

ANALYSES QUANTITATIVES DE LA MUSIQUE PAR DES MESURES REALISEES SUR DES SIGNATURES SPECTRO-TEMPORELLES DU SON

Laurent Pottier

CIEREC-Université Jean Monnet
laurent.pottier@univ-st-etienne.fr

Na Wang

Faculté des Sciences-Université Jean Monnet
École des Mines de Saint-Etienne

RÉSUMÉ

L'objectif de ce projet est de développer des outils basés sur une technique d'analyse automatique d'extraits musicaux que nous avons développée et qui permet de caractériser les qualités sonores perçues sur des échantillons de courtes durées, à l'échelle de quelques secondes. Cette analyse est à la fois une analyse spectrale et temporelle et permet de caractériser un extrait sonore du point de vue du timbre global perçu à court terme ainsi que des variations au cours du temps de son flux énergétique. Elle représente sa signature sonore à un moment donné.

Différentes mesures ont été réalisées sur des signatures et permettent d'établir des similitudes, des distances entre les signatures d'un même morceau de musique ou en comparant des musiques différentes. En opérant sur des corpus de référence et en réalisant des apprentissages automatiques, nous cherchons à réaliser différentes applications, concernant l'analyse musicologique, les recommandations pour le streaming et l'élaboration de nouvelles techniques de synthèse sonore.

1. INTRODUCTION

La musicologie a développé des outils permettant d'analyser les musiques instrumentales représentées par des partitions. Nous manquons cruellement d'outils pour représenter et analyser objectivement les musiques qui ne disposent pas de partitions, comme la plupart des musiques électroacoustiques, mais aussi pour celles pour lesquelles la partition n'apporte pas toutes les informations utiles pour l'analyse.

Traditionnellement, les paramètres de la musique occidentale étaient la hauteur, l'intensité des notes et leur répartition au cours du temps. Le reste était l'affaire du timbre. Pourtant dans l'orchestre, un son de violon peut présenter des caractéristiques acoustiques très variables, selon la note jouée, la position de l'archet, les modes de jeu. Aujourd'hui avec les outils électroniques d'amplification et de traitement, les choses sont encore plus compliquées. Par exemple parler du timbre de la guitare électrique est indissociable des traitements qui peuvent lui être associés. Distorsion, *fuzz*, *phasing*, *flanger*, *harmonizer*, réverbération sont autant de

traitements qui induisent des modifications profondes des caractéristiques acoustiques du son de l'instrument.

Avec l'apparition des musiques électroacoustiques, tout son peut devenir musique. S'il est parfois possible d'identifier la source (une casserole, un bruit de train, la voix...), les sons sont rarement utilisés pour leur caractère anecdotique, mais plus pour leur qualité acoustique. Avec l'utilisation des synthétiseurs et des processeurs de signaux, il est possible de créer des sons totalement nouveaux. Enfin, dans les musiques actuelles, le travail en studio devient une part importante de la création, un travail qui modifie la qualité du son en créant divers plans sonores par le mixage.

Aujourd'hui, des mondes sonores totalement nouveaux peuvent être explorés grâce aux technologies numériques. Si dans beaucoup de musiques commerciales le contenu repose surtout sur la mélodie et les textes, des musiciens plus aventureux cherchent aussi à sortir des sentiers battus pour explorer de nouveaux territoires sonores.

Notre recherche porte sur l'identification et la caractérisation du timbre global à court terme de la musique, cherchant à déterminer ce qui fait qu'il nous est possible de reconnaître un artiste à l'écoute d'un extrait musical de seulement une ou deux secondes. Notre objectif est d'essayer de trouver comment caractériser ce timbre, mélange de sons et de rythmes, pour pouvoir le décrire, faire des comparaisons, classer des œuvres.

2. ETAT DE L'ART

La question du timbre s'est souvent limitée à la question de l'instrumentation, le timbre étant défini comme la qualité du son spécifique de l'instrument, indépendamment de sa hauteur et de son intensité. Depuis les travaux d'Helmholtz [8], la description acoustique de ce qui pourrait caractériser le timbre a été associée à la fois à la répartition de l'énergie dans le spectre et la présence des transitoires (attaque, transitions entre les notes). Jean-Claude Risset a montré le caractère dynamique du timbre [18] et Emile Leipp [10] puis Michèle Castellengo [3] ont montré le rôle que peut jouer le sonagramme pour le décrire, mais cette description est finalement inadaptée pour rendre compte des caractéristiques sonores d'une musique faisant intervenir des sons variés superposés. La FFT contient beaucoup trop de données pour être exploitable telle quelle.

Les travaux d’Ernst Terhardt [20], repris par Gérard Assayag à l’IRCAM en 1985 quand il a réalisé le programme Iana ont permis d’extraire de ces analyses spectrales des données condensées, les partiels les mieux perçus par l’audition humaine. Ces outils continuent aujourd’hui à être employés par des compositeurs/trices dans des productions musicales.

« Les combinaisons instrumentales forment des sonorités timbrales complexes complémentées par les traitements et la synthèse en temps réel. Pour cela, les flux audio sont analysés par le programme Iana qui permet d’extraire en temps réel les pics spectraux perceptivement prégnant, puis le résultat est transmis à SuperCollider chargé de produire une resynthèse additive. » à propos d’Estremo d’ombra (2015) [12].

Depuis une dizaine d’années, des travaux ont été engagés dans le domaine des MIR (Music Information Retrieval) pour mesurer des paramètres du timbre, avec la construction de descripteurs audio, comme ceux qui ont été mis au point au milieu des années 2000 avec le MIR Toolbox sous MatLab par Olivier Lartillot et ses collaborateurs [9] ou plus récemment dans les logiciels Sonic Visualizer [4] ou MaxMSP, avec Zsa-descriptors [11] notamment. Certains de ces descripteurs peuvent même être directement intégrés dans une page WEB pour fonctionner en temps réel [21]. Ils permettent d’extraire des représentations spectrales des données d’un ordre supérieur, qui aident à comprendre la façon dont notre perception peut appréhender cette musique.

Des technologies numériques sophistiquées de type *Deep Learning* permettent maintenant d’extraire à partir de fichiers audio des informations musicales complexes, comme le Convolutional Neural Networks (CNNs) et Long Short-Term Memory (LSTM) qui respectivement permettent de calculer des descripteurs de hauteur, de tempo et de basse, pour chercher à tagger automatiquement des musiques sur Internet, ou permettent la classification des phonèmes vocaux chantés, la reconnaissance des instruments de musique et le marquage automatique de la musique [15].

Sur la figure suivante, on observe des matrices d’autosimilarité, les deux axes indiquant le temps, qui permettent de comparer la valeur spectrale moyenne à chaque instant avec celles de tous les autres instants de la pièce. Plus un pixel est sombre, plus les spectres sont proches. Évidemment la diagonale donne une ligne noire.

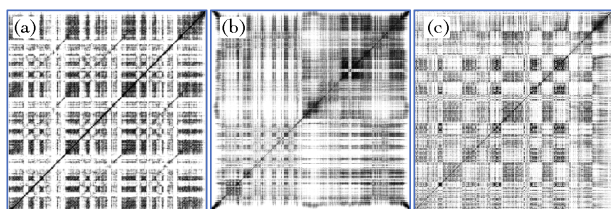


Figure 1. Matrices d’autosimilarité (FFT et timbre) de (a) *Summertime Blues* (E. Cochran, 1958, 2’04,) (b) *Close to the Edge* (Yes, 1972, 18’16), (c) *Sopiana* (F.-B. Mâche, 1980, 11’36).

3. NOTRE PROJET (1^{ÈRE} PHASE)

3.1. Étapes antérieures

Depuis plusieurs années nous sommes à la recherche d’outils permettant de caractériser le sonore, notamment dans le cadre d’un projet mené par le groupe « Analyse de la Musique Electroacoustique (AME) », soutenu par la société française d’analyse musicale¹.

Un premier travail a été réalisé en 2012 (non publié) pour caractériser graphiquement divers sons de guitare extraits de plusieurs pièces des Beatles.

3.1.1. Représentations graphiques de sons de guitare

Comme indiqué précédemment, dans les musiques populaires, le timbre de la guitare électrique peut correspondre à des caractéristiques acoustiques extrêmement variées selon les traitements qui lui sont associés. Sur cinq exemples extraits de plusieurs morceaux différents, la partition donne certaines informations qui sont des pistes pour l’interprètes mais qui sont loin d’exprimer clairement le timbre à produire.

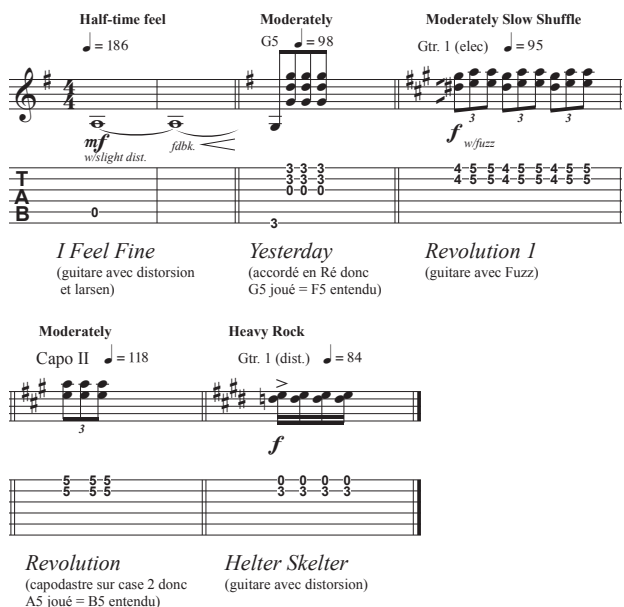


Figure 2. Notation de différentes parties de guitare chez les Beatles.

Les termes employés (*fuzz*, *distorsion*, *slight distorsion*) indiquent seulement des traitements à employer pour transformer le son de la guitare. D’autres termes, comme *Heavy Rock* ou *Slow Shuffle*, sont plutôt des indications de styles et font donc référence à une culture qui peut ou non donner des informations sur le type de timbre à produire.

Dans les outils disponibles dans un logiciel comme Audiosculpt², il est possible d’extraire à partir du son des informations comme les hauteurs perçues (chords), les partiels ou les pics dominants. Pour les cinq mesures de

¹ cf. <http://ame.sfam.org>.

² © IRCAM.

guitares indiquées précédemment, ces informations permettent de mettre en évidence des différences entre les contenus sonores perçus.

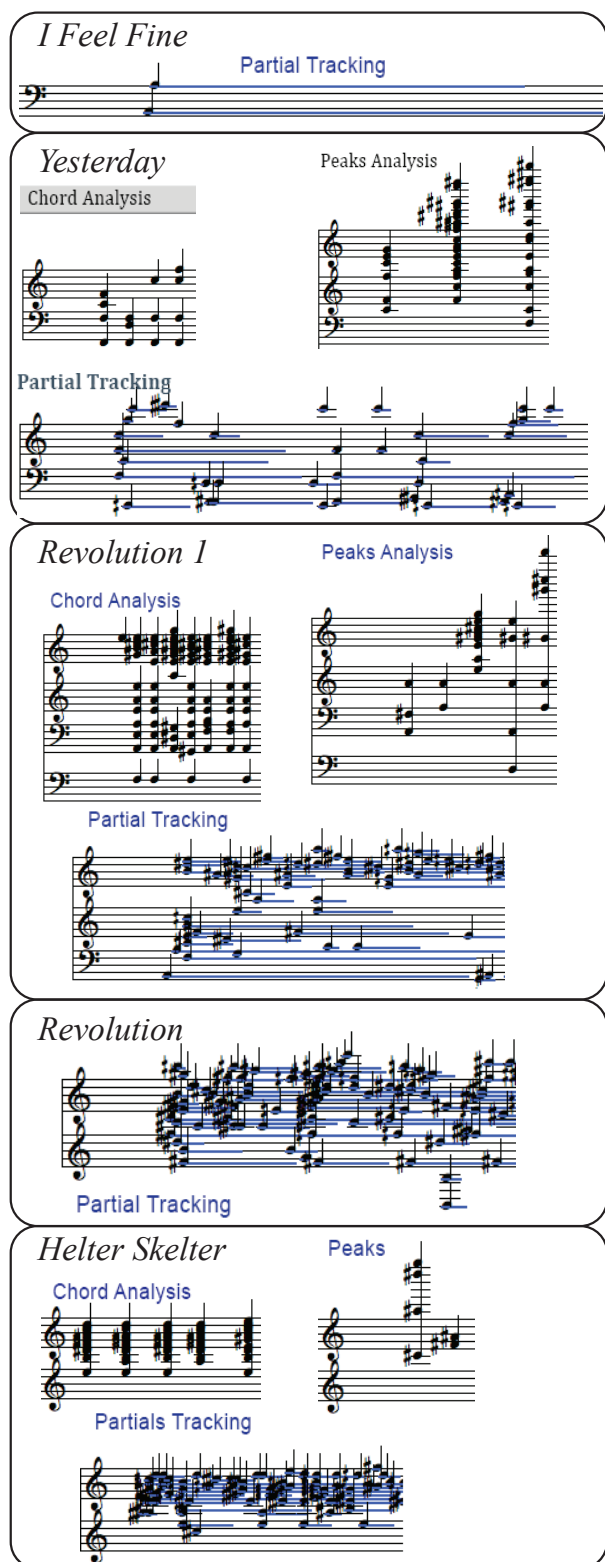


Figure 3. Différentes analyses du timbre perçu. Ces analyses sont relativement peu lisibles. Couplées avec un dispositif de synthèse, elles permettent toutefois

d'écouter la façon dont elles extraient de l'information du spectrogramme, et de décider si ces informations nous semblent pertinentes à l'audition, en fonction de ce que l'analyste cherche à faire ressortir de ces données³.

Nous avons cherché des méthodes pour représenter graphiquement la répartition de l'énergie d'une façon qui soit plus facile à appréhender. A cette occasion nous avons réalisé des analyses de l'énergie RMS dans 27 bandes de fréquences de largeur 1/3 d'octave. Cette division de l'axe des fréquences est proche de celle des Mels [17] et des Barks [24], mais elles sont plus précises dans les fréquences graves, car si les échelles Mels et Barks sont adaptées à la physiologie de l'oreille, la plupart des musiques actuelles et notamment les musiques électroniques produisent beaucoup d'énergie dans les basses fréquences qui ne sont pas perçues en concert que par les oreilles. Ces travaux ont été présentés lors de divers conférences : en janvier 2013 au CIEREC, université Lyon-Saint-Étienne, en mai 2013 lors des JIM 2013 à l'université Paris VIII [2] et lors d'un symposium en mai 2014 au CeReNeM de l'université d'Huddersfield.

Les graphes suivants présentent le temps en abscisse et la répartition de l'énergie dans les différentes bandes de fréquence en ordonnée. La couleur représente la bande de fréquence considérée (du vert clair pour les graves, jusqu'au violet pour les aigus) et la largeur de chaque bande la valeur de l'énergie, au cours du temps.

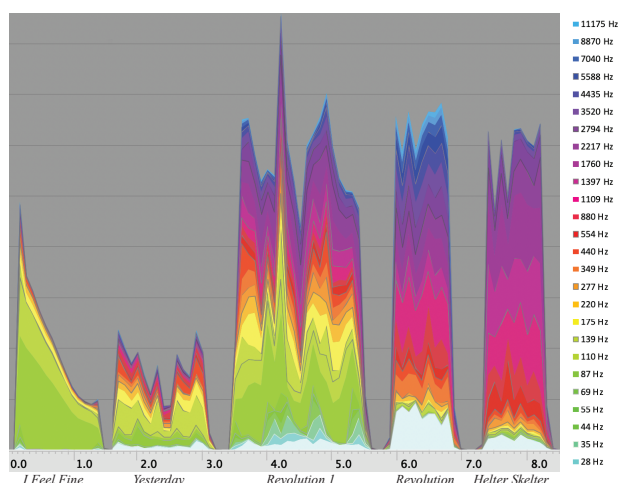


Figure 4. Représentation graphique des variations d'énergie RMS selon les différentes bandes de fréquence au cours du temps.

On voit comment l'usage de la distorsion et de la pédale fuzz produit de l'énergie dans les très hautes fréquences.

3.1.2. Analyse d'Entwurzelt (2012) d'Hans Tutschku

Dans un deuxième temps, nous avons poursuivi ces recherches en cherchant à comparer quantitativement ces tableaux énergétiques, lors d'une étude sur la structure de la pièce *Entwurzelt* d'Hans Tutschku. Entre autres paramètres, l'étude du « mode », valeur la plus énergétique parmi les 27 bandes a permis de souligner la

³ <https://musinf.univ-st-etienne.fr/recherches/analyses/guitBea.html>

structure de la pièce et certaines caractéristiques du timbre dans ses différentes sections. Ces résultats ont été présentées lors de la conférence EMS (Electronic Music Studies) à Berlin en juin 2014.

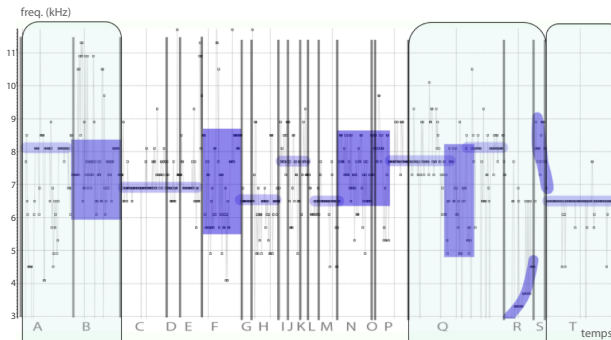


Figure 5. Caractérisation spectrale des différentes sections de la partie électronique de la pièce *Entwurzelt* d'Hans Tutschku. Chaque point indique la fréquence (Hz) de la bande la plus énergétique à un instant donné. Durée totale de la pièce : 16'.

Dans la figure ci-dessus, les lettres indiquent les différentes sections de la pièce. Nous avons surligné les zones dans lesquelles évoluait les modes spectraux. Cette représentation permet de mettre en évidence des proximités de comportement. Par exemple les sections C, D et E ont toutes la même bande unique dans laquelle est concentrée l'énergie. Dans la section F par contre ce mode se déplace sur une dizaine de bande, ce qui indique une section beaucoup plus contrastée. Dans les sections R et S, on observe nettement un glissement de l'énergie des fréquences graves vers les médiums puis des aigus vers une région où elle va se stabiliser (section T).

4. PROJET ACTUEL

Depuis l'été 2017, nos travaux portent sur la caractérisation de la « signature sonore » des musiques en réalisant une analyse spectrale condensée à court terme.

En plus des questions relatives à l'analyse des musiques, nos recherches s'intéressent aussi à la question de la recommandation des musiques sur Internet. En effet, le développement des sites de streaming audio sur Internet a changé notre comportement de consommateur dans le domaine de la musique. Face à la surabondance de données, les entreprises de streaming audio développent des outils de recommandations qui peuvent aider les consommateurs à découvrir de nouvelles musiques. Une entreprise comme Niland [19] utilise des algorithmes qui calculent des distances entre morceaux de musique, basées sur des analyses spectrales et des algorithmes d'apprentissage. Ces technologies sont principalement destinées à déterminer le genre de la musique, l'instrumentation, le tempo, voire l'humeur qu'il produit sur l'auditeur. Ils fonctionnent assez bien sur des musiques populaires assez standard (pop-rock) mais ne sont pas

vraiment adaptés à des musiques moins standardisées, comme la musique électroacoustique par exemple.

4.1. La technologie Niland

En juillet 2017, nous avons réalisés des tests avec divers morceaux de musique, à comparer avec ceux de la base de données musicale de la société 1DTouch⁴ avec les technologies Niland, pour évaluer perceptivement si les calculs de distances que réalisaient les algorithmes en question permettaient d'obtenir de bons résultats en termes de similarité.

Une recherche sur le standard *Rock Around the Clock* (Bill Haley, 1956) a donné par exemple comme résultats des pièces de Chuck Berry, Cliff Richard et Bill Haley. Une pièce du groupe Blessed Virgin (rock stéphanois non édité, 1984) a présenté des œuvres assez proches également, comme *The Ganjas* (2016) des Cure, mais nettement plus édulcorées. L'examen de *Bourée* (1969) une pièce de rock progressif avec flûte traversière de Jethro Tull (d'après J. S. Bach), a donné une œuvre fort éloignée, *On the Green Light* (1966) de The Spencer Davis Group. Un extrait de *L'Œil écoute* (1973) de Bernard Parmeggiani (musique électroacoustique) a donné *les Batteries du Premier Empire* par Les Equipages de la Flotte de Toulon et *Oleaje* par Sonidos de la Naturaleza Relax (musique ambiante) et enfin *Nidiaye* de Toure Kunda (world musique) a donné *Gitara U Noći* de Razni Izvođači. Dans ces deux derniers cas, avec des musiques utilisant des sonorités assez particulières, les correspondances trouvées ont été extrêmement éloignées des sources proposées. On constate donc que les critères de reconnaissance et de calcul de distance fonctionnent bien quand certains instruments phare et le style de musique sont reconnus, la qualité de la voix joue aussi beaucoup, mais pris globalement le timbre ne joue pas un très grand rôle.

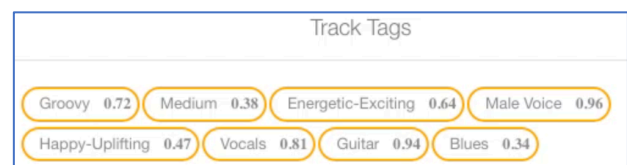


Figure 6. Les tags utilisés avec les valeurs comparant *Rock Around the Clock* et *Tu parles trop* de Samy Daussat

4.2. Technique d'analyse

Notre objectif était donc de mettre au point une technique d'analyse permettant de condenser l'information contenue dans un spectrogramme à l'échelle de quelques secondes pour n'en retenir que l'essentiel du point de vue du spectre et de ses variations temporelles en relation avec la perception.

Sur des échantillons sonores de courte durée (une à trois secondes), nous sommes à même de reconnaître à l'oreille le style, l'instrumentation, et même l'œuvre si

⁴ <http://1dtouch.com/>, plateforme de streaming multimédia équitable.

elle nous est familière. Notre analyse, à la fois spectrale et temporelle, doit donc permettre de caractériser simultanément un extrait sonore du point de vue du timbre et de son flux temporel, d'où son appellation comme **signature sonore**.

4.2.1. Protocole

Comme dans les exemples précédents, nous avons divisé le son en 27 canaux, répartis approximativement en 1/3 d'octave, puis nous mesurons l'énergie RMS de chaque bande à différents taux d'échantillonnage (de plus en plus fine lors de la montée vers les hautes fréquences).

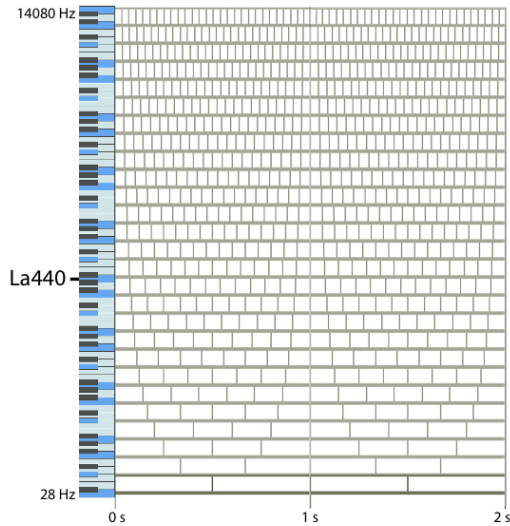


Figure 7. Structure de la table d'analyse.

| | | SR : | 44100 | SR : | 44100 | MIDI | écart |
|----------|-----------------|-------------------|---------------|---------------|-------------------|-------------|---------|
| n° bande | intervalle freq | 1/3 octave exacts | raie FFT 4096 | nbre de raies | 1/3 octave approx | note approx | 1/2 ton |
| 1 | 27-34 | 27.5 | 2 | 1 | 21.5 | 17 | 7.0 |
| 2 | ... | 34.6 | 3 | 1 | 32.3 | 24 | 5.0 |
| 3 | | 43.7 | 4 | 1 | 43.1 | 29 | 3.9 |
| 4 | | 55.0 | 5 | 1 | 53.8 | 33 | 3.2 |
| 5 | | 69.3 | 6 | 2 | 64.6 | 36 | 5.0 |
| 6 | | 87.3 | 8 | 2 | 86.1 | 41 | 3.9 |
| 7 | | 110.0 | 10 | 3 | 107.7 | 45 | 4.5 |
| 8 | | 138.6 | 13 | 3 | 140.0 | 49 | 3.6 |
| 9 | | 174.6 | 16 | 4 | 172.3 | 53 | 3.9 |
| 10 | | 220.0 | 20 | 5 | 215.3 | 57 | 3.9 |
| 11 | | 277.2 | 25 | 7 | 269.2 | 60 | 4 |
| 12 | | 349.2 | 32 | 9 | 344.5 | 65 | 4 |
| 13 | | 440.0 | 41 | 10 | 441.4 | 69 | 4 |
| 14 | | 554.4 | 51 | 14 | 549.1 | 73 | 4 |
| 15 | | 698.5 | 65 | 16 | 699.8 | 77 | 4 |
| 16 | | 880.0 | 81 | 22 | 872.1 | 81 | 4 |
| 17 | | 1108.7 | 103 | 26 | 1109.0 | 85 | 4 |
| 18 | | 1396.9 | 129 | 34 | 1388.9 | 89 | 4 |
| 19 | | 1760.0 | 163 | 43 | 1755.0 | 93 | 4 |
| 20 | | 2217.5 | 206 | 53 | 2217.9 | 97 | 4 |
| 21 | | 2793.8 | 259 | 68 | 2788.5 | 101 | 4 |
| 22 | | 3520.0 | 327 | 85 | 3520.7 | 105 | 4 |
| 23 | | 4434.9 | 412 | 107 | 4435.8 | 109 | 4 |
| 24 | | 5587.7 | 519 | 135 | 5587.9 | 113 | 4 |
| 25 | | 7040.0 | 654 | 170 | 7041.4 | 117 | 4 |
| 26 | ... | 8869.8 | 824 | 214 | 8871.7 | 121 | 4 |
| 27 | 11k-14k | 11175.3 | 1038 | 269 | 11175.7 | 125 | 4 |
| | | 14080.0 | 1307 | | 14071.9 | 129 | |

Tableau 1. Détail des 27 bandes de fréquences, pour un taux d'échantillonnage de 44,1kHz et une taille de la FFT de 4096 échantillons.

Pour simplifier les calculs, le programme regroupe des raies de la FFT, ce qui ne donne pas exactement des bandes en tiers d'octave pour les fréquences graves. Dans le tableau ci-dessus, la colonne de droite indique (en demi-tons) les intervalles de chaque bande. A partir de la note MIDI 60 (do3) tous les intervalles font 4 demi-tons.

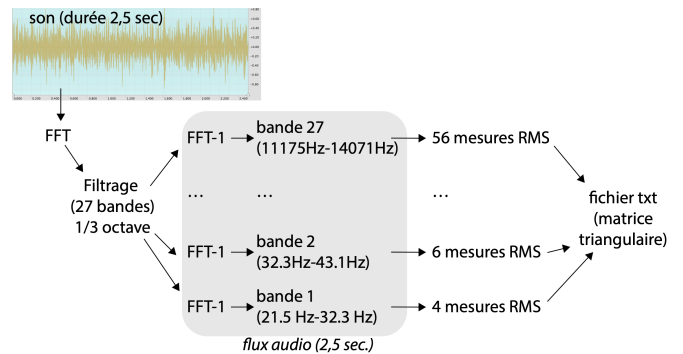


Figure 8. Les étapes de l'analyse.

Nous réalisons ensuite un affichage graphique de cette signature, à partir d'un code couleur allant du vert au bleu, (de -96 dB à 0 dB).

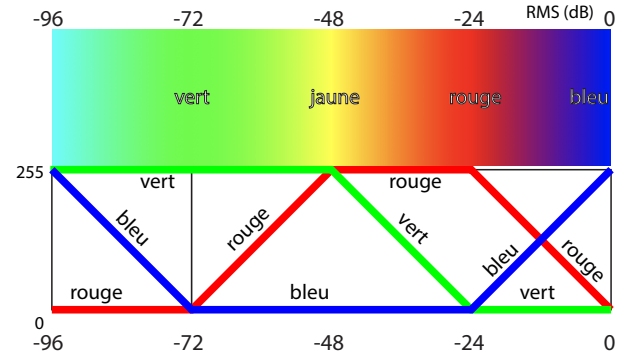


Figure 9. Correspondances couleurs/énergie RMS de chaque bande.

Un second graphique (graphes inférieurs sur la figure suivante) est également réalisé en effectuant un tri sur l'axe temporel de façon à obtenir une sorte de profil spectral global de la signature.

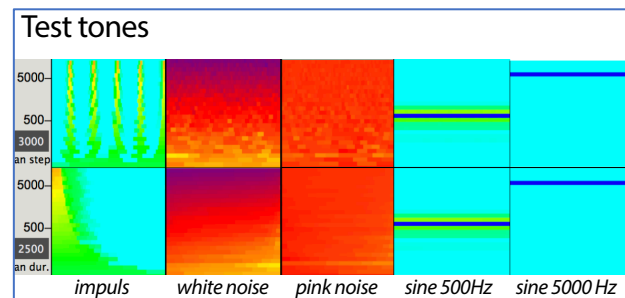


Figure 10. Signature sonore de divers sons tests.

Nous avons également réalisé deux autres représentations graphiques placées à droite des précédentes dans les figures suivantes. En haut à droite, les bandes sont triées verticalement par énergie décroissante, en bas à droite, ces mêmes bandes sont triées de plus par énergie décroissante horizontalement.

L'énergie se retrouve alors concentrée en haut à gauche de la figure.

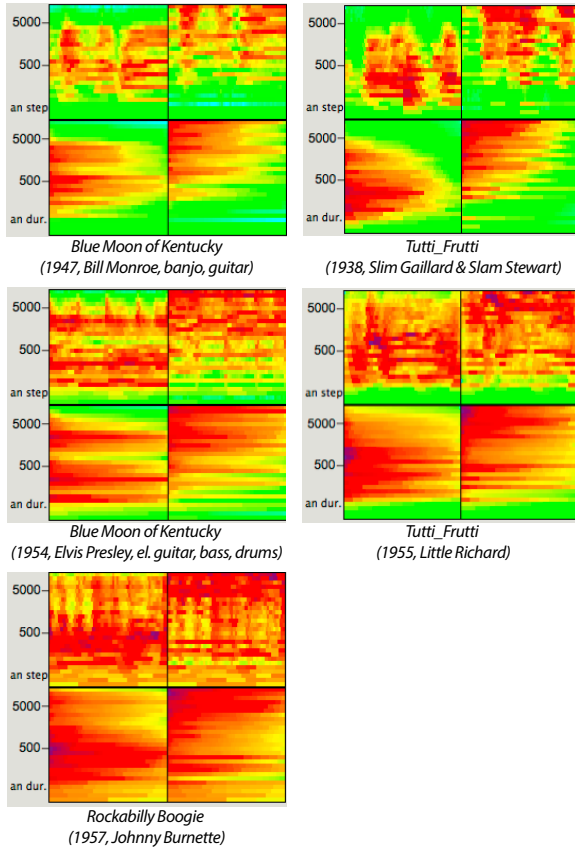


Figure 11. Signatures sonores de plusieurs pièces pionnières du rock'n roll (issues du corpus de [7]).

Dans ces quelques signatures sonores, on constate en dix ans un enrichissement très marqué du spectre, à la fois dans les fréquences graves et dans les fréquences aiguës, principalement lié à l'électrification progressive des instruments et notamment à l'usage des effets de distorsion du son sur la guitare.

4.2.2. Langages utilisés

La version 2018 du programme d'analyse a été réalisée en Max⁵. Nous avons réalisé un outil qui fait les analyses des signatures en temps réel, pendant la lecture et l'écoute du fichier audio, qui affiche leurs représentations graphiques au fur et à mesure et enregistre les fichiers d'analyse (format texte) sur disque. Le programme peut faire une analyse simple, l'analyse d'un morceau complet, ou l'analyse d'un répertoire contenant tout un corpus d'œuvres.

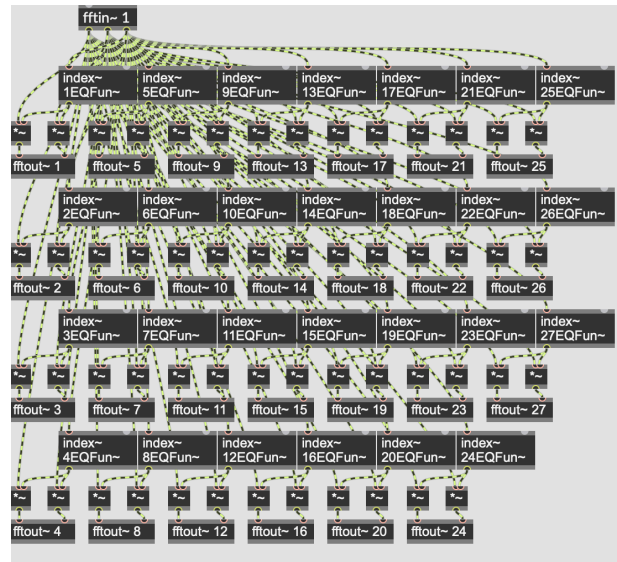


Figure 12. La partie DSP faisant la FFT et la FFT inverse pour générer les flux audio de chaque bande

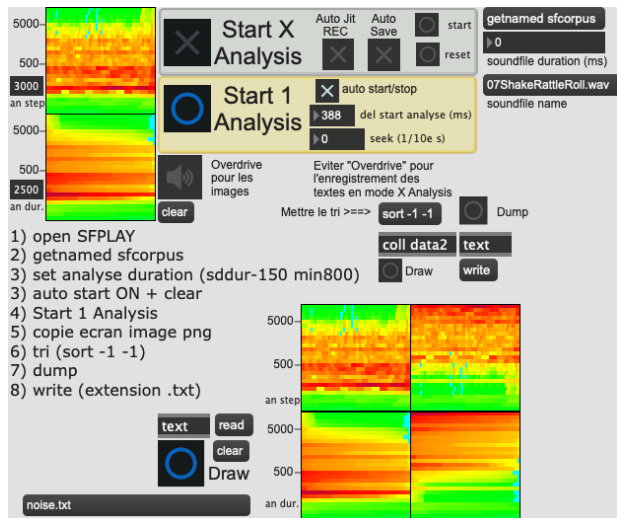


Figure 13. La visualisation successive des signatures en temps réel (pendant la lecture des sons), et les différentes modes : analyse simple (*I Analyses*), analyse d'un morceau entier (*X analysis*), d'un répertoire (*getnamed sfcopus*).

En janvier 2019, Mahdi Mohammed-Bashir a réalisé en Python une version du programme d'analyse en utilisant la bibliothèque librosa⁶. Il l'a ensuite intégrée dans une page WEB en utilisant la plateforme Node.js⁷.

Le programme comporte également les trois modes, il génère les fichiers textes des analyses, les graphiques des analyses et en option les fichiers audio générés par les fft inverses de chaque bande de fréquence.

⁵ © cycling74.com

⁶ <https://librosa.github.io/librosa>

⁷ <https://nodejs.org/en/>

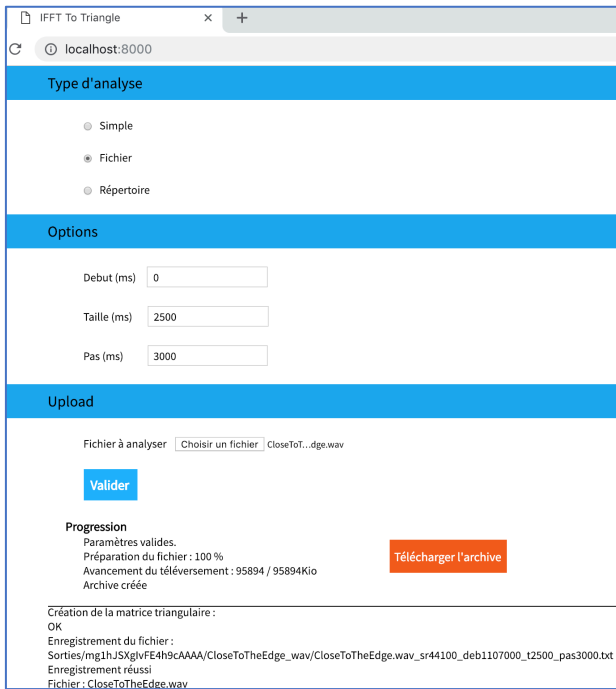


Figure 14. Version Web/Python/Node.js de l'analyseur.

4.3. Descripteurs audio

En 2018, nous avons programmé en Python plusieurs descripteurs audio [23] calculés à partir de nos signatures en nous inspirant notamment des formules indiquées par Geoffroy Peeters [13] ou des programmes réalisés pour Max par Emmanuel Jourdan et al [11].

Nous avons calculé ainsi le **centroïde spectral** (en Hz), centre de gravité de l'énergie sur l'axe des fréquences et **deux écarts-types** (*ecartlow* et *ecarthigh*) indiquant la déviation moyenne de l'énergie au-dessous et au-dessus du centroïde, ainsi que la **penne spectrale** moyenne (*slope*), l'**asymétrie spectrale** (*skewness*) qui donne une mesure de l'asymétrie du spectre autour de sa valeur moyenne (une asymétrie nulle indique une distribution symétrique, une asymétrie négative implique plus d'énergie à des fréquences inférieures à la valeur moyenne). Nous avons aussi calculé la « **planéité** » **spectrale** (*flatness* cf. équation 1), mesure de la distance entre un bruit (valeur = 1) et un son pur (valeur = 0), obtenue en calculant le rapport entre la moyenne géométrique et la moyenne arithmétique du spectre, et la « **crête** » **spectrale** (*crest*), un autre descripteur lié également à la planéité du spectre et calculé par le rapport entre la valeur maximale obtenue dans le spectre et la moyenne arithmétique du spectre.

$$flatness = \frac{\sqrt[k]{\prod_{k=0}^{K-1} S(k)}}{\frac{1}{K} \sum_{k=0}^{K-1} S(k)} \quad (1)$$

| titre | centroid (Hz) | spectral_slope | spectralflatness | spectral_crest |
|-------------------------|---------------|----------------|------------------|----------------|
| 0-Impuls/impuls2hz.txt | 3334.6 | 0.0038 | 0.83 | 29.51 |
| 0-Silence/silence.txt | 0.0 | 0.0000 | 0.00 | 0.00 |
| 1-Noise/noise.txt | 3971.3 | 0.0060 | 0.67 | 3.60 |
| 2-Pink/pink.txt | 1919.4 | -0.0011 | 0.99 | 1.65 |
| 3-Sin500/sinus500.txt | 492.4 | -0.0061 | 0.00 | 27.01 |
| 4-Sin5000/sinus5000.txt | 4978.6 | 0.0095 | 0.00 | 27.01 |

Figure 15. Valeurs de quelques descripteurs spectraux sur les signatures de sons tests.

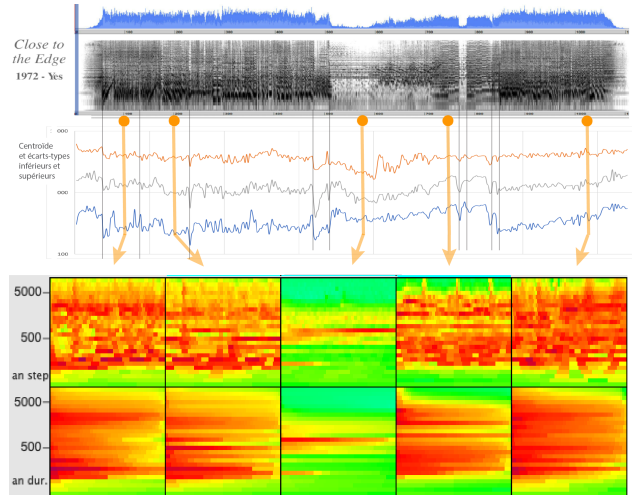


Figure 16. *Close to the Edge*, de haut en bas : enveloppe, spectrogramme, graphe du centroïde et des écarts, signatures à quelques instants clés.

En poursuivant l'étude de l'évolution du son de la guitare dans le domaine des musiques rock, nous avons constaté que le paramètre *spectral flatness* donnait une très bonne idée de la saturation du spectre.

Nous avons effectué toutes les signatures sonores de chaque morceau d'un corpus de plus d'une centaine de pièces de Rock Métal (issues du corpus de [1]), pour comparer la saturation du son dans les différents morceaux et observer les variations au sein d'un même morceau. Sur les dix-huit premiers morceaux de ce corpus, le « spectral flatness » met assez bien en évidence les morceaux très denses, très saturés (voir les parties surlignées sur le graphique de la figure suivante).

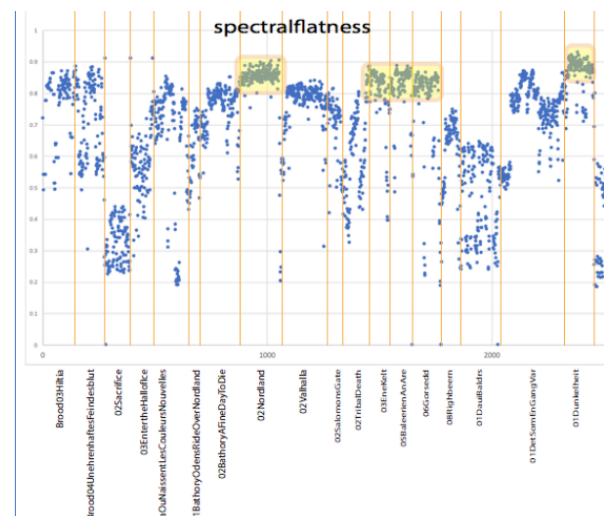


Figure 17. Calcul du paramètre *spectral flatness* sur des morceaux de musique Black Métal.

En calculant ensuite les moyennes de toutes les signatures sonores de chaque morceau du corpus, nous avons ensuite trié ces œuvres par ordre de *flatness* décroissant (cf figures suivantes).

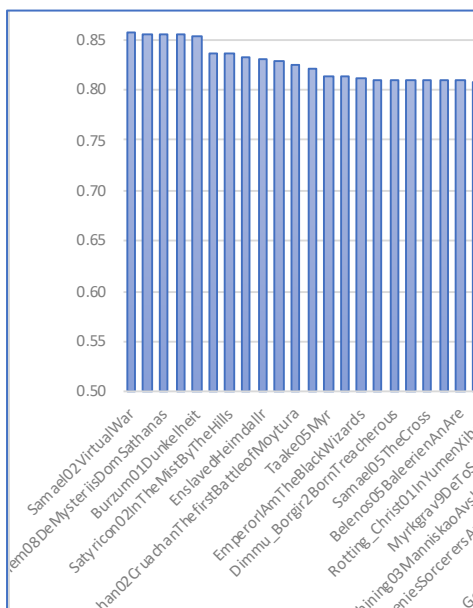
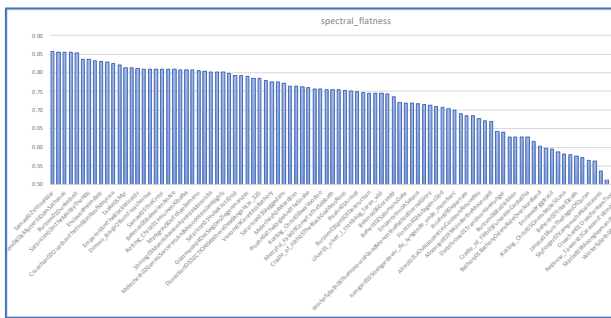


Figure 18. Tri des œuvres de Rock Metal selon la valeur de *flatness*. en haut toutes les pièces, en bas un zoom sur les valeurs de *flatness* les plus élevées.

Le morceau avec la valeur de *flatness* la plus élevée, *Virtual War* (2009) du groupe Samael⁸, présente une signature extrêmement dense totalement saturée, proche du bruit rose, mais avec des régions encore plus énergétiques.

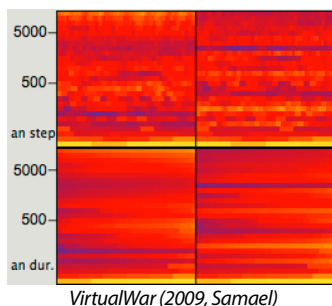


Figure 19. Signature totalement saturée.

5. PERSPECTIVES

- Nous avons surtout étudié pour l'instant des descripteurs spectraux. Nous nous intéressons maintenant aux descripteurs rythmiques permettant de mesurer les variations temporelles du flux spectral selon les différentes bandes de fréquence.
- Nous cherchons des méthodes qui nous permettront de calculer des distances entre les signatures qui correspondent le mieux avec la perception, selon plusieurs critères possibles.
- En établissant un certain nombre de descriptions sémantiques des musiques analysées, nous souhaitons développer des algorithmes de fouille de données et d'apprentissage portant sur la base de données de musiques de l'entreprise 1DTouch dans le but de réaliser des recommandations audacieuses sur des musiques présentant des qualités sonores similaires.
- Nous avons aussi commencé à explorer les possibilités de développer des techniques de synthèse à partir des signatures sonores. Un premier prototype a été réalisé pour synthétiser les signatures en générant des bandes de bruits modulés par les données des signatures et qui permettent à l'oreille de comparer ces signatures synthétisées avec les sons d'origine.

6. REMERCIEMENTS

Ce projet est encore en phase de démarrage. Nous tenons à remercier la Fondation UJM qui a financé son lancement et nos collaborateurs : la compagnie 1DTouch, le laboratoire Vibrations Acoustique (INSA-Lyon), le laboratoire Hubert Curien (UJM) et GRAME (Lyon). Un grand merci en particulier à Fabrice Muhlenbach et Pierre Maret qui ont aidé à démarrer les opérations d'apprentissage ainsi qu'à Joseph Chataignon (premiers développements autour de la base de données MySQL), David Boit (DSI, configuration de la base) et Mohammed-Bashir Mahdi (développement de la première version Python et Web via Node.js de l'analyse).

7. REFERENCES

- [1] Bera, C. "Identités nationales et racines culturelles du Black Metal". Mémoire de Master, UJM, Saint-Etienne, 2013.
- [2] Bonardi, A. ; Bossis, B. ; Couprie, P. ; Dufeu, F. ; Malt, M. ; Pottier, L. "Outils pour l'analyse de la musique électroacoustique". Actes des Journées d'Informatique Musicale (JIM 2013), Univ. Paris VIII, p.199-200.
- [3] Castellengo, M. *Écoute musicale et acoustique. Avec 420 sons et leurs sonagrammes décryptés*. Eyrolles, Paris, 2015.
- [4] Cook, N. & Leech-Wilkinson, D. *A musicologist's guide to Sonic Visualiser*. King's College London, 2009.

⁸ <https://www.youtube.com/watch?v=3qYHxzK0vjE>

URL :

http://www.charm.rhul.ac.uk/analysing/p9_1.html.

- [5] Diefenbach, D. ; Lhérisson, P.-R. ; Muhlenbach, F. ; Maret P. "Computing the Semantic Relatedness of Music Genre using Semantic Web Data". *SEMANTiCS 2016*. URL : <http://ceur-ws.org/Vol-1695/paper23.pdf>
- [6] Fletcher, H. & Munson, W. A. "Loudness, its definition, measurement and calculation". *Bell Labs Technical Journal*, vol. 12, ISBN 4, 1933, p. 377–430.
- [7] Gilles, G. "Les représentations de la sauvagerie dans le rock'n'roll américain des années 1950, entre mythes et réalités", université Paris-Est – Marne-la-Vallée, 2012.
- [8] Helmholtz, H. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. New York, Dover, 1863, R/1954.
- [9] Lartillot, O. & Toivainen, P. "A Matlab Toolbox for Musical Feature Extraction From Audio". *International Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, 2007.
- [10] Leipp, E. *Acoustique et Musique*. Presse des Mines, ParisTech, 4^e édition de 1984, 2010.
- [11] Malt, M. & Jourdan, E. "Zsa.Descriptors: a library for real-time descriptors analysis". 5^e Sound and Music Computing Conference (SMC), Berlin, 2008, p.134-137.
- [12] Morciano L. *Écriture du son, du temps et de l'espace dans l'interaction entre instruments et dispositifs numériques synchrones*. Thèse de doctorat SACRe, Univ. PSL, Paris, 2018, p.94-95.
- [13] Peeters, G. "Descripteurs audio : de la simple représentation aux modèles de connaissances. Geste sonore et paramètres. L'analyse musicale à l'heure des outils multimédia". 2015 <hal-01254068>
- [14] Peeters, G. ; Giordano, B. L. ; Susini, P. ; Misdariis, N. ; McAdams S. "The timbre toolbox: Extracting audio descriptors from musical signals". *The Journal of the Acoustical Society of America*, vol. 130, n°5, 2011, p.2902–2916.
- [15] Pons, J. ; Slizovskaia, O. ; Gong, R. ; Emilia Gomez E. ; Serra, X. "Timbre Analysis of Music Audio Signals with Convolutional Neural Networks". *EUSIPCO*, Grèce, 2017.
- [16] Pottier, L. "Local Sound Signatures for Music Recommendations". *Proceedings of Timbre 2018: Timbre Is a Many-Splendored Thing*, 5-7 Juillet 2018, Montréal, Québec, Canada, 2018, p.65-66. https://www.mcgill.ca/timbre2018/files/timbre2018/timbre2018_proceedings.pdf
- [17] Rabiner, L. & Juang B. *Fundamentals of speech recognition*. New York, Prentice-Hall, 1993.
- [18] Risset, J.-C. "An Introductory Catalogue of Computer Synthesized Sounds". (v. o. 1969), The Historical CD of Digital Sound Synthesis, *Computer Music Current 13*, Wergo, Mayence, WER 2033-2 (CD), rééd. 1995.
- [19] Tardieu, D. (Co-fondateur & CEO de Niland). "L'intelligence artificielle va changer la façon dont vous écoutez de la musique". URL : http://www.huffingtonpost.fr/damien-tardieu/intelligence-artificielle-recommandation-musique_b_9700808.html
- [20] Terhardt, E. ; Stoll, G. ; Seewa, M. "Algorithm for extraction of pitch and pitch salience from complex tonal signals". *Journal of the Acoustical Society of America*, vol 71, n°3, 1982, p.679- 688.
- [21] Rawlinson, H. ; Segal, N. ; Fiala J. "Meyda: an audio feature extraction library for the Web". *Web Audio Conference, IRCAM*, Paris, 2015.
- [22] Serrà J., Serra X., ; Andrzejak R. G. "Cross recurrence quantification for cover song identification". *New Journal of Physics*, Vol. 11, n°9, 093017, 2009.
- [23] Wang, N., *Spectral and rhythmic signatures for musical recommendations*, Master 1 Internship Thesis, Cyber-Physical-Social-Systems (CPS2) Faculté des Sciences et Techniques, Université Jean Monnet & École des Mines, Saint Etienne, juin 2018.
- [24] Zwicker, E. & Feldtkeller R. *Psychoacoustique : l'oreille, récepteur d'information*. Traduit de l'allemand par Christel Sorin, Masson, Paris, 1981.