (Zillmann, 1971) on peut penser qu'il y a une contamination possible de l'excitation (arousal) d'une écoute à l'autre. Ainsi si le média A active grandement le sujet cette activation persiste lors de l'écoute de B et ainsi de suite. On précisera que cet arousal est hédoniquement neutre, c'est-à-dire qu'il n'affecte pas la valence (positive ou négative) mais seulement l'intensité. On peut ici donc se demander si la proximité entre B et C ne viendrait pas d'un tel parasitage.

Outre cela, un seul « label » ne semble pas suffisant pour dénoter de l'humeur induite par un objet musical (Schmidt, Turnbull, et Kim, 2010)⁵⁴ et c'est justement ce que l'on retrouve dans cette étude. Instinctivement, on peut penser que le système sera d'autant moins soumis au piège de la polysémie des termes que le nombre de qualificants sera restreint (voire réduit à un seul mot) et que ces derniers seront univoques, introduire plusieurs labels pourrait donc, dans un objectif d'informatisation, s'avérer contre-productif.

Notons aussi que l'annotateur ne reporte peut être pas l'humeur adéquate. Il peut y avoir dissociation entre l'humeur ressentie - impactée par le stimulus musical - et celle que l'auditeur

⁵³ Bien qu'encore une fois aucune conclusion ne soit permise, cela crée l'un des axes de navigation des moteurs de recherches qu'est la navigation de proche en proche. Cette proximité de construit permet en plus d'espérer pouvoir répondre aux briefings du type « musique ressemblant à X ou Y ».

⁵⁴ "Because of the time-varying nature of music, developing systems to automatically organize an entire song or clip using a single mood label, as in prior approaches [...], undoubtedly leads to imprecise classifications."

annote du fait qu'il peut savoir (par exemple) que son « avis » n'est pas généralisable (Johnson-Laird, Oatley, 2008, 104-105⁵⁵ ; Lundqvist et al., 2009⁵⁶). *A contrario*, le biais peut venir d'une difficulté à différencier l'émotion ressentie (au travers de la modification de son humeur) de l'émotion perçue et véhiculée par la musique (Lundqvist et al., 2009)⁵⁷ *a fortiori* quand celles-ci sont très éloignées (Trochidis, Delbé, et Bigand, 2011)⁵⁸.

De plus, malgré la consigne, les sujets ont utilisé indistinctement adjectifs (ex. triste) et substantifs (ex. tristesse). Parmi les limites de l'algorithme développé, il était impossible de les regrouper sans supervision. *De facto*, aucun adjectif n'est synonyme d'un nom et inversement. Une étape de normalisation des lexèmes semble donc s'imposer. Celle-ci unifierait les qualifiants sous l'une ou l'autre des typologies.

Toutefois on pourrait se demander si cette différence de forme ne serait pas justement due (sans doute de manière inconsciente) à des sujets qualifiants d'une part la musique et d'autre part leur propre humeur. En français, si l'on parle de soi, de son humeur, on peut utiliser une expression de la forme « je ressens de la tristesse » (mettant en jeu un substantif), alors que si l'on qualifie la musique, on utilisera plus communément une expression de la forme « la musique est triste » (mettant en jeu un adjectif). La normalisation des lexèmes suivant l'une ou l'autre des typologies ferait donc, suivant cette hypothèse, une confusion de ces deux concepts (et tendrait potentiellement à réduire le consensus obtenu en induisant du bruit).

Revenons enfin sur les résultats : les extraits B et C sont bien qualifiés par rapport à une émotion, une impression émotionnelle de « tristesse ». Néanmoins, si l'on se rapporte à A, qualifié par « étrange » et « mouvement », aucun des 2 n'est sémantiquement lié à une émotion particulière. Tout au plus, mouvement suscite une activation physiologique forte, étrange suscite une surprise. En dehors de cela, ces deux termes ne sauraient se rapprocher ni d'une émotion primaire, ni d'une émotion complexe aussi floue soit-elle. On est donc dans un non-sens manifeste.⁵⁹

Conclusion

Epistémologiquement parlant, l'approche SIC que nous proposons complète les travaux de sciences physiques et de psychologie. Elle s'intéresse à la construction de sens en contexte d'écoute et au processus de qualification complexe qui en résulte. L'intégration de ces résultats offre des perspectives fructueuses d'optimisation pour les bibliothèques musicales.

On découvre qu'à partir d'expressions de qualifiants en conditions de liberté (open-ended), on peut reconstruire (à moindre effort) des clusters offrant des qualifiants à consensus intersubjectif suffisant pour une qualification efficace de l'humeur de la musique. Dans le futur, il serait intéressant, pour rendre cela computable, d'essayer de rapprocher cela d'un espace représentant la musique sur un plan d'arousal-valence (Thayer, 1989) permettant ainsi à

⁵⁵ “A subtle distinction occurs between the emotions that listeners experience from listening to a piece of music and their judgments about what emotion the music expresses. Listeners can tell that a piece is intended to convey happiness, though they themselves are irritated by it.”

⁵⁶ “There are, however, a number of methodological problems with self-reports. One is that participants may answer according to what they feel is the expected, socially most acceptable response.”

⁵⁷ “Another problem is that listeners may find it difficult to discriminate felt emotions from the emotions perceived in the music.”

⁵⁸ “Many studies indicate the important distinction between one's perception of the emotion(s) expressed by music and the emotion(s) induced by music.”

⁵⁹ Cela nous amène à penser que la qualification de l'humeur (par abus de langage de l'émotion) de la musique n'est pas suffisante pour rendre compte de ce qui est qualifiable dans l'évocation. Une autre notion, pourquoi pas d'atmosphère pourrait aider à diminuer ce problème. Ainsi, on pourrait voir la distanciation B-C par rapport à A comme le fait que se ne sont pas les mêmes aspects qualifiants qui ont été évalués dans l'un et l'autre cas.

l'ordinateur de n'avoir à traiter que des nuages de points et des formes pour lier les morceaux entre eux et la requête utilisateur au morceau.

Références bibliographiques

Alpert, J.I. & Alpert, M.I. (1990), « Music influences on mood and purchase intentions », *Psychology and Marketing*. Vol.7(2), 109–133.

Barrett, L.F. (1996), « Hedonic tone, perceived arousal, and item desirability : Three components of self-reported mood », *Cognition & Emotion*, Vol.10(1), 47–68.

Barthet, M., Fazekas, G. & Sandler, M. (2012), « Multidisciplinary perspectives on music emotion recognition: Implications for content and context-based models », *Proc. CMMR'12*, 492–507.

Bateson, G. & Drosso, F. (1995), *Vers une écologie de l'esprit. Tome I*, Paris, France : Seuil.

Bateson, G., Lot, L. & Simion, E. (2008), *Vers une écologie de l'esprit. Tome II*, Paris, France : Seuil.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J. & Dacquet, A. (2005), « Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts », *Cognition & Emotion*, Vol.19(8), 1113-1139.

Corby, O., Dieng-Kuntz, R., Faron-Zucker, C., Gandon, F. & Giboin, A. (2005), « Le moteur de recherche sémantique Corese », *Actes Raisonner le web sémantique avec des graphes*, Nice, France.

Chai W. & Vercoe B (2000) « Using User Models in Music Information Retrieval Systems ». *Proc. ISMIR'00*.

Debaecker, J. (2012), « Reconnaissance des émotions dans la musique », *Les Cahiers du numérique*, Vol.7(2), 135–156.

Ertzscheid, O. (2008), « Moteurs de recherche : des enjeux d'aujourd'hui aux moteurs de demain », [en ligne], Paris, France.

Fabbri, P. (1968), « Considérations sur la proxémique », *Langages*, Vol.3(10), 65-75.

Gaume, B. (2004), « Balades aléatoires dans les petits mondes lexicaux », *Information interaction intelligence*, Vol.4(2), 31–90.

Gaume, B. (2008), « Mapping the forms of meaning in small worlds », *International Journal of Intelligent Systems*, Vol.23(7), 848–862.

Ghezaiel, L.B., Latiri, C., Ahmed, M.B. & Goudier-Khouja, N. (2011), « Un réseau proxémique pour la recherche d'information: Application à la maladie des dystonies ».

Jeanneret, Y. (2007), *Y a-t-il (vraiment) des technologies de l'information?*, Villeneuve d'Ascq, France : Presses Universitaires Septentrion.

Johnson-Laird, P.N. & Oatley, K. (2008), « Emotions, Music, and Littérature », *In* : Lewis, M., Haviland-Jones, J.M. & Barrett, L.F., *Handbook of emotions. 3rd Edition*, New York, USA : Guilford Press, 102-113.

Kim, Y.E., Schmidt, E.M., Migneco, R., Morton, B.G., Richardson, P., Scott, J., Speck, J.A. & Turnbull, D., « Music emotion recognition: A state of the art review. », *Proc. ISMIR'10*, 255–266.

Labour, M. (2011), *MEDIA-REPERES : Une méthode pour l'explicitation des construits de sens au visionnage*. Valenciennes : Univ. Lille Nord de France.

Lamprier, S., Amghar, T., Levrat, B. & Saubion, F. (2010), « Organiser les résultats d'une recherche d'information. Clustering, répartition de l'information et facilité d'accès », *Document numérique*, Vol.13(1), 9-39.

Lee, J.H. & Downie, J.S. (2004), « Survey Of Music Information Needs, Uses, And Seeking Behaviours: Preliminary Findings », *Proc. ISMIR'04*, 5

Leleu-Merviel, S. (1997), *La Conception en communication : méthodologie qualité*, Paris, France : Hermès.

Leleu-Merviel, S. (2003), « Le désarroi des maîtres du sens à l'ère du numérique », *Actes H²PTM'03*, 17-34.

Leleu-Merviel, S. (2003), « La structure du Aha. De la fulgurance comme percolation », *Actes H²PTM'05*, 59-76.

Leleu-Merviel, S. (2010), « Le sens aux interstices, émergences de reliances complexes », *Actes. Colloque Complexité'10*.

- Li, T. & Ogihara, M. (2004), « Content-based music similarity search and emotion detection », *Proc. ICASSP'04*, 705.
- Lundqvist, L.O., Carlsson, F., Hilmersson, P. & Juslin, P.N. (2009), « Emotional Responses to Music : Experience, Expression, and Physiology », *Psychology of Music*, Vol.37(1), 61-90.
- Meyer, L.B. (2008), *Emotion and Meaning in Music*, Chicago, USA : University of Chicago Press.
- Mugur-Schachter, M. (2006), *Sur le tissage des connaissances*, Paris, France : Hermès.
- Ploux, S. & Victorri, B. (1998), « Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes », *Traitement automatique des langues*, n° 39, 161–182.
- Rastier, F. (2001), « Signification, sens et référence du mot », *Journal of Linguistic*, Vol.4(1), 9-30.
- Schedl M., Flexer A. (2012) « Putting the User in the Center of Music Information Retrieval. » *Proc. ISMIR'12*, 385–390.
- Schmidt, E.M., Turnbull, D. & Kim, Y.E. (2010), « Feature selection for content-based, time-varying musical emotion regression », *Proc. International conference on Multimedia information retrieval*, 267–274.
- Shannon, C.E. (1948), « A mathematical theory of communication », *The Bell System Technical Journal*, Vol.27(3), 379-423.
- Shannon, C.E. (1964), *The mathematical theory of communication*. Urbana, USA : University of Illinois Press.
- Tat, K.Q. (2011), *Recherche d'information sur le web (RIW) et moteurs de recherche: le cas des lycéens*, Cachan, France : ENS Cachan.
- Thayer, R.E. (1989), *The Biopsychology of Mood and Arousal*, New York, USA : Oxford University Press, 250 p.
- Trochidis, K., Delbé, C. & Bigand, E. (2011), « Investigation of the relationships between audio features and induced emotions in Contemporary Western music », *Proc. 8th Sound and Music Computing Conference*.
- Wittezaele, J.-J. (2006), « L'écologie de l'esprit selon Gregory Bateson », *Multitudes*, Vol.24.
- Yvart F.-X. (2004), *L'émotion musicale : du rôle du contexte socio-émotionnel au partage social de l'émotion musicale*. Villeneuve d'Ascq, France : Univ. Lille 3.
- Yvart, W. (2013), *La musique au mètre : petit guide des musiques à l'usage des professionnels*, Villeneuve d'Ascq, France : Presses universitaires du Septentrion.
- Zillmann, D. (1971), « Excitation transfer in communication-mediated aggressive behavior », *Journal of Experimental Social Psychology*, Vol.7(4), 419-434.