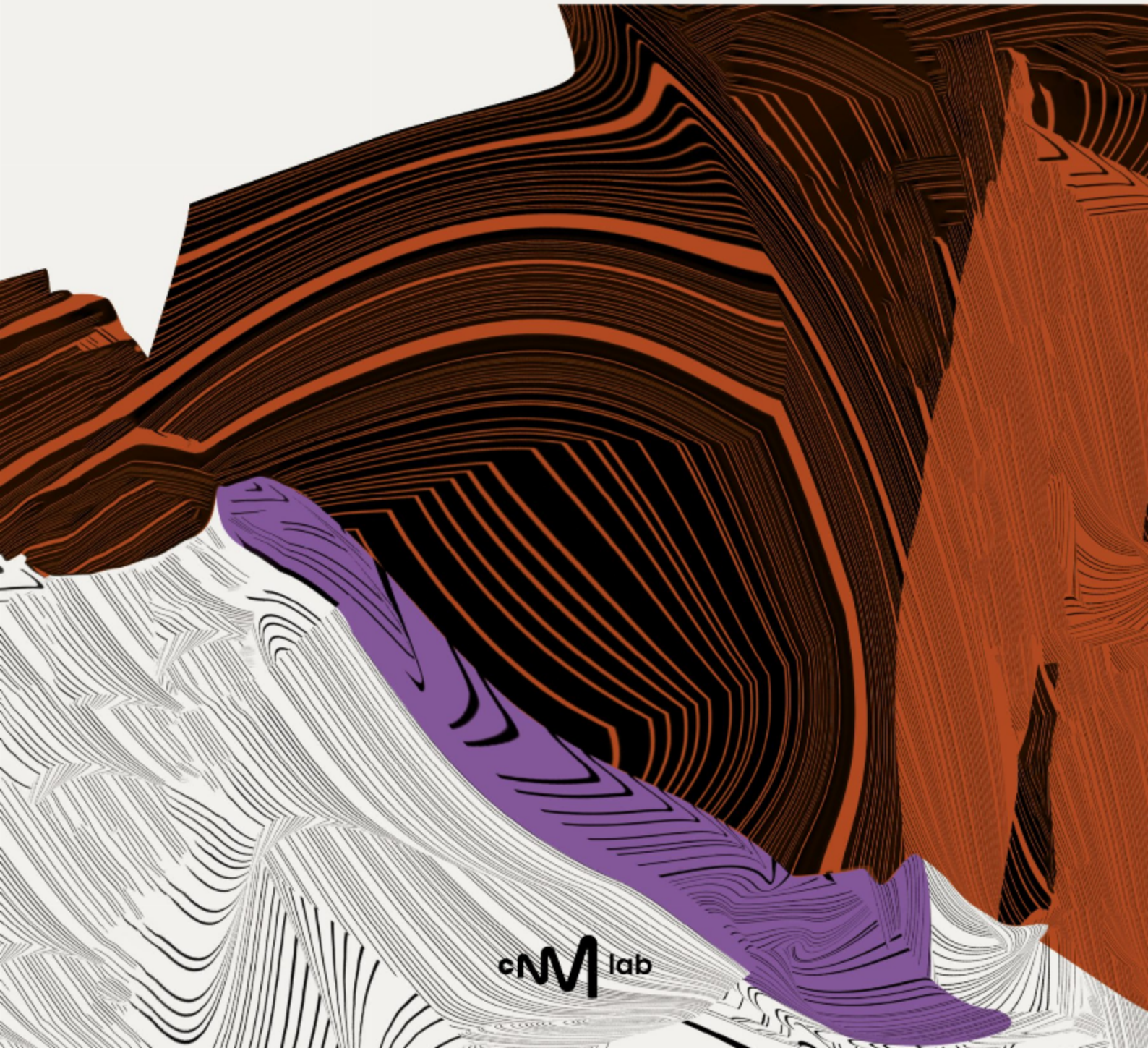
L'analyse musicale et sonore

De la recherche à la recommandation

Par Laurent Pottier



Pottier Laurent

Laurent Pottier est professeur de musicologie à l'université Jean Monnet (Saint-Étienne). Il a créé en 2011 le master professionnel « Réalisateur en informatique musicale ». Il a enseigné à l'Ircam (1992-1996), puis a dirigé le secteur recherche au GMEM à Marseille (1997-2005). Il a collaboré avec de nombreux compositeurs contemporains. Ses recherches portent actuellement sur la préservation et l'analyse des musiques utilisant des instruments électroniques, ainsi que sur le développement de lutheries numériques.

Introduction

Ce texte témoigne de travaux de recherche menés depuis de nombreuses années au sein du département de musicologie de l'université de Saint-Étienne, qui ont pour but l'étude et l'analyse des musiques des XX^{e} et XXI^{e} siècles, en particulier celles qui utilisent des instruments de musiques électroniques, ou, plus généralement, les technologies électronumériques. Nous cherchons à décrire et à mesurer ce qui est perçu à l'écoute d'un morceau de musique, à comprendre les caractéristiques du son, afin d'analyser, de comparer et de classer des œuvres musicales selon des critères auditifs. Cela suppose de se situer au carrefour de plusieurs disciplines scientifiques : la musicologie, l'acoustique, l'informatique et le traitement du signal. En parallèle au caractère purement musicologique de nos travaux, notre volonté est de concevoir des outils qui permettraient de travailler sur de vastes bases de données de morceaux de musique avec pour objectif d'améliorer les techniques de recommandation musicale. À l'heure actuelle, les fonctionnalités proposées par les plateformes de streaming sont assez limitées au regard de la sonorité, du timbre qui caractérise tel ou tel morceau de musique. C'est pourquoi nous développons des logiciels à destination des internautes, qui visent à augmenter et diversifier les résultats obtenus lors de la recherche musicale suivant des critères psychoacoustiques. En affinant ces outils et en les combinant, nous cherchons à réaliser des classifications de la musique qui s'appuieraient sur le timbre et, plus précisément, sur la notion de « signature sonore ».

Cet article aborde la démarche scientifique qui nous a conduits à étudier les principaux paramètres du son que les technologies actuelles permettent de mesurer. Dans cette optique, nous retrouverons ici le contexte et les enjeux de ces outils, mais également les caractéristiques techniques de ceux-ci, afin de mieux appréhender leur fonctionnement et leurs possibles usages.

1. Rappels historiques

La musicologie

La musicologie est une discipline universitaire dont l'une des activités consiste à décrire et analyser la musique. Si, historiquement, la musique occidentale était principalement représentée par des notes figurant sur une partition, le timbre [1] a pris de nouvelles dimensions dès le début du XX^{e} siècle, notamment par le biais de certaines œuvres de compositeurs comme Arnold Schönberg (1874-1951) explorant les possibilités de l'atonalité, Claude Debussy (1862-1918) produisant des musiques impressionnistes, Igor Stravinsky (1882-1971) [2] ou encore Edgar Varèse (1883-1965) intégrant dans l'orchestre des instruments de percussion et des instruments électroniques.

Après les années 1950, le travail des compositeurs sur le timbre a subi de profondes mutations. Ils ont exploré de nouveaux modes de jeu sur les instruments de musique, comme le recours aux pianos préparés [3] ou aux sons multiphoniques pour les instruments à vent ou à cordes. L'innovation a aussi concerné les voix, les compositions pouvant intégrer des cris, des gémissements, des souffles, des rires [4], ou encore l'écriture à travers l'exploration de la microtonalité. Iannis Xenakis (1922-2001) a produit des œuvres orchestrales à partir des lois de la thermodynamique avec sa musique « stochastique » formée de nuages de sons dont les paramètres reposaient sur la théorie des probabilités. Les pratiques développées par les compositeurs qui ont fondé l'école spectrale avec l'ensemble de création musicale L'Itinéraire dans les années 1970 ont permis de composer le son, d'après des analyses spectrales d'échantillons audio, en faisant fusionner les instruments de l'orchestre, produisant un « continuum harmonie-timbre [5] ». Plus récemment, des compositeurs contemporains ont créé un courant de composition « qui a fait de la saturation [instrumentale] un défi, comme un nouveau champ [6] » à explorer.

Mais c'est l'utilisation des technologies électroniques et numériques qui a le plus profondément modifié la nature des musiques produites depuis les années 1950. Pierre Schaeffer (1910-1995) a inventé la musique concrète consistant à composer de la musique directement par montage de sons enregistrés provenant de n'importe quel matériau sonore. Karlheinz Stockhausen (1928-2007) a été le principal pionnier de la musique électronique, dont tout le matériau est produit entièrement par des moyens électroniques (oscillateurs, générateurs de bruits, filtres). À la fin des années 1950, les travaux de Max Mathews aux laboratoires Bell ont ouvert la voie à l'analyse, à la synthèse et au traitement du son sur ordinateur.

En parallèle, dans les musiques populaires pop-rock, l'utilisation de l'électricité pour amplifier et transformer les sons des instruments a permis de donner au travail de production en studio un rôle très important pour définir les identités sonores des artistes. La naissance du rock'n'roll a fait suite à l'amplification de la guitare, dont le volume a ainsi pu rivaliser avec les sons de la batterie. L'apparition des effets (échos, réverbération, filtres, wah-wah, distorsion) a fait de la guitare un instrument incontournable, et avec l'augmentation des décibels en concert sont apparus les groupes de hard rock, de punk, de heavy metal, de musique industrielle et de musique noise, faisant passer le son saturé, chargé de bruit, au premier plan.

Si, par le passé, la musique pouvait être décrite par les notes, les rythmes et les instruments utilisés, il faudrait désormais aussi décrire les traitements audio utilisés, les techniques de synthèse employées et les valeurs des paramètres choisies pour qualifier précisément ce que l'auditeur est susceptible d'entendre dans la musique.

Mais ici, plutôt que de décrire les processus de production du son, nous cherchons à décrire les caractéristiques acoustiques de ce qui est entendu, ce qui constitue le timbre, non pas d'un instrument, mais d'un extrait musical, correspondant souvent à la superposition de plusieurs sons.

L’acoustique et la psychoacoustique

L’acoustique est la science du son. C’est une discipline scientifique relevant du domaine de la physique. Elle est divisée en trois secteurs : la production du son, sa propagation et sa réception ou ses effets. L’acoustique musicale est le domaine de l’acoustique consacré à la place et à l’utilisation du son dans l’élaboration et la perception de la musique. C’est aussi l’ensemble des lois physiques qui régissent les fondements de la théorie musicale.

Selon François Delalande [7], le mot « son » a de nombreuses affectations : en musique, on peut parler d’un accord de trois « sons » pour indiquer que trois notes différentes sont jouées simultanément. Avec l’enregistrement, on parle maintenant du « son » d’un disque, qui caractérise le travail réalisé en studio. En électroacoustique, un « son » est un « objet sonore », un échantillon à partir duquel on crée une œuvre. Dans le domaine des musiques instrumentales, on peut distinguer le « son » d’un instrument (le timbre), le « son » d’un instrumentiste, le « son » d’un ensemble orchestral, le « son » d’un enregistrement et jusqu’au « son » d’une chaîne hi-fi. Toutes ces acceptions du terme « son » restent extrêmement qualitatives.

La plupart des études portant sur le son en acoustique musicale s’intéressent aux timbres des instruments. Mais, comme nous l’avons déjà indiqué, l’évolution des techniques d’enregistrement en studio, le développement des musiques électroniques, utilisant des sons produits par synthèse, ou des musiques électroacoustiques, pour lesquelles la composition est basée sur des sons d’origines très variées, ont fortement modifié la notion de timbre [8]. Cette dernière ne peut plus être liée uniquement à l’origine instrumentale des sons, mais plutôt à un ensemble de qualités acoustiques qu’il s’agit de définir.

La qualité et la nature du son que nous percevons dépendent d’une combinaison de différents paramètres acoustiques, mais les études pour analyser et comprendre les liens entre les valeurs mesurées de ces paramètres et la perception par un auditeur sont assez récentes et très partielles. Le champ d’études est immense et reste à explorer, pour mieux comprendre comment les sons de différentes musiques se ressemblent, pour pouvoir établir des classifications ou pour analyser finement la musique.

Jusqu’à une période récente, les seules descriptions possibles du son étaient liées à son spectre et aux sonagrammes et étaient principalement qualitatives, mais l’acoustique et la psychoacoustique sont des domaines scientifiques qui ont bénéficié de nombreuses avancées technologiques ces dernières décennies avec le développement de l’informatique et des techniques de traitement du signal. On dispose dorénavant d’outils permettant de mesurer un certain nombre de paramètres du son à l’aide de ce qu’on appelle communément les « descripteurs audio ». Chacun de ces descripteurs met en évidence une certaine caractéristique du son. En étudiant les valeurs obtenues par l’analyse, on peut mesurer des distances dans des extraits d’œuvres de musique pour analyser et comparer ces œuvres.

2. Les outils numériques dans l'analyse acoustique

La connaissance des caractéristiques des sons s'est développée avec l'apparition des outils électroniques au milieu du XX^{e} siècle (oscilloscopes, spectroscopes) et surtout avec l'avènement de l'informatique et des outils numériques d'analyse du son.

Le spectre et le sonagramme

Un son dans l'air peut être défini comme une perturbation oscillatoire qui se propage à partir d'une source. Les variations de la pression de l'air engendrées par une onde sonore au cours du temps peuvent être captées par un microphone et converties en courant électrique. La visualisation de cette onde ne permet pas de comprendre la nature du son, sauf dans le cas de sons très simples.

La première avancée majeure pour décrire le son a été offerte par les représentations spectrales du son, grâce aux transformées de Fourier (1768-1830). Ce mathématicien a montré que tout signal complexe pouvait être décomposé comme somme de signaux élémentaires. Les transformées de Fourier sont des opérations mathématiques qui permettent de produire cette décomposition. Elles génèrent un « spectre de raies », c'est-à-dire un tableau comprenant, pour chaque bande de fréquence couvrant la plage des fréquences audibles, des valeurs d'amplitude et de phase associées. On passe ainsi d'une représentation temporelle du son (amplitude du signal en fonction du temps) à une représentation fréquentielle (amplitude du signal en fonction de la fréquence). Cette décomposition est similaire à celle que l'on obtient avec un prisme qui décompose la lumière, et elle est réversible, sans perte : en additionnant un ensemble de signaux sinusoïdaux (sons purs) dont les fréquences, amplitudes et phases correspondent aux valeurs de chaque raie du spectre, il est possible de reconstituer à l'identique le son original. Les transformées de Fourier sont le point de départ de nombreuses techniques actuelles d'analyse et de transformation des sons.

Ces décompositions du son établissent une correspondance entre le domaine temporel de l'onde sonore et le domaine fréquentiel dans lequel le spectre affiche les différents partiels qui composent un son. Cette correspondance est essentielle, car l'oreille humaine entend les sons dans le domaine fréquentiel, alors que ceux-ci sont stockés en mémoire sur l'ordinateur dans le domaine temporel.

Le spectre de fréquence (spectre de raies) est un tableau ou un graphique obtenu en ayant réalisé une FFT [9] sur une petite portion d'un son numérique (voir fig. 1). L'amplitude peut être exprimée sur une échelle linéaire, proportionnelle aux variations de la pression de l'air, ou sur une échelle logarithmique, en décibels.

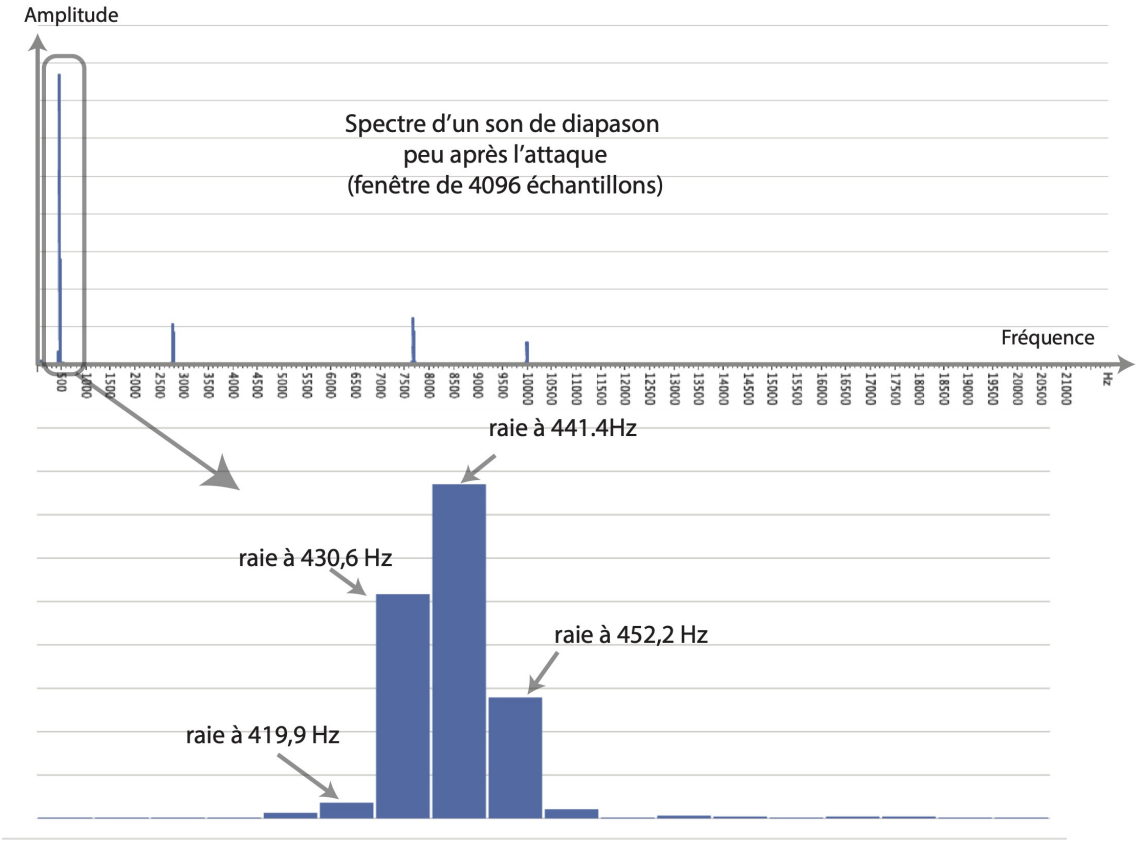
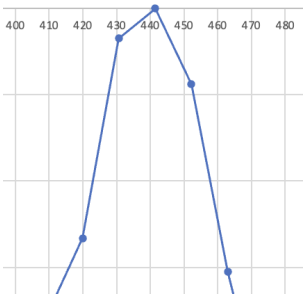


Figure 1. Spectre de raies obtenu par FFT d'un son de diapason (fenêtre de 4096 échantillons).

L'observation du spectre permet de repérer les pics, les raies, dont les amplitudes sont supérieures à celle de leurs voisines (quatre pics détectés dans le spectre du diapason). La fréquence d'un pic est ici donnée avec une précision à environ 10 Hz près. Toutefois, une estimation par calcul de la courbe idéale qui passerait par les valeurs des amplitudes des raies voisines est souvent proposée par les logiciels d'analyse pour obtenir une estimation plus précise de cette fréquence (voir fig. 2).



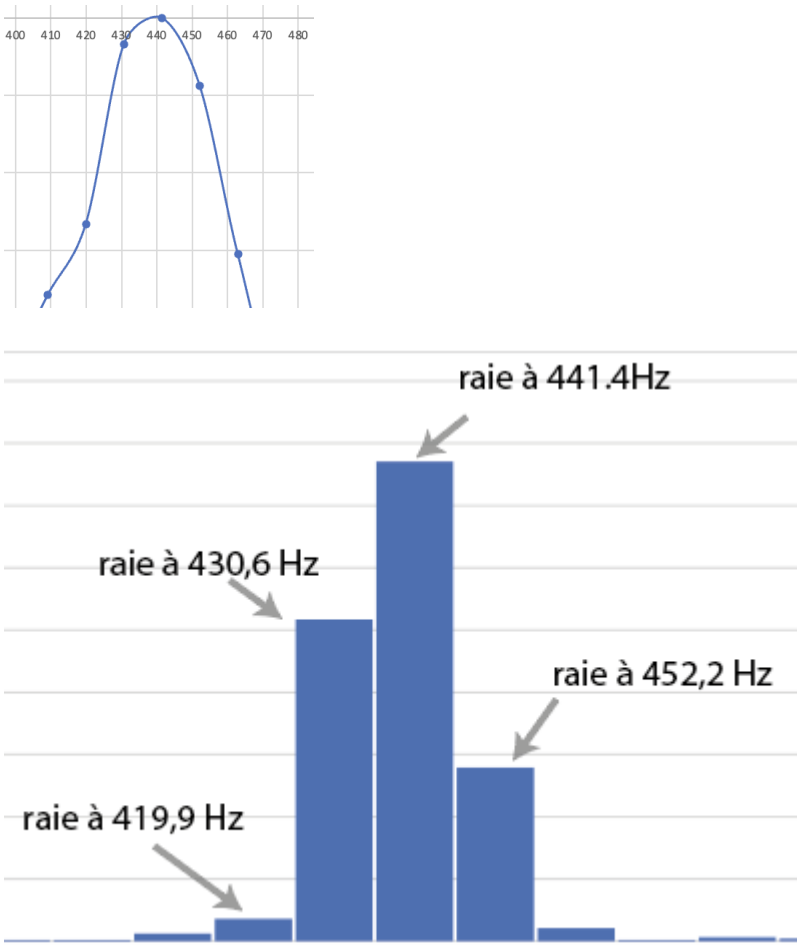


Figure 2. À gauche et au centre : quelques raies entourant un pic, valeur du pic : 441,4 Hz ; à droite : valeur estimée du pic : 440 Hz.

S'il existe une valeur dont toutes les fréquences des pics observés soient multiples, on a affaire à un son périodique (aussi appelé « son harmonique »). Cette valeur théorique est la fréquence fondamentale (F0 exprimée en hertz) du son et correspond à la hauteur de la note qui sera entendue et notable sur la partition (440 Hz par exemple pour un *la* 3, au milieu de la clé de *sol*) [10]. Les différents pics sont alors des harmoniques [11].

Le spectre donne des indications sur la nature du son à un instant donné. Il ne suffit donc pas à décrire le timbre d'un instrument. Le timbre d'un instrument dépend de la façon dont les pics observés dans le spectre varient et peuvent se déplacer au cours du temps. C'est en créant un sonagramme qu'il est possible d'obtenir une représentation qui intègre le temps. Le sonagramme est une représentation graphique issue de la superposition latérale temporelle de plusieurs analyses FFT successives (analyses de Fourier à fenêtre glissante), l'axe des abscisses représentant le temps, et l'axe des ordonnées la fréquence. L'amplitude est donnée par la couleur ou l'intensité de gris. Un sonagramme permet de visualiser les composantes partielles du son, ou « partiels ». Un partiel est une composante fréquentielle continue présente dans un sonagramme et correspond à une succession de pics de fréquences très proches présents dans les analyses spectrales successives qui constituent le sonagramme.

La figure 3 montre certaines similitudes entre une partition et le sonagramme représentant cette partition interprétée au violon. La ligne inférieure du sonagramme en escalier donne les fréquences fondamentales des notes jouées par le violon et est exactement superposable aux notes placées sur la partition. Les autres lignes qui se développent parallèlement au-dessus sont les autres partiels. Leurs fréquences sont ici toutes multiples de la fondamentale. Ce sont donc des partiels harmoniques. Les partiels suivent des trajectoires horizontales très rectilignes, ce qui indique que le son n'est pas modulé, même pour les notes un peu longues. Les lignes verticales, moins précises, font ressortir les transitoires, parties bruitées très courtes qui surviennent lors du passage d'une note à la suivante.

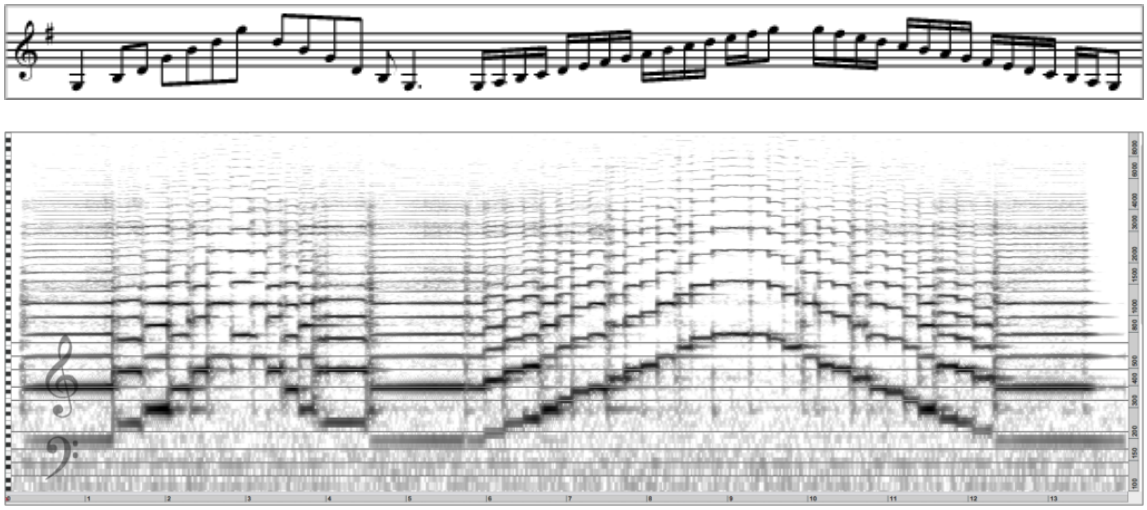


Figure 3. En haut, une phrase de violon en notation traditionnelle ; en bas, le sonagramme de la même phrase de violon (durée : 13 secondes, logiciel Audiosculpt, fenêtres d'analyse de 4 096 échantillons).

Les travaux d'Émile Leipp (1913-1986), qui a fondé en 1963 le laboratoire d'acoustique musicale dans le département de mécanique de la faculté des sciences de Paris, ont permis de décrire des sons grâce aux observations réalisées sur des sonagrammes [12] . Ces représentations donnent des informations visuelles sur le timbre du son d'une manière assez proche de la perception humaine. Toutefois, le sonagramme n'affiche pas des informations quantitatives, et les FFT qui servent à produire cette représentation contiennent beaucoup trop de données numériques pour pouvoir être exploitées telles quelles pour expliquer de quoi est fait le timbre.

> La déception des compositeurs qui ont tenté de créer des timbres par synthèse sonore n'a d'égal que le désarroi des acousticiens qui, au terme de soixante années de recherches en psychoacoustique, font état du faible avancement de leurs connaissances et constatent que le timbre reste un « mystère », un attribut auditif mal compris. Pour eux, le concept de timbre est mal défini. Il est vrai que la notion de timbre est riche et complexe et que le terme utilisé est manifestement porteur de sens multiples [13] .

Comme l'indique Michèle Castellengo, qui a dirigé le Laboratoire d'acoustique musicale (LAM) d'Émile Leipp à partir de 1982 et créé la classe d'acoustique musicale au Conservatoire national supérieur de musique et de danse de Paris en 1989, le timbre est

difficile à définir et à quantifier.

Les travaux et expérimentations de Jean-Claude Risset (1938-2016) ont ouvert de nouvelles voies dans la recherche sur le timbre dès les années 1960. Il a entrepris de produire par synthèse sur ordinateur les sons des instruments de l'orchestre afin de comparer la qualité des sons d'origine et leur resynthèse. Si le son produit par synthèse est similaire à l'original, cela signifie que le son d'origine a été décrit avec justesse [14]. Ces travaux ont été poursuivis, notamment à l'Ircam, par ceux de Xavier Rodet sur la synthèse de la voix chantée [15] et de Jean-Baptiste Barrière sur celle des sons résonants [16]. Ces différents travaux ont débouché sur la production d'outils pour analyser et synthétiser des sons sur ordinateur, que de nombreux compositeurs ont utilisés – et continuent d'utiliser – depuis. Une des applications célèbres des outils sur la synthèse de la voix a été la reconstitution de la voix d'un castrat pour le film *Farinelli* réalisé par Gérard Corbiau en 1994 [17].

L'intelligence artificielle et les recommandations en ligne

Dans les années 2000, le développement de l'intelligence artificielle (IA) et des descripteurs audio a permis des avancées importantes dans le domaine de l'étude du timbre. Depuis quelques années, les techniques de l'IA peuvent être utilisées pour la reconnaissance d'images ou la reconnaissance et la synthèse de la voix parlée, mais aussi pour l'élaboration d'applications à des fins d'expérimentation créative (peinture, vidéo, musique), donnant alors cours à des exploitations marchandes par les plateformes y ayant recours, notamment pour leur recommandation algorithmique de musique à leurs utilisateurs [18].

L'intelligence artificielle a aussi été associée à des activités de recherche portant sur l'analyse du son, sur l'étude de larges bases de données musicale et sur tout un ensemble d'informations collectées sur le web. Elle permet d'associer des morceaux de musique à des descriptions textuelles, ainsi que de construire des relations concernant les choix des internautes sur leurs consultations de musiques sur Internet.

> [Les systèmes de recommandation mis en place pour diffuser du contenu sur Internet par ces entreprises fonctionnent] généralement par une optimisation d'une mesure de précision de l'adéquation entre un utilisateur et un produit. Cependant, plusieurs travaux de recherche ont montré que l'optimisation de la précision ne produisait pas les recommandations les plus utiles pour les utilisateurs. Un système trop précis peut contribuer à confiner les utilisateurs dans leur propre bulle de choix [19].

Les systèmes de recommandation sur Internet réalisent des indexations de la musique à partir de différents types d'informations prélevées, soit sur le signal audio lui-même, comme la structure rythmique, la mélodie et les caractéristiques du timbre, soit sur des représentations symboliques, généralement issues de pages web, comme des informations sur le contexte culturel ou politique d'un artiste ou des descriptions textuelles de la musique.

Dans l'offre des outils de recommandation disponibles en 2017, nous avons testé les technologies développées par l'entreprise Niland ayant produit un programme disponible sur Internet qui effectuait un calcul de distance entre le morceau de musique qui lui était soumis et l'ensemble des morceaux présents dans la base de données de la plateforme de streaming équitable 1DTouch [20]. Plusieurs facteurs étaient pris en compte pour ce calcul de distance : le genre (plus d'une vingtaine de genres répertoriés dont un pour la musique classique et quatre pour la musique électro…), les instruments présents (batterie, guitare, percussions, cuivres, cordes et vents), l'humeur (agressive, dramatique, énergétique, dansante, joyeuse, etc.), la voix (femme, homme, chœur ou instrumental) et le tempo.

morceau	1er choix	proximité	2e choix	proximité		proximité		proximité		proximité
Hount Do g - Big Mama Thornton (1952)	*Hound Dog* Billy Boy Arnold	**95%**	*Hound Dog* Elvis Presley	**92%**	*Hound Dog* Buddy Johnson	**90%**	*Tara* Amadou Ndiaye Samb	73%	*Sugar Mama* John Le Hooker	72%
La Crise - Blessed Virgin (groupe stéphanois enregistrement live non commercialisé) (1982)	*The Ganjas* The Cure	69%								
Bourée - Jethro Tull - rock progressif	*On the Green Light* The Spencer Davis Group	71%								
L'Oeil écoute - Bernard Parmegiani - musique électroacoustique	Batteries du Premier Empire Les Equipages de la Flotte de Toulon	77%	Vagues Musique Relaxante Relax	76%						
Paesana-Palabre - Sixun - Jazz Fusion	*SFING* Alain Caron	75%	*Crosstown Traffic* Groove Catchers	75%						
Hot Fun - Stanley Clarke - Jazz-rock	*Overdrive* Widemann, Vander, Top, Lockwood	72%	*Coriza* Dos Cafundós	70%						
Timewind - Klaus Schulze - rock planant	Tamas 432HZ 432Hz Yoga	77%	*La chasse gallery* Malicorne	76%						
Kassandra - François Bernard Mâche - musique contemporaine mixte	*Geneviève de Brabant, Scene 1* : Présentation Orchestre Lyrique de la RDF	73%	*Le voyage dans la Lune: Récit VI* Orchestre Jean-Paul Kreder	72%						
Autobahn - Kraftwerk - musique électronique	Intro Mike-L	68%	Chimère Maelstrom	64%	*La jambe de bois* Serge Gainsbourg	64%				

Tableau 1. Test du logiciel Niland pour le calcul de similarités entre des morceaux choisis (colonne de gauche) et ceux de la base de données 1DTouch.

Les tests que nous avons réalisés ont donné des résultats très convaincants (voir p. 150, tableau 1) sur des œuvres musicales célèbres (« Hound Dog »), assez bons sur des musiques inconnues des bases de données (Blessed Virgin), mais situées dans un registre rock très marqué, assez mauvais pour des musiques de rock progressif (Jethro Tull) et totalement déplacés pour des musiques électroacoustiques (Bernard Parmegiani), voire des musiques de groupes de musique électronique connus (Kraftwerk).

C'est pourquoi, en 2017, après avoir échangé avec Pierre-René Lhérisson et le responsable de la société 1DTouch, nous avons eu le projet de développer des outils pour essayer de diversifier les offres de streaming de la plateforme, en faisant intervenir l'analyse audio des musiques, à travers la notion de signature sonore qui est totalement absente de ces outils de recommandation, pour lesquels les critères de reconnaissance et de calcul de distance fonctionnent bien quand certains instruments phares et le style de musique sont reconnus. La qualité de la voix et le tempo se révèlent être des éléments importants dans le calcul, mais, pris globalement, le timbre ne joue aucun rôle significatif dans ces outils.

Les descripteurs audio

Les descripteurs audio sont des paramètres issus d'algorithmes utilisés depuis ces dernières décennies pour décrire le timbre, mais chacun d'entre eux n'en fait ressortir qu'un aspect très réducteur. C'est la combinaison d'un assez grand nombre de ces descripteurs qui doit permettre de représenter les qualités acoustiques d'un son que l'oreille humaine perçoit.

Souvent, ces descripteurs sont issus de formules mathématiques assez simples, il est donc nécessaire de les retravailler afin de les rendre mieux adaptés à la perception.

Les premières publications significatives concernant ces descripteurs ont été réalisées par des chercheurs tels John Grey [21] ou David Wessel [22] dans les années 1970. Elles portaient sur l'étude du timbre de sons isolés d'instruments de musique de l'orchestre occidental. Les descripteurs employés étaient généralement la distribution de l'énergie dans le spectre (représentée par le barycentre de l'énergie selon les partiels du son), les caractéristiques de l'attaque du son (liée à la durée des transitoires ou la durée de l'attaque) et le flux spectral (correspondant aux variations de l'enveloppe spectrale au cours du temps).

Depuis une vingtaine d'années, de nombreux descripteurs audio ont été mis au point [23] et sont maintenant disponibles dans différents environnements de programmation comme Matlab, Sonic Visualizer ou MaxMSP.

L'Ircam a développé un programme très sophistiqué intitulé Orchidea [24] (orchestration assistée par ordinateur) et qui utilise les descripteurs pour prédire le timbre produit par différentes associations possibles d'instruments de l'orchestre. L'utilisateur peut proposer le résultat sonore qu'il souhaite atteindre en fournissant un fichier audio de référence, et l'ordinateur lui offre ensuite une partition d'orchestre complète en fonction des instruments qui ont été choisis pour la composition.

> Réalisé à la suite du logiciel « Orchidée », « Orchids » est le premier système complet pour l'orchestration temporelle assistée par ordinateur et l'optimisation de mélanges de timbres. Il fournit un ensemble d'algorithmes permettant de reconstruire n'importe quelle cible sonore évoluant dans le temps par une combinaison d'instruments ou échantillons, selon un ensemble de critères psychoacoustiques. Il peut aider les compositeurs à obtenir des couleurs de timbre inouïes en fournissant une multitude de solutions efficaces qui recréent au mieux cette cible sonore. […] Ce système fournit plusieurs algorithmes d'approximation permettant d'optimiser conjointement plusieurs propriétés de timbre. Les avantages du système « Orchids » résident dans le fait que cette approximation peut être faite séparément sur des formes temporelles, valeurs moyennes ou écarts-types (ou toute combinaison des trois) de chaque descripteur psychoacoustique. En outre, les utilisateurs peuvent également définir une déformation temporelle manuelle, et même effectuer une recherche multicible à l'intérieur de multiples segments sonores, offrant ainsi des réalisations de pièces orchestrales complètes en quelques secondes [25].

Concernant notre projet, nous avons réalisé un programme dans MaxMSP utilisant la fonction « ircamdescriptor~ », développée à l'Ircam pour le calcul d'une cinquantaine de descripteurs, pour analyser plusieurs morceaux de musique. Les descripteurs que nous avons sélectionnés dans ce programme sont les suivants : *Loudness, SpFlatness, SpCrest, PercSpCentroid, PercSpSpread, PercSpkewness, PercSpKurtosis, PercSpRolloff, PercSpVariation, PercSpDecrease*. Selon la nature du descripteur, les calculs sont réalisés soit directement sur le signal audio, soit sur une analyse spectrale du son (FFT), ou encore sur des analyses spectrales successives au cours du temps. Les analyses spectrales peuvent être

prises dans leur ensemble, mais leurs raies peuvent aussi être regroupées sur des intervalles d'égale largeur sur une échelle logarithmique (on utilise par exemple des bandes de fréquences divisées en tiers d'octave). Certains descripteurs sont calculés à partir d'une extraction des harmoniques qui composent le son (pour des sons monophoniques périodiques). Les descripteurs dits « perceptuels » (*perceptual*) sont calculés sur des FFT ayant subi un filtrage produisant des pondérations liées à la sensibilité de l'oreille humaine, après un regroupement des raies de ces FFT sur des échelles Bark [26] ou Mel [27].

Voici une rapide présentation de ces descripteurs :

- *Loudness* (*perceptual*) : somme des énergies (amplitudes au carré) observées dans chaque bande (Bark) de fréquences ;
- *SpFlatness* : rapport entre la moyenne géométrique (racine n^e du produit des n amplitudes de chaque raie de la FFT) et la moyenne arithmétique des amplitudes de chaque raie (pour un bruit blanc, ce paramètre est proche de 1 [100 %] ; pour un son pur ou un son harmonique, il est proche de zéro) ;
- *SpCrest* : rapport entre l'amplitude de la raie la plus forte et la moyenne des amplitudes de l'ensemble des raies (pour un bruit blanc, la valeur est proche de 1) ;
- *PercSpCentroid* : moyenne arithmétique des fréquences pondérées par les amplitudes correspondantes (de nombreux auteurs associent ce paramètre [exprimé en Hz] à la brillance du son, un centroïde élevé indique de l'énergie dans les fréquences medium et aiguës) ;
- *PercSpSpread* : écart-type (en hertz) à la moyenne (donc au centroïde) (ce paramètre mesure l'étalement du spectre) ;
- *PercSpSkewness* : mesure le degré d'asymétrie du spectre (une valeur positive indique plus d'énergie dans les graves et vice-versa) ;
- *PercSpKurtosis* : mesure la finesse ou la dispersion du spectre (une valeur élevée indique un spectre resserré) ;
- *PercSpRolloff* : indique la fréquence (en hertz) en dessous de laquelle est concentrée 95 % de l'énergie du spectre ;
- *PercSpVariation* : représente la quantité de variation (en pourcentage) observée entre une analyse FFT et la suivante ;
- *PercSpDecrease* : mesure la décroissance du son au cours du temps et permet de différencier les sons entretenus et les sons percussifs.

De nombreuses références à ces outils sont disponibles dans les publications rédigées par des chercheurs travaillant sur l'analyse musicologique des œuvres et s'intéressant à la description du timbre. Lors du plus grand congrès mondial qui s'est tenu à ce jour sur la question du timbre, « Timbre 2018 : Timbre Is a ManSplendored Thing » à l'université McGill de Montréal au Québec [28], le terme *centroïde* (le plus répandu des descripteurs) a été cité dans plus d'une dizaine de communications pour comparer des échantillons audio.

Les descripteurs sont utiles pour faire ressortir certains traits saillants concernant des sons isolés. Ils permettent de placer des sons instrumentaux dans un espace de timbres comme l'ont fait Grey, Wessel ou Krumhansl [29], ce qui peut aider un compositeur à choisir les

instruments qu'il utilisera dans une composition. En revanche, pour des sons complexes ou des extraits d'œuvres musicales, ces descripteurs nécessitent des aménagements.

L'application de plusieurs descripteurs sur une œuvre très contrastée comme « Pop Plinn [30] » (1971) d'Alan Stivell, une des premières rencontres entre musique celtique et musique pop-rock, peine à mettre en évidence les contrastes de timbres (guitare électrique, harpe celtique, bombarde) qui sont pourtant très clairs à l'écoute.

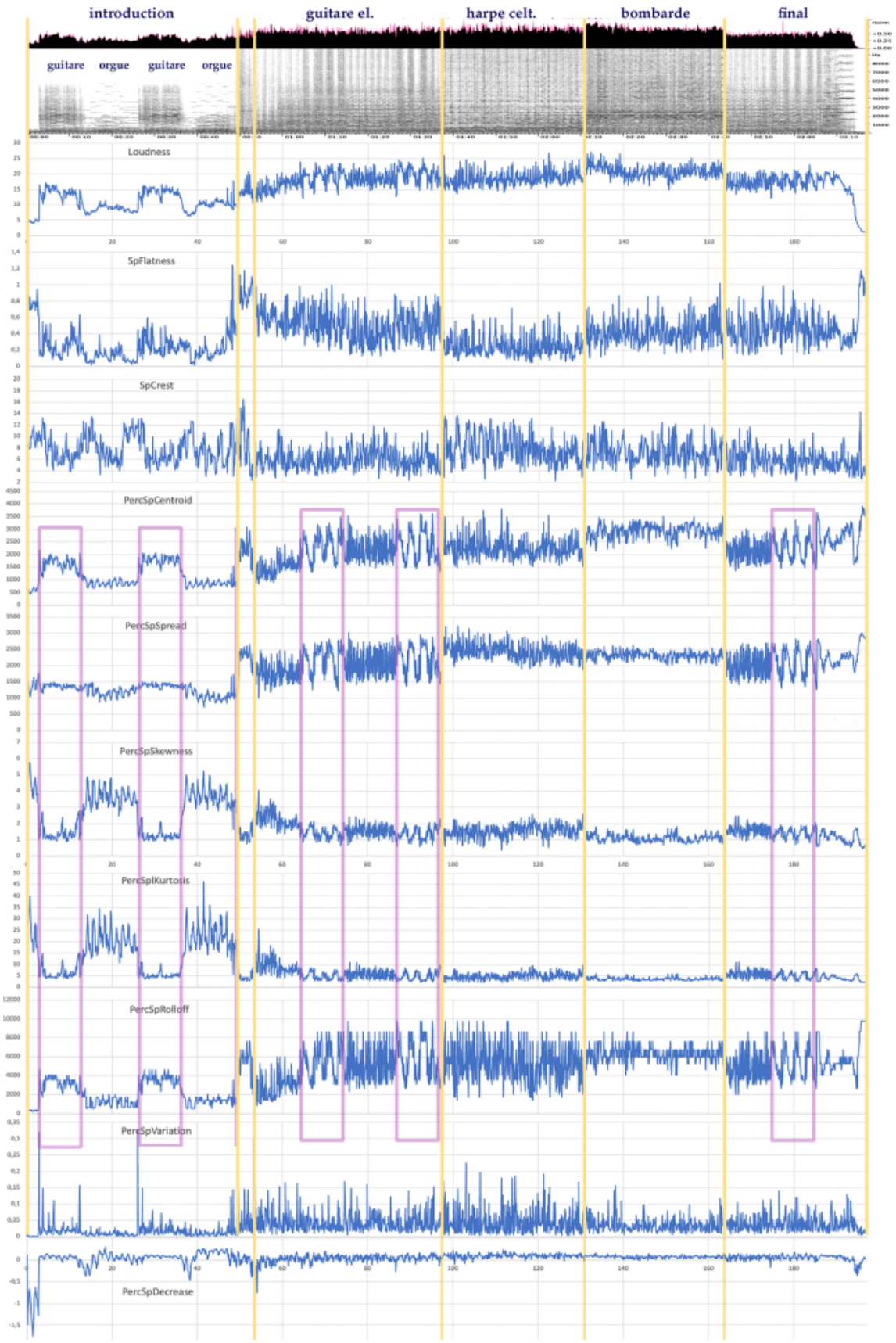


Figure 4. Graphes de plusieurs descripteurs audios calculés sur « Pop Plinn » (l'axe horizontal représente le temps en secondes).

Sur la figure 4, on constate que plusieurs descripteurs sont redondants (*PercSpCentroid, PercSpSpread, PercSpSkewness, PercSpKurtosis, PercSplRolloff*). Ils font principalement ressortir l'alternance guitare/orgue dans les cinquante premières secondes (introduction) et mettent en évidence l'opposition refrain/couplet dans le solo de guitare (55 s-97 s) et dans le final (162 s-185 s). Les autres paramètres ne donnent pas vraiment d'informations significatives en rapport avec la perception, alors qu'à l'écoute, les sons de guitare électrique, de harpe celtique et de bombarde sonnent de façons très différentes. On constate donc que malgré l'utilisation d'un nombre important de descripteurs, les graphes ne permettent pas de faire ressortir les variations de timbre qui sont assez claires à l'audition. C'est pourquoi nous avons voulu mettre au point de nouveaux outils plus adaptés à ce qu'un auditeur peut percevoir d'un extrait musical.

Comme nous l'avons indiqué, les descripteurs audio sont plutôt conçus pour analyser des sons d'instruments de musique. Ils donnent des informations assez floues sur des polyphonies instrumentales, mélanges complexes de timbres.

Les signatures sonores

Dans le cadre de notre projet, nous travaillons sur des échantillons sonores prélevés dans des œuvres musicales ou sur des sons d'origines diverses, de deux secondes environ. À cette échelle temporelle, il est possible de reconnaître immédiatement à l'écoute, pour des musiques qui nous sont familières, le nom du compositeur, de l'œuvre ou du groupe qui les interprète. Nous utilisons alors la notion de « signature sonore » pour indiquer une représentation spectro-temporelle condensée du son que nous réalisons sur ordinateur [31]. À côté du terme de « signature sonore » sur lequel porte notre travail, on trouve souvent dans la littérature actuelle, principalement en IA, celui d'« empreinte sonore » (*footprint*), utilisé par des programmes comme Shazam, pour identifier automatiquement des musiques à partir d'un extrait audio. Les données numériques correspondant à ces empreintes ne sont pas accessibles, elles interviennent dans des processus automatiques d'apprentissage, à l'intérieur de « boîtes noires » dans lesquelles il est difficile de savoir ce qui se passe.

L'identification audio ou la prise d'empreintes digitales est un scénario de récupération nécessitant une grande spécificité et une faible granularité. Le but ici est de récupérer ou d'identifier le même fragment d'un enregistrement musical donné avec certaines exigences de robustesse (par exemple, bruit d'enregistrement, codage). Des approches bien connues telles que celle proposée par Wang [32] ont été intégrées dans des systèmes disponibles dans le commerce, tels que Shazam, Vericast ou Gracenote MusicID [33].

Dans notre projet, la signature est obtenue après réduction d'une succession de FFT pour en extraire une matrice triangulaire de 27 bandes de fréquences (par tiers d'octaves, de 20 Hz à 14 kHz), chaque bande k étant décrite par un ensemble de N cellules temporelles contenant l'énergie RMS de chaque cellule ($N=k\times 2+4$). Ce tableau est ensuite représenté graphiquement en deux versions : la signature (axe horizontal : le temps ; axe vertical : la fréquence en échelle logarithmique) et le spectre moyen (axe horizontal : l'énergie ; axe vertical : la fréquence [34]) (voir p. 153, fig. 5).

L'objectif est de pouvoir décrire ces signatures et de les comparer. Pour cela, nous avons expérimenté plusieurs descripteurs audio d'après les formules énoncées par Geoffroy Peeters dans ses articles de 2004 et 2011 [35] sur un ensemble de sons tests et de sons issus d'un large corpus de morceaux, principalement issus du genre pop-rock.

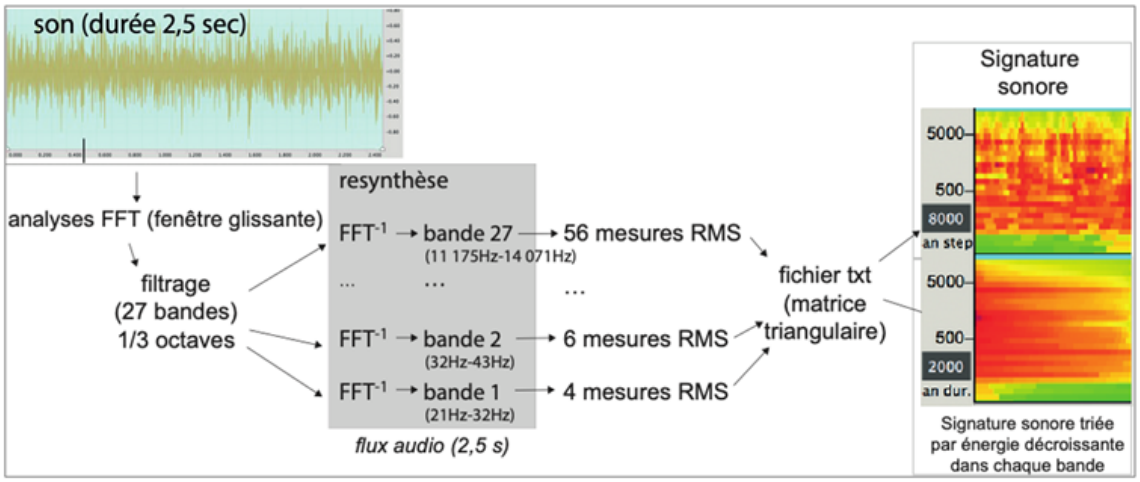


Figure 5. Les étapes pour le calcul d'une signature sonore.

Les descripteurs reconsidérés

Les descripteurs audio sont définis par des formules mathématiques qui peuvent donner des résultats assez différents selon la nature de la fonction d'analyse spectrale utilisée, du taux d'échantillonnage, etc. Nous avons réalisé des adaptations de ces formules pour les appliquer à nos signatures et les adapter à la perception, par exemple en prenant des échelles logarithmiques pour les fréquences, en accordant plus d'importance aux bandes de fréquences que favorise l'audition humaine et en choisissant des découpages du temps différents selon les bandes de fréquences utilisées.

Les descripteurs audio sur lesquels porte notre étude et qui nous ont donné le plus d'informations sur le timbre perçu sont les suivants :

- le centroïde des fréquences (calculé d'après le logarithme des fréquences, une échelle proportionnelle à celle des notes de musique), que l'on pourrait qualifier de centroïde MIDI (*mc_centr*), en référence aux numéros de notes codées par le système MIDI ;
- les écarts moyens des fréquences du spectre par rapport au centroïde MIDI, vers les graves (*mc_ec_low*) et vers les aigus (*mc_ec_high*). Ils apportent des précisions sur l'apparence générale du spectre ;
- les descripteurs *skewness, kurtosis, rolloff, flatness* et *crest* auxquels nous avons ajouté le paramètre *maxfreq25*, *midi-sp-slope* et *flux_tps*. Le paramètre *maxfreq25* indique l'amplitude maximale observée dans la bande n° 25 (située vers 8 kHz) et représente donc la quantité d'énergie présente dans les très hautes fréquences, *midi-sp-slope* calcule la pente moyenne du spectre, et *flux_tps* indique les variations (en pourcentage) des amplitudes au cours du temps dans les différentes bandes.

3. Exemples d'application des outils

Nous présentons tout d'abord quelques signatures de sons témoins caractéristiques : un son pur (500 Hz), une suite d'impulsions (2 Hz), un bruit blanc et un bruit rose.

Pour chaque signature nous utilisons deux représentations graphiques : la première est la signature elle-même ; la seconde est une représentation spectrale moyenne de la signature : pour chaque bande de fréquences (horizontale) les valeurs d'amplitudes sont triées de la gauche vers la droite par amplitude décroissante. L'axe horizontal ne représente donc plus le temps (voir tableau 2, fig. 7).

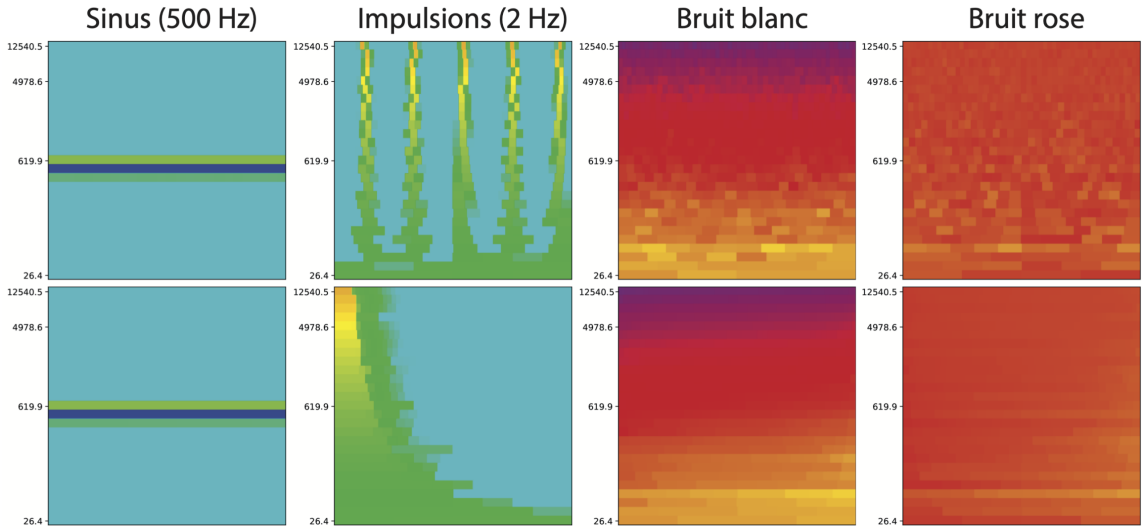


Figure 6. Les signatures de quatre sons témoins (durées 2,5s) et échelle des couleurs à droite.

	mc_ec_low (MIDI)	mc_centr (MIDI)	mc_ec_high (MIDI)	midi_sp_slope	skewness	kurtosis	sp_rolloff (Hz)	sp_flatness (%)	sp_crest	maxfreq25	flux_tps	ecartTpsAmps
Sinus (500 Hz)	71	71	75	-0.0001	17.29	1360.3	492	0%	26.99	0.00	0%	0%
Impulsions (2 Hz)	59	94	116	0.0000	0.77	2.3	9957	82%	27.18	0.01	10%	13%
Bruit blanc	69	101	119	0.0017	0.53	2.0	9957	66%	3.51	0.21	1%	1%
Bruit rose	45	77	107	0.0001	1.71	5.0	6273	99%	1.42	0.03	3%	2%

Tableau 2. Les valeurs des descripteurs sur les quatre sons témoins. En orange les valeurs élevées, en bleu les valeurs faibles.

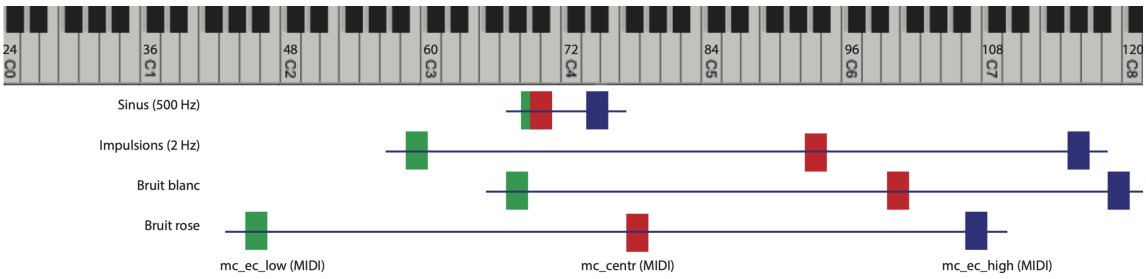


Figure 7. Le centroïde MIDI (en rouge) des sons témoins et les écarts inférieurs (en vert) et supérieurs (en bleu).

« Pop Plinn » d'Alan Stivell, une structure contrastée

Les valeurs des descripteurs calculées sur les signatures réalisées toutes les quatre secondes dans le morceau « Pop Plinn » sont affichées dans la figure 8.



Figure 8.Quelques-uns des descripteurs audio issus des signatures réalisées sur « Pop Plinn ».

Sur cette figure, on constate que les descripteurs utilisés montrent beaucoup mieux que les descripteurs standards de la figure 4 les variations globales de timbre qui caractérisent ce morceau.

En observant le centroïde et les écarts à la moyenne, on remarque ainsi une très nette progression, régulière, vers les aigus, vers plus de brillance et d'énergie jusqu'à 2'45", retour de la guitare dans la partie finale (*coda*).

La pente spectrale (*slope*) montre une forte opposition entre le son de la guitare, riche en harmoniques, mais contenant beaucoup d'énergie dans les graves (valeurs négatives de *slope*), et le son de la bombarde, qui présente un formant [36] très marqué situé entre 1 500 et 3 000 Hz comme l'indiquent les spectres réalisés à deux endroits dans le morceau (voir fig. 9).

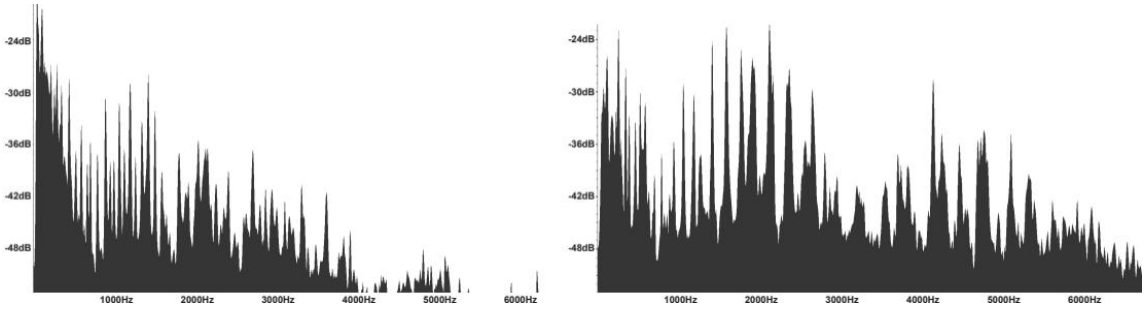


Figure 9. Analyses spectrales à 1 min 05 (solo de guitare à gauche) et à 2 min 15 (solo de bombarde à droite) – fenêtre d'analyse de 4096 échantillons.

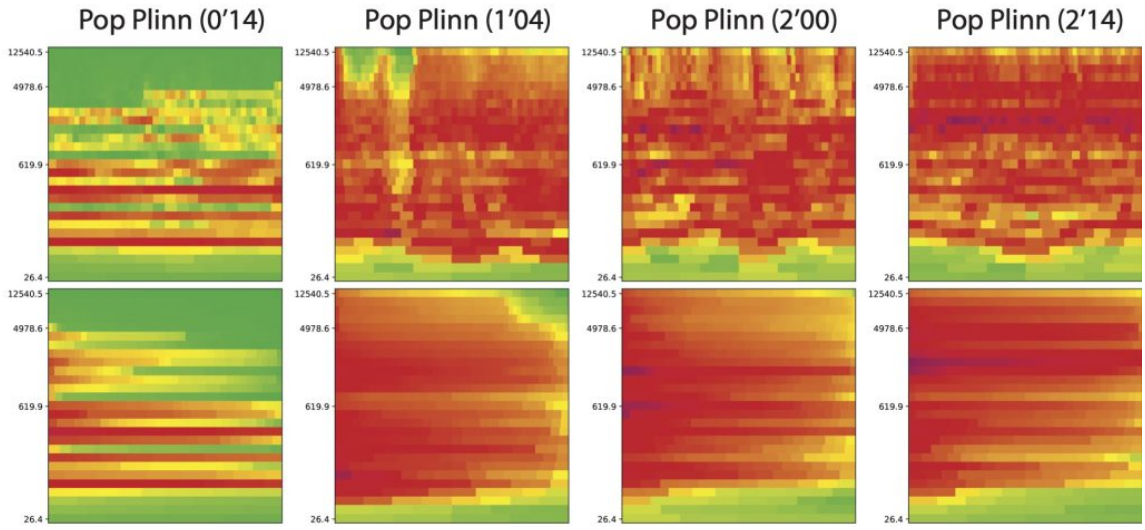


Figure 10 : quatre signatures extraites de Pop Plinn (de gauche à droite, orgue seul, chorus guitare, chorus harpe, chorus bombarde).

Les paramètres *flatness* (saturation) et *crest* (crête) mettent en évidence la richesse spectrale de la guitare qui remplit l'espace sur l'axe des hauteurs, ainsi que sur l'axe temporel. La harpe, en revanche, laisse de l'espace sur le plan rythmique, et on constate sur ces signatures que les temps marqués par la batterie sont beaucoup plus visibles dans les lignes supérieures du graphe (voir fig. 10).

Développement de la saturation dans le rock, du rythm'n'blues au black metal

Les paramètres *flatness* et *crest* mettent ainsi en évidence une tendance qui s'est développée depuis les années 1950 avec l'électrification de la guitare et la naissance du rock'n'roll, puis avec l'apparition du hard rock au début des années 1970 et le développement du rock metal ensuite : développement de la distorsion, production de sons saturés (notamment guitare et voix) et de rythmes de batterie de plus en plus rapides. Dans un texte paru en 2019 [37], nous avons montré l'évolution de la saturation dans le rock'n'roll et l'avons quantifiée avec ces deux paramètres. Nous avons également établi un classement d'une centaine de morceaux de black metal, du plus saturé au moins saturé. Le morceau le plus saturé avec une valeur de *flatness* de plus de 85 % est le morceau « Virtual War » (2009) du groupe Samuel.

Représentation multidimensionnelle des descripteurs sur Led Zeppelin

Nous nous sommes intéressés à la saturation dans la musique, en étudiant quatre morceaux [38] du groupe Led Zeppelin à l'origine du mouvement hard rock apparu au tout début des années 1970 [39]. Nous avons calculé, toutes les cinq secondes et sur toute la durée des morceaux, les signatures et les descripteurs associés. Nous avons ensuite combiné l'ensemble des descripteurs sur une seule représentation graphique à deux dimensions, en utilisant la technique d'analyse en composantes principales (ACP) [40], puis nous avons éliminé graduellement les descripteurs les plus redondants pour garder ceux qui offraient une représentation graphique en bonne adéquation avec la perception.

Le graphe (voir p. 156, fig. 12) met en évidence deux dimensions diagonales (voir fig. 11), que nous avons pu relier à une dimension son « sombre/brillant », représentée par le centroïde, la pente et l'écart moyen inférieur, et une dimension son « clair/saturé », représentée par les paramètres *crest*, *kurtosis*, *flatness* et *maxfreq25*. Des ellipses colorées ont été dessinées à la main pour mettre en évidence les regroupements des signatures de chaque morceau. Pour les quatre morceaux, on observe une ellipse oblique caractéristique située dans la partie inférieure gauche qui correspond à une forte saturation du son.

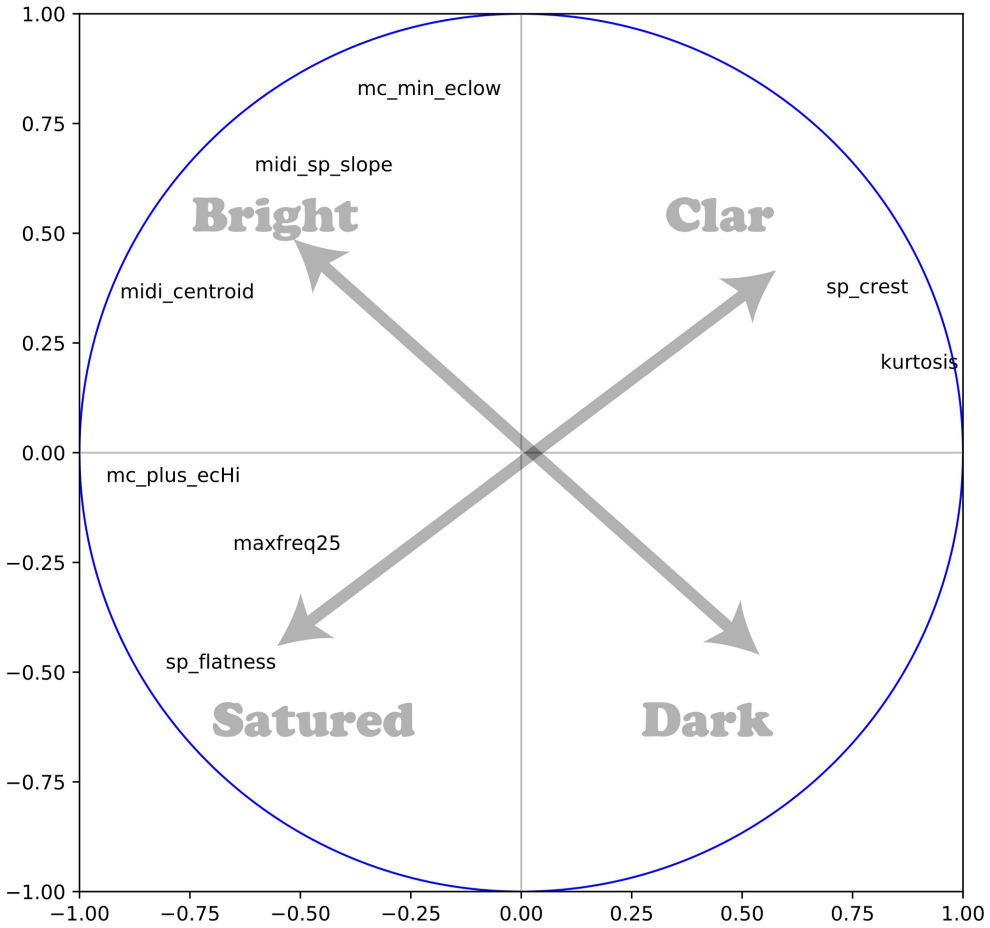


Figure 11. Disposition des descripteurs sur leurs axes d'influence dans le cadre d'une analyse ACP.

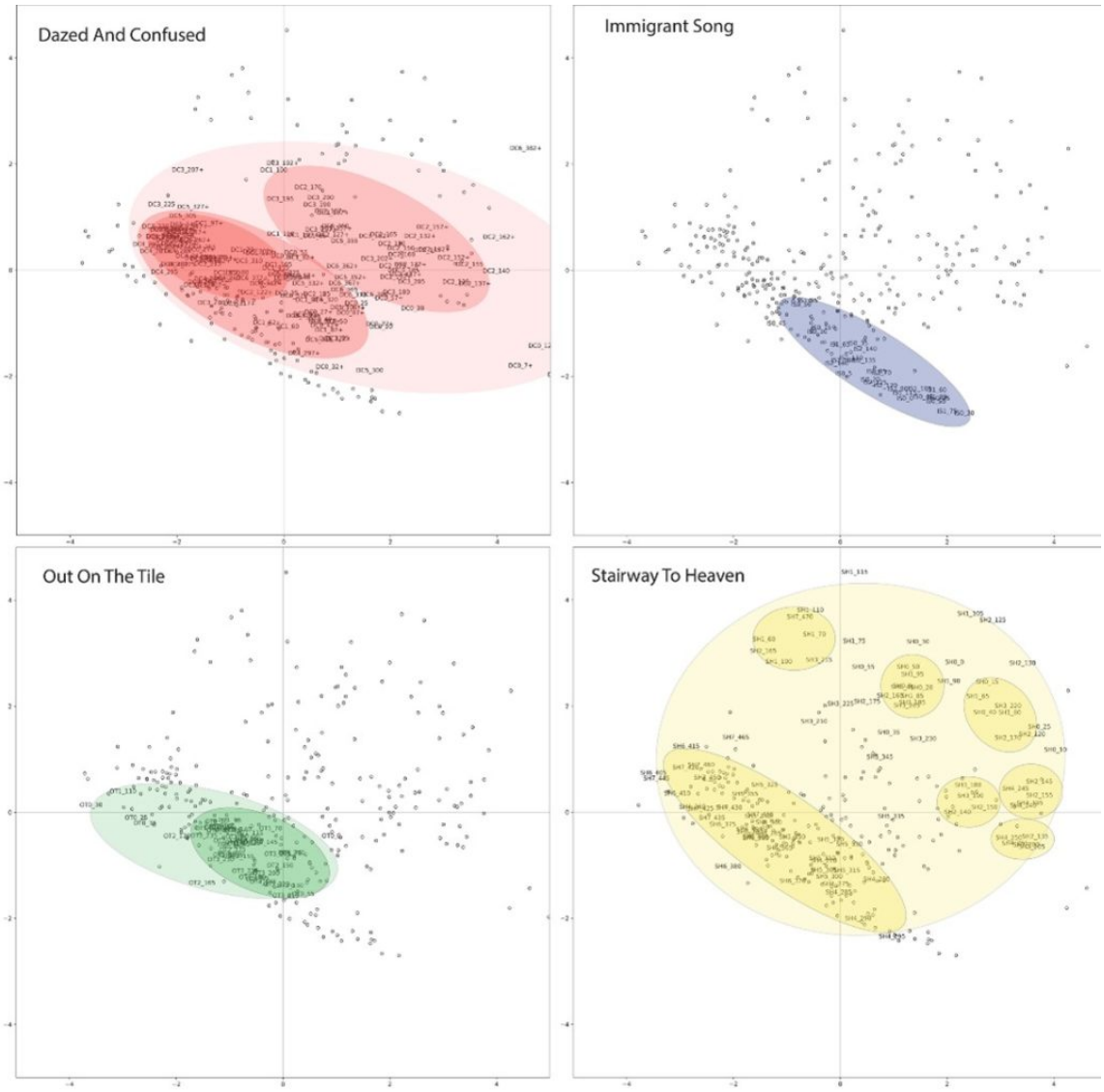


Figure 12. Positionnement des signatures sonores de quatre morceaux de Led Zeppelin après une ACP globale.

Selon les cas, les points de certains morceaux sont très rapprochés et placés dans la partie du plan dirigée vers la saturation et vers le côté sombre, comme pour « Immigrant Song ». Au contraire, pour « Stairway to Heaven », l'espace est rempli dans toutes les directions.

Nous avons ensuite ajouté à ce graphique les points issus des signatures d'autres œuvres pour voir où elles se situent par rapport à ces œuvres de référence du genre hard rock (voir fig. 13a). En ajoutant le morceau « Virtual War », le plus saturé de la liste de black metal, nous le trouvons sans surprise dans le quart inférieur gauche, bien en dessous des quatre autres morceaux. Nous avons également placé dans cet espace le morceau « Pop Plinn » (voir fig.13b). Les différentes sections de celui-ci apparaissent clairement dans des zones distinctes, mais la partie jouée par la bombarde sort de façon très nette de l'espace de représentation initial (en haut à gauche), témoignant d'un timbre tout à fait inhabituel dans les musiques populaires actuelles.

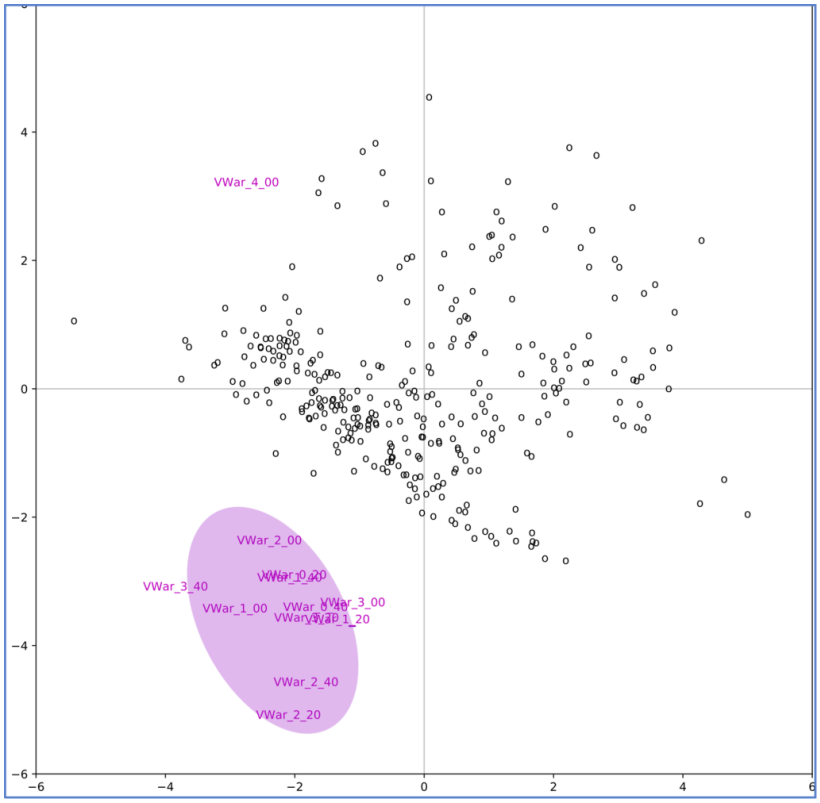

VWar_4_00
VWar_2_00
VWar_3_40
VWar_1_00
VWar_3_00
VWar_2_40
VWar_2_20



Figure 13. Positionnement des signatures de (a) Virtual War (à gauche) et (b) Pop Plinn (à droite) dans l'ACP de Led Zeppelin.

4. Perspectives

Comme nous l'avons vu, les outils mathématiques et informatiques permettent de mesurer les paramètres du son qui ont des liens avec la perception lors de l'écoute. Ils peuvent trouver de nombreux champs d'application en musicologie, mais ils ne sont actuellement pas suffisamment adaptés pour offrir une représentation intéressante d'un timbre global d'échantillons sonores prélevés dans des morceaux de musique à l'échelle d'environ deux secondes de signal audio. C'est pourquoi nous avons proposé de calculer des « signatures sonores » qui contiennent des informations condensées, tout en étant assez riches pour permettre le calcul de descripteurs audio adaptés qui donnent des informations plus pertinentes vis-à-vis de la perception. Nous avons notamment évoqué dans ce texte l'utilisation de deux paramètres (*flatness* et *crest*) qui mettent très bien en évidence la notion de la saturation en musique.

En ce qui concerne la classification d'œuvres musicales ou les calculs de similarité et de distance entre des œuvres, les analyses en composante principale (ACP) présentent des fonctionnalités très intéressantes. Elles permettent de positionner graphiquement dans un espace à deux dimensions les différentes signatures sonores d'une œuvre musicale, prélevées par exemple à des intervalles de quatre à dix secondes sur l'ensemble d'un morceau. Elles offrent la possibilité d'envisager des segmentations pour faire ressortir la structure de l'œuvre en fonction des variations globales du timbre. Lorsque l'on effectue une ACP à partir des descripteurs d'une œuvre, ou d'un corpus d'œuvres, cela produit une représentation de l'ensemble des données sur un plan (espace bidimensionnel), tout en conservant le maximum d'information. En fonction des critères perceptifs que l'on cherche à étudier, on choisit ainsi une liste de descripteurs particuliers que l'on souhaite mettre en avant. Cette répartition des différentes dimensions dans un espace 2D permet ensuite d'ajouter de nouvelles œuvres pour les comparer au corpus initialement utilisé, afin de faire ressortir des similarités ou des différences selon ces critères. Sur l'ACP réalisée sur les quatre œuvres de Led Zeppelin, nous n'avons présenté que deux exemples d'œuvres à comparer, mais nous avons néanmoins effectué des comparaisons sur une dizaine de musiques de styles très variés. L'utilisation de la bombarde dans « Pop Plinn » d'Alan Stivell a montré que cette section du morceau sortait des limites affichées du graphique et se trouvait à une très grande distance de toutes les œuvres que nous avions préalablement étudiées.

Actuellement, l'un de nos objectifs est de prolonger ce travail en effectuant des analyses sur des corpus de plus grande dimension. Dans cette optique, nous collaborons avec les chercheurs du projet Wasabi [41] qui ont réalisé un graphe de connaissances décrivant la discographie de plus de 77 000 artistes, représentant plus de 200 000 albums et deux millions de chansons. Ces chercheurs ont conçu une plateforme [42] qui récupère sur Internet les métadonnées de n'importe quel artiste, analyse les résultats et les affiche grâce à différentes techniques de visualisation. Des techniques de miniatures audio [43] sont utilisées pour résumer un album, grâce à un lecteur web audio intégré dans la visualisation. Il est ainsi possible pour l'internaute d'avoir à sa disposition toute la discographie d'un artiste dont il peut très facilement parcourir des extraits.

Pour aider l'internaute à explorer des catalogues musicaux, nous avons travaillé avec Guillaume Pellerin, chercheur à l'Ircam, afin de rendre possible le calcul et l'affichage en temps réel dans le navigateur de différents descripteurs audio. Une première version de cet outil [44] a permis de produire quelques représentations graphiques synchronisées avec le lecteur audio pour des chansons du groupe Queen : le centroïde spectral ; l'aplatissement spectral ; la pente spectrale ; les attaques ; le spectrogramme ; l'enveloppe d'amplitude colorisée selon les valeurs du centroïde spectral. Le dispositif permet également à l'internaute de réaliser des annotations sur des segments sélectionnés sur l'une des courbes affichées. Ces types de visualisation peuvent être utiles pour la recherche d'artistes, l'identification des périodes importantes de leur production ou l'exploration de leur discographie. Ils permettent également de faire des recherches en privilégiant certaines caractéristiques acoustiques des musiques que l'on souhaite étudier, en combinant plusieurs descripteurs dans une même représentation graphique pour rendre l'affichage plus

explicite, comme dans la figure suivante (fig. 14).

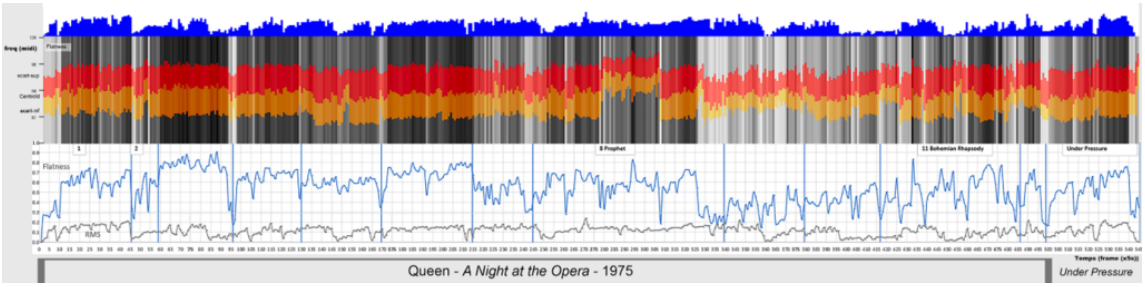


Figure 14. Représentations graphiques multidimensionnelles de certaines propriétés acoustiques des enregistrements A Night at the Opera et Under Pressure du groupe Queen : en haut et en bleu : RMS ; au milieu : saturation (en niveau de gris), centroïde entouré d'écarts-types du spectre (en jaune et rouge) ; les deux courbes inférieures donnent respectivement la saturation et le niveau RMS.

Il y a là un travail de recherche passionnant à réaliser pour effectuer des rapprochements entre des œuvres pouvant provenir de styles de musique très différents et qui présentent des similarités globales de timbre. Nous projetons d'enrichir la base Wasabi avec des signatures portant sur l'ensemble des morceaux de la base, ce qui permettra ensuite à l'internaute de se déplacer dans ces représentations graphiques multidimensionnelles pour rechercher des œuvres dont la structure en matière d'évolution timbrale l'intéresse.

Conclusion

Nous envisageons de continuer à explorer les possibilités offertes par ces outils : d'une part, pour des travaux d'analyse en musicologie et, d'autre part, pour participer à l'amélioration et à la diversification des programmes effectuant des recommandations musicales sur Internet. Notre conviction est que ces outils, à la différence de la plupart des dispositifs d'intelligence artificielle qui cherchent à proposer les morceaux correspondant le mieux aux goûts des utilisateurs, mais qui reposent généralement sur des écoutes statistiquement récurrentes, devraient pouvoir être programmables, de façon à ce que l'utilisateur puisse lui-même choisir les critères qu'il souhaite mettre en avant dans ses choix de recommandation. Il nous faut maintenant mettre au point des outils programmables, ergonomiques, et qui donnent la possibilité d'explorer l'univers de la musique sur Internet en prenant en compte, comme un critère parmi d'autres, des caractéristiques de timbres dans la musique. Chaque morceau peut être représenté par une ou plusieurs signatures sonores selon sa complexité, chaque signature étant associée à un ensemble de mesures numériques liées à différents descripteurs perceptifs optimisés.

Les recherches que nous menons et que mènent les psychoacousticiens permettent d'affiner les connaissances dans le domaine de l'étude des œuvres musicales, dans le but de mieux comprendre ce qui fait les spécificités des musiques d'aujourd'hui et de demain. Les technologies que nous développons peuvent aussi être employées pour aider à la création musicale, le compositeur étant ainsi à même de mesurer de façon inédite les caractéristiques

auditives de ce qu'il élabore durant le processus de création [45] .

1. De façon générale, le timbre est défini comme la qualité du son spécifique d'un instrument de musique, indépendante de sa hauteur et de son intensité. Le timbre est lié à de nombreux paramètres, dont le spectre des fréquences, les transitoires et bien d'autres encore.
2. « Pour Debussy, et plus encore pour Stravinsky, la musique était conçue dès le départ avec sa couleur instrumentale », Delalande F., Le son des musiques, Paris, Buchet-Chastel, 2001, p. 28.
3. Technique élaborée par Henry Cowell (1897-1965) et reprise par John Cage (1912-1992) qui consiste à altérer le son d'un piano en introduisant divers objets parmi ses cordes.
4. Voir par exemple Berio L., Sequenza III (8'), Londres, Universal Edition, 1966.
5. Murail T., « Spectres et lutins (1982) », dans D. Cohen-Levinas (dir.), L'Itinéraire. La Revue musicale, no 421-422-423-424, Paris, 1991, p. 309-322.
6. Cendo R., « Saturation instrumentale : organisation et nouveaux enjeux pour la composition », Les mardis de la saturation #3, EHESS, NESAP, CDMC, 2010, en ligne : www.cdmc.asso.fr/fr/actualites/saison-cdmc/mardis-saturation.
7. Delalande F., « Qu'est-ce que le "son" ? » dans Le son des musiques, op. cit., p. 13-22.
8. Chion M., « La dissolution de la notion de timbre », Analyse musicale, no 3, 1986, p. 7-8.
9. Une FFT (fast fourier transform) est un algorithme permettant de réaliser sur ordinateur une transformée de Fourier discrète en économisant des calculs, à condition d'avoir une durée de fenêtre d'analyse égale à une puissance de deux échantillons. On utilise en général 4 096 échantillons pour des sons musicaux, soit 0,093 s (T) pour un taux d'échantillonnage de 44 100 Hz, norme du CD audio. La transformée donne alors un tableau de 2 050 lignes dont chaque ligne est une raie calculée de 10,75 Hz (f = 1/T) en 10,75 Hz jusqu'à la moitié du taux d'échantillonnage.
10. Dans le cas du diapason, à part le pic à 440 Hz, les autres pics observés sont très aigus et s'atténuent rapidement. On peut en faire abstraction à l'écoute et considérer que le son produit est presque un son pur. Si l'on considère des sons isolés, on peut les classer en plusieurs catégories. La première catégorie correspond aux sons élémentaires, dont les plus simples sont l'onde sinusoïdale (ou son pur), dont s'approche le diapason, et le dirac (ou impulsion), un son de durée courte (coup de feu, clap de cinéma par exemple). Les autres sons sont des sons complexes qu'on peut classer en trois sous-catégories : (1) les sons harmoniques dont l'onde sonore se répète périodiquement à une certaine fréquence (sons à hauteur déterminée) ; (2) les sons inharmoniques dont les partiels identifiables à l'analyse ne sont pas multiples d'une même fréquence fondamentale (il est difficile de leur attribuer une hauteur comme c'est le cas des sons de cloches) ; (3) les bruits, des sons dont le nombre de partiels est trop important pour qu'on puisse les identifier à l'analyse (par exemple une cymbale, une caisse claire avec timbre) ou qui varient trop vite dans le temps (comme les sons transitoires).
11. La fréquence fondamentale correspond théoriquement au premier harmonique du son (F0 x 1) et c'est souvent le partiel le plus fort, mais pas nécessairement. Parfois la fréquence fondamentale est totalement absente des analyses spectrales.
12. Leipp É., Acoustique et musique, Paris, Masson, 1971.
13. Castellengo M., Écoute musicale et acoustique, Paris, Eyrolles, 2015, p. 287-288.
14. Risset J.-C., An Introductory Catalog of Computer-Synthesized Sounds, Murray Hill, Bell Laboratories, 1969, réédité avec exemples sonores sur CD dans The Historical CD of Digital Sound Synthesis, Mayence, Allemagne, Wergo, 1995.
15. Rodet X. et BennettG., « Synthèse de la voix chantée par ordinateur », Conférences des journées d'étude, Festival international du son, Paris, 1980, p.73-91.
16. Potard Y., Baisnée P.-F. et Barrière J.-B., « Experimenting with models of resonance produced by a new technique for the analysis of impulsive sounds », dans Proceedings of the 1986 International Computer Music Conference, La Haye, 1986, p. 269-274. Voir aussi Potard Y., Baisnée P.-F. et Barrière J.-B., « Méthodologie de synthèse du timbre : l'exemple des modèles de résonance », dans Le Timbre. Métaphore pour la composition, Paris, Christian Bourgois- Ircam, 1991, p. 135-163.
17. Depalle Ph., Garcia G. et Rodet X., « A virtual Castrato (!?) », Proceedings of the International Computer Music Conference, Aarhus, Danemark, 1994.
18. Farchy J., « Les enjeux de l'IA dans l'industrie musicale », CNMlab, mars 2022, en ligne : cnmlab.fr/recueil/horizon-la-musique-en-2030/chapitre/3.
19. Lhérisson P.-R., Système de recommandation équitable d'œuvres numériques. En quête de diversité, thèse de doctorat en informatique, Lyon, université de Lyon, 2018, p. 3.
20. 1DTouch est une plateforme multimédia dédiée à la découverte culturelle, en ligne : 1dtouch.com.
21. Voir Grey J. M., « An exploration of musical timbre », rapport no STAN-M-2, Stanford University, février 1975, en ligne : ccrma.stanford.edu/files/papers/stanm2.pdf.
22. Voir les articles de Wessel D. L., « Psychoacoustics and Music. A Report from Michigan State University », Bulletin of the Computers Arts Society, vol. 30, 1973, p. 1-2, et « Timbre space as a musical control structure », Computer Music Journal, vol. 3, no 2, 1979, p. 45-52.
23. Peeters G., « A large set of audio features for sound description (similarity and classification) in the CUIDADO project », Paris, Ircam, 2004, p. 1-25 ; Peeters G., Giordano B. L., Susini P., Misdariis N. et McAdams S., « The timbre toolbox. Extracting audio descriptors from musical signals », The Journal of the Acoustical Society of America, vol. 130, no 5, 2011, p. 2902-2916.
24. Ircam, Orchestration assistée par ordinateur (Orchids), en ligne : www.ircam.fr/projects/pages/orchestration-assistee-par-ordinateur-orchids.
25. Ibid.
26. Zwicker E. et Terhardt E., « Analytical expressions for critical-band rate and critical bandwidth as a function of frequency », The Journal of the Acoustical Society of America, vol. 68, no 5, 1980, p. 1523-1525.

27. Rabiner L. R. et Juang B.-H., Fundamentals of Speech Recognition, Englewood Cliffs, Prentice Hall, 1993.
28. Voir le programme et les captations du colloque « Timbre is a many-splendored thing », Montréal, McGill University, 2018, en ligne : www.mcgill.ca/timbre2018/program.
29. Krumhansl C. L., « Why is musical timbre so hard to understand ? », dans S. Nielzen Olsson (dir.), Structure and Perception of Electroacoustic Sound and Music, Amsterdam, Excerpta Medica, 1989, p. 47.
30. « Alan Stivell – Pop Plinn (1971) » YouTube, 7 juillet 2009, en ligne : www.youtube.com/watch?v=2f8vL4JA0Ds.
31. Pottier L. et Wang N., « Analyses quantitatives de la musique par des mesures réalisées sur des signatures spectro-temporelles du son », Actes des Journées d'informatique musicale (JIM 2019), Bayonne, mai 2019.
32. Wang A. L.-C., « An industrial strength audio search algorithm », Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003), Baltimore, 2003, p. 26-30.
33. Schedl M., Gómez E. et Urbano J., « Music Information Retrieval. Recent developments and applications », Foundations and Trends in Information Retrieval, vol. 8, no 2-3, 2014, p. 132.
34. Voir l'ensemble de signatures sonores par les laboratoires CIEREC, ECLLA et LaHC, « Projet ANALYSE », Saint-Étienne, université Jean Monnet, 2017-2022, en ligne : musinf.univ-st-etienne.fr/recherches/signature/SiteAnalyse2018R/analyses3.html.
35. Peeters G., « A large set of audio features for sound description (similarity and classification) in the CUIDADO project », art. cité, et Peeters G. et al., « The timbre toolbox », art. cité.
36. Un formant est une région qui concentre plus d'énergie qu'ailleurs. Il est caractérisé par sa fréquence centrale (en hertz), sa largeur de bande (en hertz) et son amplitude.
37. Pottier L. et Wang N., « Analyses quantitatives de la musique par des mesures réalisées sur des signatures spectro-temporelles du son », art. cité.
38. « Dazed and confused » (1969-6 min 27) de l'album Led Zeppelin I, « Immigrant Song » (1970-2 min 26) et « Out on the Tiles » (1970-4'04) de l'album Led Zeppelin III ; « Stairway to Heaven » (1971-7 min 55) de l'album Led Zeppelin IV.
39. Pottier L., « Étude des caractéristiques acoustiques de quatre pièces du groupe Led Zeppelin et comparaison avec d'autres répertoires », dans Ph. Gonin (dir), Led Zeppelin. Contexte, analyse, réception, Dijon, EUD, 2021, p. 109-132.
40. Cette technique permet de projeter les observations d'un espace à n dimensions avec n variables vers un espace à deux dimensions, tout en conservant au maximum les informations provenant des dimensions initiales.
41. Projet financé par l'Agence nationale de la recherche (ANR).
42. Base de données musicales du projet ANR Wasabi : wasabi.i3s.unice.fr.
43. Dispositif qui permet de jouer un très court extrait d'un morceau de musique.
44. Menin A. et al., « Incremental and multimodal visualization of discographies. Exploring the WASABI music knowledge base », Actes de la conférence WAC22, Antibes, 2022.
45. Ce projet est issu de travaux menés initialement dans le cadre d'un groupe de travail de la Société française d'analyse musicale (SFAM) créé en 2012 portant sur « l'analyse des musiques électroacoustiques » et regroupant des chercheurs de l'université Paris 8, de l'Ircam, de l'université Paris-Sorbonne, de l'université de Huddersfield (Grande-Bretagne) et de l'université Jean Monnet à Saint-Étienne. Il a également obtenu un financement de la Fondation UJM en 2018. Ont participé à ce projet : Fabrice Muhlenbach et Pierre Maret, enseignants-chercheurs dans l'équipe « Connected Intelligence » du laboratoire Hubert Curien (UJM) ; et des étudiants stagiaires : Joseph Chataignon (Télécom Saint-Étienne), Mohammed-Bashir Mahdi (faculté des sciences et techniques-UJM), Na Wang (faculté des sciences et techniques-UJM et Mines Saint-Étienne), Lucas Veltri et Nicolas Roche (musicologie-UJM).