

Reconnaissance des Émotions dans la Musique

Jean Debaecker

► **To cite this version:**

Jean Debaecker. Reconnaissance des Émotions dans la Musique. Analyse d'opinions sur internet., Hermès-Lavoisier, pp.135-155, 2011. sic_00682496

HAL Id: sic_00682496

https://archivesic.ccsd.cnrs.fr/sic_00682496

Submitted on 26 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LES CAHIERS DU NUMÉRIQUE

Consignes éditoriales à l'usage des auteurs

<http://lcn.revuesonline.com>

DEBAECKER JEAN

Résumé bilingue de l'article d'une dizaine de lignes maximum :

La reconnaissance des émotions dans la musique est un challenge industriel et académique. À l'heure de l'explosion des contenus multimédias, il devient nécessaire de concevoir des ensembles structurés de termes et concepts facilitant l'organisation et l'accès aux connaissances. Dans les bibliothèques musicales, l'analyse des sentiments et la classification par émotion sont très émergentes et demeurent un objet de recherche ambitieux. Dans ce chapitre, nous ferons un aperçu de l'existant et du contexte applicatif sur l'évaluation quantitative et qualitative des sentiments à travers le prisme de l'intersubjectivité du sujet. Puis, dans le cadre de notre thèse de doctorat au laboratoire GERiCO de l'université Charles-de-Gaulle Lille 3, nous présenterons le protocole expérimental mis en place ; permettant d'évaluer l'intersubjectivité du sujet ainsi que les premiers résultats obtenus.

Music emotion recognition is an industrial and academic challenge. While we observe an explosion of multimedia content, it becomes crucial to build structured sets of terms and concepts so as to help knowledge organization. In music libraries, sentiment analysis and emotion classification are emerging and remain a subject of ambitious research. In this chapter, we will do an overview of the state-of-the-art over the application context and the quantitative and qualitative assessment of feelings through the intersubjectivity of the subject. Then, as part of our PhD works at the laboratory GERiCO - university Charles de Gaulle Lille 3, we will present our experimental protocol for evaluating the intersubjectivity of the subject and the first results obtained so far.

LE RESUME ANGLAIS FIGURERA EN FIN DE NUMERO

Les auteurs envoient leur article au coordonnateur du numéro spécial avec :

- le fichier (.doc) attaché et éventuellement une copie des polices de caractères utilisées, dans le cas de symboles particuliers.
- l'annexe de fabrication
- le formulaire de cession de droits rempli et signé.

Le service éditorial Lavoisier (lcn@lavoisier.fr) reste à votre disposition pour toute information complémentaire.

Reconnaissance des Émotions dans la Musique

Introduction

La Reconnaissance des Émotions dans la Musique (R.E.M.) est un challenge important à plus d'un titre. La musique joue un rôle capital dans l'histoire de l'humanité, et ce d'autant plus avec l'arrivée du numérique, Web 2.0 et Web de données. Nous assistons à une véritable explosion des contenus multimédias. Il devient nécessaire de concevoir des ensembles structurés de termes et concepts facilitant l'organisation et l'accès aux connaissances. Dans les bibliothèques musicales, l'analyse des sentiments et la classification par émotion sont très émergentes et demeurent un objet de recherche ambitieux.

D'une part, la perception des émotions est fondamentalement subjective. Cette subjectivité varie¹ en fonction de l'individu, de sa culture, sa génération, du sexe, de sa personnalité, du contexte d'écoute et au sein même d'une même chanson (Yang *et al.*, 2008). D'autre part, nous ignorons exactement ce qui déclenche en nous des émotions : il s'agit d'un agencement complexe d'affects et de percepts difficilement définissables. Y a-t-il un élément dans la musique même, ou chez le sujet, qui permette l'éclosion de l'émotion ? Ces éléments demeurent encore relativement méconnus. Ensuite, la description des émotions elle-même n'est pas universelle. Les termes utilisés peuvent être très ambigus et, d'une personne à l'autre, plusieurs termes différents peuvent se référer à une même émotion.

Dans ce chapitre, nous ferons un aperçu de l'existant et du contexte applicatif sur l'évaluation quantitative et qualitative des sentiments à travers le prisme de l'intersubjectivité du sujet.

Puis, dans le cadre de notre thèse de doctorat au laboratoire GERüCO² de l'université Charles-de-Gaulle Lille 3, nous proposerons une présentation critique du protocole expérimental mis en place, permettant d'évaluer l'intersubjectivité du sujet ainsi que les premiers résultats obtenus.

¹ Ces variations s'intègrent dans ce que nous désignerons plus loin comme étant « les mondes possibles ».

² Groupe d'Etudes et de Recherche Interdisciplinaire en Information et COmmunication

Contexte

Enjeux applicatifs

Depuis plusieurs années, notamment depuis la création en 2000 de l'International Society for Music Information Retrieval (I.S.M.I.R.), les scientifiques et industriels cherchent à extraire des données suffisamment significatives d'un objet sonore tel qu'une chanson, un enregistrement de répondeur téléphonique, etc., afin de pouvoir les analyser et les utiliser. Par objet sonore, nous nous accorderons avec la définition de Bernard Sève (Sève, 2002) : « un ensemble organisé d'évènements sonores, et l'ensemble des effets produits, chez l'auditeur, par ces évènements. »

Au niveau applicatif, plusieurs scénarios très concrets sont envisageables, parmi eux :

- un répondeur téléphonique triant les messages enregistrés en fonction de leur humeur ;
- recherche de musique apaisante après une journée de travail ;
- recherche de musique excitante pour accompagner une activité sportive ;
- recherche d'opinions sur les réseaux sociaux.

Les métadonnées traditionnelles ne permettent pas de répondre à ce type de requêtes. (Lu *et al.*, 2006)

Une métadonnée est littéralement une donnée qui décrit une donnée. Autrement dit, une donnée qui facilite l'accès et la compréhension au contenu informationnel d'un objet. Les métadonnées d'un objet sonore numérisé répondent à une indexation documentaire à trois niveaux : « contenu – auctorialité – matérialisation ».

- le contenu concerne la description même de la chanson : titre, titre de l'album, genre musical, etc. ;
- la propriété intellectuelle, c'est-à-dire l'auteur, le copyright, le compositeur, l'interprète, etc. ;
- les caractéristiques « matérielles » : date, format, durée, fréquence, échantillonnage, etc.

Dans le cadre d'une bibliothèque multimédia, nous visons l'enrichissement des dispositifs de recherche (moteurs de recherche) fonctionnant par similarité émotionnelle. Autrement dit, l'amélioration des classifications par émotion ou humeur (en anglais : *moods*).

D'une part, nous soulevons le problème de la production et de la représentation de ce type d'information (les émotions) ; d'autre part nous questionnons son organisation et son écriture (les métadonnées). Dans quelles mesures est-il possible d'inclure dans les métadonnées d'un objet sonore une donnée portant sur l'émotion ? Ce type d'indexation pose le problème de la modélisation et de la normalisation d'une telle information.

Nous observons un glissement paradigmatique de l'indexation classique à l'indexation sociale, collaborative, certes moins institutionnelle mais plus pertinente pour certains usages. Le nombre d'échanges en ligne sur le sujet témoigne de son importance et de son actualité. En effet, les nouvelles pratiques émergentes au sein des réseaux sociaux semblent démontrer l'intérêt de notre approche : MusicCovey, Aupeo, StereoMood, Last.Fm, The Sixty One, etc.

MuMa : the Music Mashup

C'est un des objets de recherche du programme franco-allemand *Quaero* : « programme fédérateur de recherche et d'innovation industrielle sur les technologies d'analyse automatique, de classification et d'utilisation de documents multimédias et multilingues. »³ Ils ont notamment développé un moteur⁴ : *MuMa : the Music Mashup* incluant des fonctions de recherche par humeur (*moods*), genre, accord, etc. : <http://muma.labs.exalead.com>. Le fonctionnement global du dispositif est expliqué par Rémi Landais.⁵

Classification et technique multi-label

Basée sur les travaux de Juan José Burred et Geoffroy Peters (Burred *et al.*, 2009), la classification par humeur a été effectuée manuellement sur une base de 193 fichiers. Chaque fichier ayant été évalué plusieurs fois par trois annotateurs. Ils n'ont conservé que les labels (*moods*) ayant été utilisés par deux

³<http://www.quaero.fr/modules/movie/scenes/home/index.php?fuseAction=article&rubric=presentation> - visité le 20.05.2011.

⁴ En collaboration avec Exalead, IRCAM, Télécom ParisTech et Yacast.

⁵ Sur le site *Quaero*, sous l'onglet « Vidéos et démos ». Visité le 20.05.2011.

annotateurs et s'appliquant à au moins six chansons. Les auteurs précisent eux-mêmes :

« (...) in benefit of general applicability, it must avoid a too high modeling accuracy on the training set, i.e., it must avoid overfitting. The most demanding goal of a classification application is not accuracy, but generality. »

Computer Audition Lab 500

Le travail est effectué grâce à l'ensemble des données de CAL-500 et le circomplexe Valence / Arousal (voir Figure 1). Le circomplexe (Guttman, 1954) est un système de représentation circulaire intégrant des variables psychométriques rigoureusement corrélées. Le circomplexe de (Thayer, 1989) se base sur l'excitation chimique et fait figure d'autorité pour les systèmes de représentation des émotions.



Figure 1. Le circomplexe arousal-valence de Thayer

L'ensemble de données *Computer Audition Lab 500*⁶ regroupe 500 chansons de 500 artistes. Chaque chanson a été annotée par 3 personnes (Turnbull *et al.*,2008); ce qui a généré 1700 annotations musicales décrivant 500 chansons occidentales populaires. Six attributs sont utilisés : *Mood, Genre, Instrument, Song, Usage* et *Vocal*(Zhao *et al.*,2010). Ces tags furent générés sous des conditions de contrôle expérimental et sont considérés par la communauté comme faisant figure d'autorité. Ils ont ainsi pu concevoir un système capable de générer un descriptif pour une chanson au hasard à partir de leur base de données. Les mots en gras sont générés par leur système :

Nirvana-Aneurysm

This is **alternativey, rock** song that is **exciting and not lighthearted**. It features **distorted electric guitar, clean electric guitar, drum set**, a nice **distorted electric guitar solo**, and **screaming, strong vocals**. It is a song with a **fast tempo** and **with negative feelings** that you might like listen to while **driving**.

Problématique de recherche

Notre problématique est la suivante : est-il possible d'avoir une connaissance *a priori* de l'émotion en vue de son élicitation ? Nous entendons le terme *a priori* comme signifiant : universel et indépendant de l'expérience. Autrement dit, dans la pratique, est-il possible de connaître l'émotion que l'on ressentira à l'écoute d'une chanson, avant de l'écouter ? Suite à l'écoute, son élicitation est-elle possible ? C'est-à-dire, est-ce possible de formaliser une émotion dans le but de la sauvegarder et, ou, de la partager ?

Les mondes possibles de l'émotion

Dans un premier temps, il nous importe de préciser les notions suivantes : intersubjectivité du sujet, monde possible, intensité, sensation, sentiment et émotion.

L'intersubjectivité du sujet

⁶ CAL-500 est le travail libre de droit de Douglas Turnbull, Luke Barrington, David TorresetGertLanckriet.

Selon Kant, ce que vise le jugement de goût, la faculté de juger, c'est l'universalité. C'est ce qu'il nomme « jugement ratiocinant » (Kant, 1790, pp. 326). Autrement dit, c'est l'expérience toute singulière et subjective du goût que nous vivons qui doit endosser un caractère suffisamment universel pour être partagé. Il préconise l'usage de la formule *Du goût, on peut discuter*, arguant « qu'il faut qu'on ait l'espoir de parvenir à un accord ». Accord ne se reposant pas uniquement sur des fondements subjectifs (dont la validité est toute personnelle), mais bien sur un autre élément.

Cet autre élément est à penser avec (Buber, 1935), il s'agit du sujet relié aux autres sujets « à l'intérieur de ses rapports à l'objet ». En prenant en considération la pensée d'autrui dans son propre jugement, c'est « intégrer une autre sphère d'appartenance » (Husserl, 1929). C'est-à-dire, intégrer la perception d'un monde qui n'est pas le notre, mais celui d'autrui. Un monde que nous ne percevons pas directement, mais dont nous connaissons indirectement l'existence.

Ainsi l'intersubjectivité se rapporte à l'intégration au sein de son jugement personnel, du jugement d'autrui et *espérer parvenir à un accord*.

Les mondes possibles

Les technologies de recherche des émotions dans la musique représentent un challenge important, dans la mesure où l'expérience émotionnelle est très personnelle. Elle peut varier d'une chanson c à l'autre, d'un individu i à l'autre, d'un instant t à l'autre : elle est dépendante du « monde possible » \mathcal{W} dans lequel elle s'inscrit. Nos travaux ne relevant pas de la logique modale et de la sémantique des mondes possibles,⁷ nous n'en ferons pas un objet d'étude ici. Il nous importe cependant de définir et illustrer nos propos :

Par « monde possible » \mathcal{W} , nous entendons toute situation contre-factuelle modifiant les relations et états de nécessité, contingence et possibilité des objets ou individus. Autrement dit, une telle situation présente un événement tel qu'il aurait pu se présenter sous certaines conditions.

En guise d'illustration, citons cet exemple célèbre suivant : la proposition « Jules César franchit le Rubicon le 11 janvier de l'an 49 avant J.C. » est vraie dans le monde actuel \mathcal{W} . Dans un autre monde possible \mathcal{W}_i , la proposition « Jules César ne franchit pas le Rubicon le 11 janvier de l'an 49 avant J.C. » se vérifie également.

⁷Kripke S.A., *Naming and necessity*. Wiley-Blackwell, 1981.

Dans un cadre musical, il existe un monde possible W_1 (voir Figure 2) dans lequel l'individu Paul ressent de la mélancolie en écoutant le morceau *I go to sleep* (repris par le groupe Anika). Dans W_2 , le même individu Paul peut ressentir de la joie.

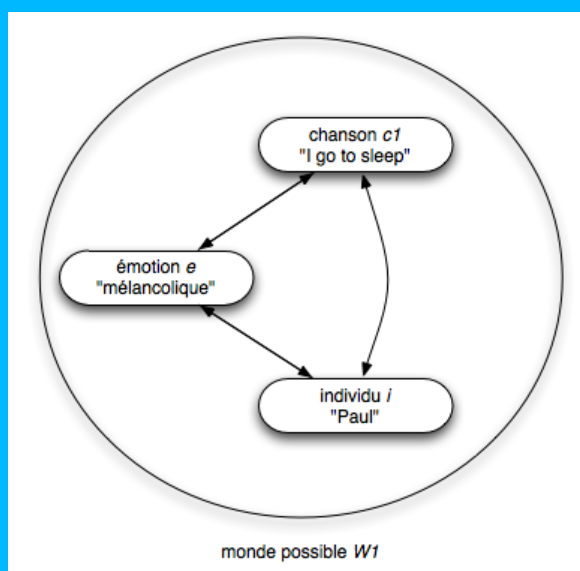


Figure 2. Le monde possible de Paul

De l'intensité

Intensité et quantité

L'artiste sait qu'une œuvre de Schubert lui procure un plaisir plus intense qu'un *jingle* publicitaire. Or, comment passer d'une intensité à une autre ? Si nous appliquons la même (dis)continuité entre les objets, il est, par exemple, incontestable qu'un nombre en surpasse un autre, cette disposition croissante est justifiée dans la série naturelle des nombres : 13 est plus grand que 10, lui-même plus grand que 7, etc.

Dans quelles mesures il est possible d'appliquer une telle série avec des intensités ?

Soit (voir Figure 3) :

- S : une sensation
- S' : une autre sensation
- ΔS : l'écart entre S et S'

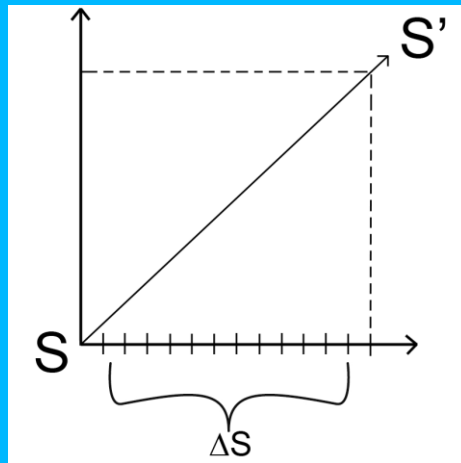


Figure 3. ΔS comme échelle quantitative ou qualitative

Si nous admettons, avec Fechner (Fechner, 1860), que S' soit plus fort que S et qu'il demeure une inégalité entre ces sensations ; alors il nous semble naturel d'établir un rapport quantitatif (ΔS) entre S et S', autrement dit de chercher de combien S est surpassé par S' (voir équation 1).

$$\{ \Delta S = [S_1 + S_2 + \dots + S_n] \} \quad (1)$$

ΔS représente une série d'intervalles juxtaposés dans l'espace. Cette traduction spatiale implique une différence mathématique et donc des variations quantitatives. Le moins demeure au sein du plus, ces variations, ces quantités peuvent croître ou diminuer.

Par exemple, dans le cadre d'une intensité de chaleur, ΔS s'exprime en degré. 25°C (S') est une intensité de chaleur plus grande que 13°C (S). Par ailleurs, l'usage d'un appareil d'enregistrement graphique montre(ra)it que cette intensité de chaleur est solidairement liée à une activité corporelle, renforçant ainsi l'idée d'un intervalle quantifiable. Les états physiques sont des états

solidaires de leur cause extérieure. Autrement dit, les sensations sont liées à leur cause extérieure. Ici, ces quantités d'intensité sont extensives et mesurables, ces efforts sont des tensions du corps. Ce sont des sensations représentatives.

Dans le cadre de l'émotion, cela revient à établir un intervalle entre, par exemple, une joie S et une joie S' plus grande. La question ici est de savoir s'il nous est possible d'atteindre une intensité de joie supérieure à la condition d'être d'abord passé par les intensités inférieures intermédiaires de la même sensation de joie. Car effectivement l'artiste ressent une plus grande intensité de joie à l'écoute de telle œuvre musicale qu'un *jingle* publicitaire. Pourquoi disons-nous d'une intensité supérieure de joie S' qu'elle est plus grande que S ?

Si pour obtenir un tas de dix pierres, il faut effectivement passer par l'addition de dix éléments rocheux ; est-ce que pour connaître une intensité supérieure de joie S' , il faut passer par l'addition d'intensités intermédiaires de joie S_n ?

Intensité et qualité

Selon Bergson (Bergson, 1889) l'erreur de Fechner réside dans la définition et représentation de ΔS . Là où ce dernier y voit la représentation d'une quantité, le premier y décèle la représentation d'une qualité. Dans le cas des intensités de chaleur, certes rien n'empêche de mesurer la sensation de chaleur par le degré de température, mais selon Bergson il s'agit là de convention. Une chaleur plus intense est réellement une chaleur autre. Nous la pensons plus intense parce que nous avons déjà éprouvé ce même changement quand nous nous rapprochons d'une source de chaleur.

L'intérêt de comparer perception de chaleur et perception de la musique réside en deux propriétés de la perception :

D'une part endossant un caractère prétendu universel, le premier lieu commun *chacun possède son propre goût* [et ce] *jugement n'a nul droit à l'assentiment nécessaire d'autrui* (Kant, 1790) amène le plus grand nombre à s'accorder commodément de dire que *Yesterday* des Beatles est plus triste que *Ob-La-Di, Ob-La-Da* ; ou qu'un frottement de mains très énergique est plus chaud qu'un frottement de mains mou.

D'autre part, la génération d'un trouble des régulations organiques est une propriété universelle partagée entre l'émotion (voir la partie définitionnelle de l'émotion ci-dessous) et la chaleur.

Sensation et sentiment

Afin d'éviter toute confusion possible, nous entendrons la sensation comme étant une *représentation objective des sens* (Kant, 1790, p.183). Et le sentiment comme *ce qui doit toujours rester purement subjectif et ne peut constituer aucune espèce de représentation d'un objet.* (Kant, 1790, p.184)

Par exemple : la couleur bleue du ciel relève de la sensation : c'est la perception objective d'un *objet du sens de la vue*. Ce qu'il y a de gracieux et de plaisant dans cette perception relève du sentiment : c'est la perception subjective *par laquelle aucun objet n'est représenté*. Ceci ne constitue pas une connaissance.

D'un côté la sensation ne promet rien, indépendamment de ce qu'elle produirait de (dés)agréable et renverrait ainsi au sujet en tant qu'il a un corps.

D'autre part, le sentiment excite chez le sujet le désir de relations semblables : retrouver la grâce et le plaisir à travers un autre objet, par exemple *l'International Klein Blue* ; dont Klein dira lui-même : « Je suis allé signer mon nom au dos du ciel dans un fantastique voyage... »

L'émotion

Approche définitionnelle

Nous entendons l'émotion⁸ comme extensive des sentiments générant un trouble de l'adaptation des conduites de l'individu. Du point de vue des sciences cognitives, elle est une manifestation psychologique et physique (Thayer, 1989) d'un processus cognitif répondant à des événements de la vie. Ce processus étant lui-même un stimulus suffisant pour déclencher une réaction biologique. « [Emotion] is nothing more than a response caused by cognitive, physiological, and/or biochemical events » (ibid).

⁸Définition de l'encyclopédie Universalis : « L'émotion apparaît quand les exigences de la situation sont disproportionnées avec les possibilités du sujet, c'est-à-dire lorsqu'il y a un décalage entre ses anticipations perceptives et cognitives et son répertoire de réponses disponibles. Tout se passe alors comme si une énergie inemployée se diffusait dans tout l'organisme, troublant aussi bien les régulations organiques et l'idéation que la vie de relation. »

Connaissance a priori de l'émotion

Dire que l'émotion est objective reviendrait à définir un ensemble $[P]$ de propriétés essentielles de la joie, par exemple :

- P_1 : tempo rapide ;
- P_2 : accords majeurs ;
- P_3 : paroles « positives ».

Une chanson C conforme aux propriétés $[P_1 ; P_2 ; P_3]$ serait alors joyeuse. Or l'émotion n'est pas une connaissance. Elle est fondamentalement subjective et insécable.

Cadre théorique

Notre hypothèse se base sur le travail d'Emmanuel Kant (Kant, 1790). Nous soutenons, en effet, que le *libre jeu de l'imagination et de l'entendement* ne peut pas être intellectuellement connu, mais seulement, senti. Aucune connaissance de l'émotion *ne précède l'expérience, et toutes commencent avec elle* (Kant, 1787). Autrement dit, il nous est impossible d'établir une connaissance *a priori* de l'émotion :

« Pour décider si une chose est belle ou ne l'est pas, nous n'en rapportons pas la représentation à son objet au moyen de l'entendement et en vue d'une connaissance, mais au sujet et au sentiment du plaisir ou de la peine, au moyen de l'imagination (peut-être jointe à l'entendement). Le jugement de goût n'est donc pas un jugement de connaissance ; il n'est point par conséquent logique mais esthétique, c'est-à-dire que le principe qui le détermine est purement subjectif. » (Kant, 1790, p.181)

La rencontre entre un objet esthétique (une chanson) et un individu relève de la représentation et de la sensation. Cette relation ne désigne ni l'individu, ni la chanson, mais l'état de conscience du sujet affecté par la rencontre :

« Mais il ne s'agit plus alors de leur relation au sentiment du plaisir et de la peine, laquelle ne désigne rien de l'objet, mais simplement l'état dans lequel se trouve le sujet affecté par la représentation. » (*ibid.*)

Ce cadre théorique diverge de celui d'Emery Schubert⁹, que nous ne partagerons pas de fait :

⁹AssociateProfessor, School of English, Media and Performing Arts.

« Based on the frequently supported assertion that there is a significant degree of agreement regarding the emotion expressed by a single piece of music, it stands to reason that a non-trivial proportion of emotion in music can be attributed to musical structure, or more specifically, to combinations of musical features. » (Schubert, 2004, p.562)

Il nous incombe de confronter ce cadre théorique à la réalité du terrain via un protocole expérimental permettant d'évaluer la subjectivité inhérente au sujet et à un objet sonore

Présentation et méthodologie du protocole : Prêter l'oreille à la science !

Dans le cadre de notre travail de thèse de doctorat, au sein du laboratoire GERüCO de l'université Charles-de-Gaulle Lille 3, nous avons mis en place un dispositif permettant l'évaluation et l'échantillonnage des émotions en fonction d'objets sonores. Le projet s'intitule *Prêter l'oreille à la science !*¹⁰, débuté le 19 mai 2011 et, est accessible en ligne à l'adresse suivante : www.earstoscience.c.la. La durée prévue pour l'ensemble de l'expérimentation est de huit mois (jusque janvier 2012). Une petite base de données musicales a été construite. Ensuite une expérimentation en ligne a été conduite.

Sélection musicale et population

Contrairement au travail de Peeters et Burred (Burred *et al.*, 2009) qui ont sélectionné 193 chansons évaluées par 3 annotateurs ; nous avons décidé de ne sélectionner qu'une seule chanson potentiellement évaluable par le plus grand nombre possible. Notre choix est motivé par des raisons techniques, pratiques et financières, parmi celles-ci :

- le coût ;
 - nous manquons de financement pour recruter des annotateurs à temps plein.
- Le temps ;
 - la sélection de plusieurs centaines de chansons est très chronophage, malgré l'accès au CAL-500 Data Set. Qui plus est, il est plus simple de convaincre plusieurs centaines d'auditeurs d'annoter une chanson ; qu'un auditeur d'annoter plusieurs centaines de chansons.
- Le contraste ;

¹⁰ Lendyourears to science ! Le site est en français et anglais.

- contraster avec le travail effectué dans *Quaero*¹¹ et CAL-500 permet un enrichissement ; et de fait de mieux harmoniser et converger les travaux de recherche sur cette thématique.

La mise en place d'un formulaire en ligne permet de palier à ces aspects technico-financiers.

Sélection musicale

Cadre théorique de sélection

La majorité des expérimentations similaires, telle que celle d'Andy M. Sarroff (Sarroff, 2009), requiert la sélection d'un morceau remplissant des conditions telles que :

- chanson familière ou pas ;
 - afin d'éviter un biais de reconnaissance.
- Représentative d'un corpus populaire ;
 - structure couplet-refrain-couplet.
- Qualité sonore ;
 - qualité CD.

Notre expérimentation sort de ce cadre conventionnel pour les raisons suivantes :

- chanson familière ou pas ;
 - nous soutenons que l'expérience esthétique est transcendante et singulière (Nicolas, 2004) ; l'écoute se renouvelle dans la répétition qui se veut altérante et générant de la différence (Deleuze, 1968).
- Représentative d'un corpus populaire ;
 - le choix d'un tel corpus n'est pas plus pertinent qu'un autre. La structure répétitive (couplet-refrain) altère le champ de perception. Nous pensons qu'aucune contrainte n'est nécessaire à ce niveau.
- Qualité sonore ;
 - l'expérience doit être valable quelque soit la qualité audio du morceau ainsi que du dispositif d'écoute : écouteurs, enceintes d'ordinateur portable, chaîne Hi-Fi, etc.).

¹¹ Leur base de donnée n'est pas encore publique.

Échantillonnage

La chanson est sélectionnée aléatoirement au sein d'une bibliothèque musicale personnelle de 25032 morceaux : *I go to sleep* reprise par le groupe *Anika*.¹² Elle fut originellement écrite et composée par Ray Davies en 1965. Le morceau est au format AAC, d'un débit de 256 kbit/s, avec une fréquence d'échantillonnage de 44,100 kHz.

Nous avons également adjoint un échantillon à écouter d'une seule note : un do dièse. Le nom de la note n'est pas révélé sur le site afin de ne pas influencer *a priori* l'écoute. Le morceau (voir Figure 4) est au format AIFF, d'un débit de 1411 kbit/s, avec une fréquence d'échantillonnage de 44,100 kHz. C'est une note continue d'une minute enregistrée au clavier : nous avons utilisé l'application miniSynth Pro ©, synthétiseur conçu par Yonac pour l'iPad © d'Apple. Ces échantillons répondent aux contraintes méthodologiques imposées.

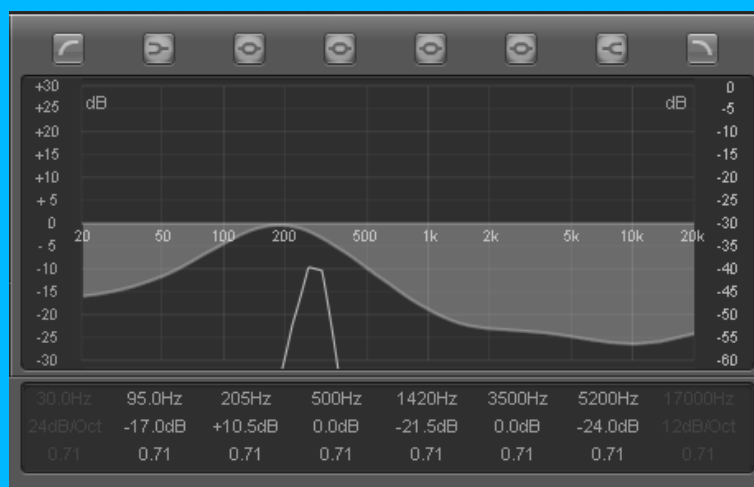


Figure 4. Fréquence exacte du Do dièse

Population

Il n'y a pas eu de sélection *de facto* : tout le monde pouvant potentiellement répondre. L'expérimentation a eu lieu en ligne. Nous l'avons annoncée en

¹²Anika *I go to sleep* - 2010 - Stones Throw Records -- Self-titled debut, Anika, produit par Geoff Barrow de Portishead.

postant un court texte introductif sur des forums (musicaux et autres), et nous l'avons envoyé sur des listes d'emails (chercheurs, collègues, amis, etc.). Le court texte introductif résume la nature de l'expérimentation et invite les personnes intéressées à se rendre sur le site web conçu à cet effet. Notre objectif était ainsi de réunir un maximum d'annotateurs.

Le site web¹³ comporte une page introductive non-technique contextualisant succinctement l'expérimentation et donnant les consignes, mais sans révéler l'objectif. Une autre page contient les deux morceaux à écouter (déposés sur SoundCloud¹⁴ ©), ainsi qu'un très court formulaire à remplir, réalisé avec Google Document.¹⁵ © Les expérimentations conventionnelles du domaine durent environ 45 minutes, le sujet parvenant difficilement à bloquer ce temps, et, ou, abandonnant en cours ; il nous importait donc particulièrement d'avoir une très courte durée d'expérimentation afin de ne pas décourager l'annotateur : cinq minutes.

Subjectivité et consigne d'écoute

Nous ne préconisons aucune consigne d'écoute, contrairement à (Yang *et al.*, 2008) : « Note that the subjects are asked to label the emotion based on their feelings of what the music sample is trying to evoke, rather than the emotion the subjects perceive at the test. We must make this distinction clear because perceived emotion and evoking emotion are not always the same. For example, a person who enjoys sorrowful tone might feel pleased when listening to sorrowful songs. »

En effet, si le sujet Paul considère *I go to sleep* comme étant mélancolique, mais ressent du plaisir à écouter des chansons mélancoliques ; alors que ressent-il exactement ? De la mélancolie ou du plaisir ? Nous pensons que cette distinction ne doit pas intervenir dans notre protocole d'évaluation. Effectivement, nous souhaitons que l'auditeur annote la chanson avec l'émotion ressentie lors de l'écoute, sans analyse aucune de sa part. Un biais linguistique est inévitable. Un auditeur anglais réagira plus facilement aux paroles d'un morceau anglais, contrairement à un auditeur non anglophone. Et en pratique ce fut le cas, notre 197^{ième} annotateur précise : « It has an ambivalent emotional tone. The lyrics are romantic, but the instrumental accompaniment is not. The instrumental accompaniment is pretty mechanical generally. Piano part is sometimes a bit cheerful, sometimes a bit mournful. »

¹³ <http://www.earstoscience.c.la>

¹⁴ <http://www.soundcloud.com>

¹⁵ <http://www.docs.google.com>

Notre choix (également motivé par la domination de la langue anglaise dans la musique) est de ne pas prendre cet élément en compte : nous ne voulions pas nous focaliser sur un point particulier¹⁶ de la chanson, mais l'entendre dans sa globalité.

Nous ne donnons aucune limite d'écoute. Les annotateurs sont autorisés à :

- écouter le morceau plus d'une fois ;
- effectuer plusieurs évaluations (dans le temps).

Nous pensons que retirer ces limites permettra de refléter au mieux les émotions de l'auditeur.

Enfin, la perception des émotions étant très subjective nous avons autorisé les annotateurs à commenter, motiver leur(s) choix et à les préciser ; voire même à effectuer des préconisations. Ce qui fut le cas dès le premier auditeur. Il a noté l'absence de la catégorie « Désagréable / Malsain / Dérangeant / Stressant / Angoissant ». La pertinence de cette classe nous a semblé évidente, et de fait nous l'avons intégré (sans changer les termes de l'auditeur) avec succès comme nous le verrons plus loin dans l'analyse des premiers résultats. La qualité et la précision des résultats occupent une place centrale dans notre analyse.

Métadonnées

Après examen des métadonnées retenues par Burred et Peeters, il nous importait de faire face à la polysémie et confusion possible des adjectifs employés. Pour ça nous avons construit des grandes classes regroupant plusieurs termes (français, traduits en anglais) :

- colère / agressif / angry / aggressive;
- bizarre / étrange / weird ;
- calme / apaisant / doux / décontacté / quiet / gentle / cool ;
- gai / festif / joyeux / positif / joyful / festive / happy / positive ;
- enchantant / magique / enchanting / magical ;
- grandiloquent ;
- mécanique / robotique / mechanical / robotic ;
- puissant / fort / powerful / strong ;
- romantique / passionné / romantic / passionate ;

¹⁶ Mais ceci représente autant d'autres perspectives de recherche passionnantes.

- triste / mélancolique / douloureux / sad / melancholic / doleful
- sophistiqué / élégant / sophisticated / elegant
- désagréable / malsain / dérangent / stressant / angoissant / unpleasant / disturbing / stressful / scary

Premiers résultats

La durée prévue de l'expérimentation est de 8 mois : de mai 2011 à janvier 2012. Nous aurions pu opter pour une approche dite synchronique. En effet, l'intersubjectivité du sujet s'illustre aisément : deux annotateurs différents évaluant les objets sonores au même instant t suffisent. Théoriquement, un même annotateur suffit à travers les mondes possibles (par exemple : t_1 ; t_2). Il nous a paru cependant plus pertinent d'adopter une approche dite diachronique avec un maximum d'annotateurs afin de conforter nos propos et renforcer l'intersubjectivité du sujet. Nous gagnerons en précision avec un nombre croissant d'annotateurs sur une longue durée.

À la date du 25 mai 2011, 265 annotateurs ont participé. Ce taux de participation exceptionnel est beaucoup plus élevé que prévu, et nous incite à analyser et publier ces premiers résultats, ainsi que le protocole.

I go to sleep

Les résultats sont dans l'ensemble hétérogènes (voir Figure 5). Les trois annotations les plus sélectionnées sont :

- triste / mélancolique (...): 39% ;
- mécanique / robotique (...): 27% ;
- étrange / bizarre (...): 25%

Les utilisateurs peuvent cocher plusieurs cases, donc les pourcentages peuvent être supérieurs à 100%. Nous notons qu'en quatrième position se retrouve actuellement la classe « calme / apaisant (...) » qui était alors en première position au bout de 145 annotateurs. Nous nous emploierons à surveiller les résultats futurs pour noter toute évolution importante du même niveau. Ces résultats (et leurs évolutions inhérentes) tendent à montrer l'instabilité des sentiments, ou plutôt à montrer la stabilité de l'intersubjectivité du sujet : un consensus clair et unanime ne ressort pas nettement.

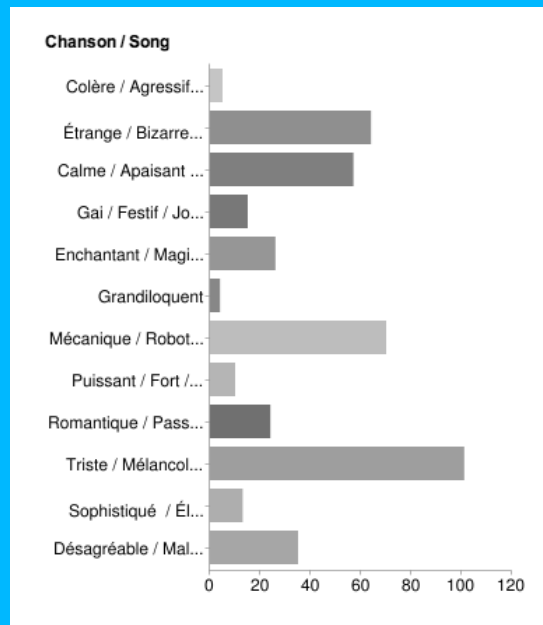


Figure 5. Répartition du taux de participation pour la chanson

Do dièse

Les résultats sont dans l'ensemble plus homogènes (voir Figure 6). Les trois annotations les plus sélectionnées sont :

- désagréable / malsain (...) : 44% ;
- mécanique /robotique (...) : 25% ;
- étrange / bizarre (...) : 15%

Les utilisateurs peuvent cocher plusieurs cases, donc les pourcentages peuvent être supérieurs à 100%. Nous notons le succès manifeste de la recommandation du premier auditeur sur la classe « désagréable ».

Ces premiers résultats sont difficiles à analyser actuellement. En effet, nous ignorons si cette sensation unanime est due au son même : la qualité du son du clavier, il en aurait été probablement été autrement avec une simple note de piano¹⁷ ; ou si elle s'en dissocie. Le do# est connu pour être une note dite « triste », or cette catégorie ne recueille que 13% des choix (lui conférant tout de

¹⁷ Point souligné notamment par le 167^e annotateur.

même la quatrième position). Cependant, nous notons que l'unanimité des choix est plus simple sur une seule note que sur une chanson.

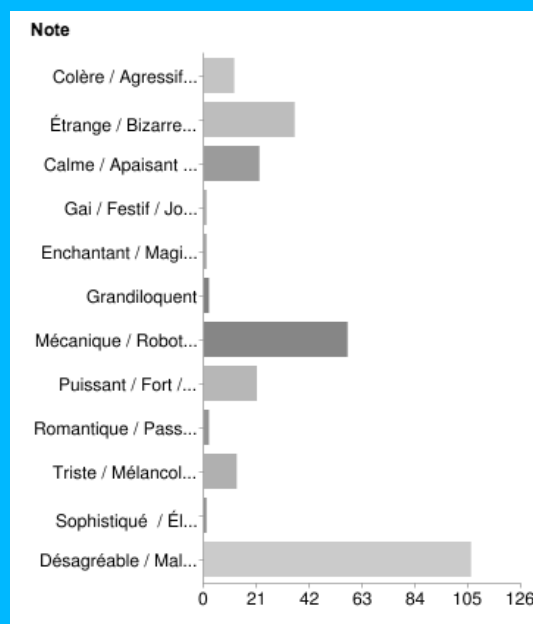


Figure 6. Répartition du taux de participation pour la note.

Cela met en relief plusieurs éléments : d'une part nous pourrions penser *a priori* qu'une note continue stabiliserait et homogénéiserait les sentiments : en effet, les différentes parties constituantes d'une œuvre musicale peuvent susciter différentes émotions. Or, le 77^{ième} annotateur précise : « Calme au début car son grave mais sa longue tenue la rend par la suite angoissante ». Ceci montre que la répétition génère de la différence, et transcende l'écoute de l'auditeur. C'est là tout le travail des musiques répétitives (Steve Reich, Gavin Bryars, Éliane Radigue The Field, etc.).

Évaluation critique et perspectives

Le protocole doit répondre à trois propriétés minimales : transmissibilité, reproductibilité et effectivité. Les deux premiers traits constituent l'objet de ce chapitre. Quant au troisième point : est-ce que le protocole répond à la problématique ? L'expérience isolée *Prêter l'oreille à la science!* nécessite une évolution vers une expérimentation globale. En effet, nous sommes ici dans un

bloc solide et chronologique : composition – expérimentation – évaluation. Bloc insécable sans interaction dynamique entre les éléments. Pour l'expérimentation globale, nous préconisons un triptyque identique, différencié en ceci :

– interaction dynamique :

- les trois étapes évolueront de concours, par exemple : des résultats partiels significatifs questionneront la composition de l'échantillonnage.

– Une base musicale plus importante.

– La présence d'un groupe témoin :

- un expérimentation menée en parallèle servira de jalon afin de gagner en précision statistique.

L'expérimentation suivante se concentrera principalement sur une partie similaire au *Do dièse*.

À travers ce chapitre nous avons essayé de démontrer l'impossibilité d'un consensus unanime eu égard à l'élicitation *a priori* des émotions dans la musique. Cet apport théorique basé sur le travail de Kant et Bergson nous semble indispensable à l'heure actuelle, dans la mesure où émerge : (1) une forte activité industrielle intense sur la classification des émotions (notamment à travers les réseaux sociaux et le web sémantique) ; (2) une recherche universitaire et académique de plus en plus intense.

En effet, comme nous l'avons vu, les travaux récents sur le sujet ne mentionnent aucunement ce bagage kantien et bergsonien, qui pourtant nous apparaît indispensable. À cet égard, et via notre protocole d'expérimentation, nous maintenons fermement notre hypothèse de départ : il est impossible d'avoir une connaissance *a priori* de l'émotion en vue de son élicitation. Nous soutenons que l'élicitation n'est pas une propriété universelle et absolue. Autrement dit, il est impossible de formaliser une émotion dans le but de la sauvegarder et, ou, de la partager. Notre cadre théorique nous semble pertinent pour démontrer que l'élicitation des émotions n'est pas une propriété universelle et absolue. Non-propriété qui tend résolument à se vérifier à travers les premiers résultats de notre protocole expérimental d'écoute *Prêter l'oreille à la science*. Ces éléments s'affineront dans de futurs travaux de recherche que nous défendrons lors de notre thèse de doctorat.

Remerciements

Nous remercions pleinement les auditeurs anonymes pour leur précieuse contribution. Nous remercions également nos proches et collègues pour avoir rendu cette expérimentation possible.

Bibliographie

- Bergson H., *Essai sur les données immédiates de la conscience*, Paris, 1989.Édition critique par Bouaniche A., sous la direction de Worms F.,Presses Universitaires de France, 2007.
- Buber M., *Ich und Du.*, 1935.Trad. Fr. Aubier-Montagne, Paris, 1992.
- Burred J.J., Peeters G., « An adaptive system for music classification and tagging », *Welcome to the 3rd International Workshop on Learning Semantics of Audio Signals*, page 3, 2009.
- Deleuze G., *Différences et répétitions*, Presses Universitaires de France, Paris,1968.Réédition 2003.
- Fechner G., *Elemente der Psychophysik*, 1860. Trad. Adler H.E., Howes D.H., Boring E.G.,*Elements of psychophysics*.Holt, Rinehart and Winston New York, 1966.
- Guttman L., *A new approach to factor analysis : the Radex*. In P.F. Lazarfeld (Ed), *Mathematical thinking in the social sciences* (pp. 258-348). Glencoe : The Free Press, 1954.
- Husserl E., *Les méditations cartésiennes : introduction à la phénoménologie*, Paris. 1929. Nouvelle édition, Vrin, 2000.
- Kant E., *Kritik der Urteilskraft*, Riga, 1790 ;trad. fr. deRenaut A.,*Critique de la faculté de juger*,GF Flammarion, Paris, 1995.
- Kant E.,*Kritik der reinen Vernunft*, Riga,1781-1787 ;trad. fr. de Renaut A.,*Critique de la raison pure*, GF Flammarion, Paris, 2006.
- Lu Li, Liu D., ZhangH.J., « Automatic mood detection and tracking of music audio signals », *Audio, Speech, and Language Processing, IEEE Transactions*,5–18, 2006.
- Nicolas F., *Écouter, lire et dire la musique : théorie de l'écoute musicale*. 2004
- SaroffA.M.,*Subjective evaluation of spatial impression in reproduced stereophonic music*. 2009.
- Sève V.,*L'altération musicale*, Paris, Éditions du Seuil. 2002.
- Schubert. E.,« Modeling perceived emotion with continuous musical features », *Music Perception*, 2004.
- Thayer R.E., *The Biopsychology of mood and arousal*, Oxford, U.K. : Oxford Univ. Press, 1989.

- Turnbull D., Barrington L., Torres D., Lanckriet G., « Semantic annotation and retrieval of music and sound effects », *IEEE Transactions on Audio, Speech and Language Processing*, p467–476, February 2008.
- Yang Y-H., Lin Y-C., Su Y-F., Chen H.H., « A regression approach to music emotion recognition », *IEEE Transaction on audio, speech, and language processing*. Vol. 16. N °2, 2008.
- Zhao Z., Wang X., Xiang Q., Sarroff A.M., Li Z., Wang Y., « Large-scale music tag recommendation with explicit multiple attributes », *MM'10*, Firenze, Italy., October 25–29, 2010.