

LES DESCRIPTEURS AUDIO AU SERVICE DE L'ANALYSE COMPAREE DES CANAUX STEREOPHONIQUES DANS LES MUSIQUES POPULAIRES ENREGISTREES

Judikael Levin
Université Rennes 2
judilvn@gmail.com

Bruno Bossis
APP Université Rennes 2
bruno.bossis@univ-rennes2.fr

RÉSUMÉ

La spatialisation sonore dans les musiques populaires enregistrées reste une pratique encore peu étudiée. Avec l'émergence des technologies de diffusion audio 3D, l'étude de cette pratique sur l'ensemble du corpus est nécessaire afin de faciliter l'adoption de ces nouveaux outils créatifs.

Cet article présente l'état actuel d'une méthode d'analyse en cours de développement. Elle a pour but d'étudier à grande échelle la diversité stéréophonique des productions musicales populaires, des années 1950 à aujourd'hui, tous styles musicaux confondus. Elle entre dans un projet de classification des différents usages de l'espace sonore et de la spatialisation dans cet immense corpus. La classification sera fondée sur une échelle d'abstraction.

Cette méthode explore les possibilités analytiques des représentations sous-symboliques que sont les descripteurs audio pour les productions phonographiques des musiques populaires. Nous la présenterons à partir de deux descripteurs audio : le barycentre spectral et le spectral rolloff. Les deux canaux de la stéréophonie des morceaux étudiés seront analysés individuellement. Nous en tirerons des résultats issus de calculs statistiques et les comparerons. Nous tenterons ensuite de rapprocher ces résultats avec certaines réalités physiques. Ce sont ces résultats qui serviront ensuite d'indicateurs d'usages de la stéréophonie pour analyser de grands corpus. Elle est expérimentée ici à partir de deux corpus que nous nommerons « corpus étalon » et « corpus expérimental ». Le premier regroupe plusieurs versions d'une même composition dont le placement des sources dans la stéréophonie varie. Le second présente cinquante titres issus de différentes époques et tendances stylistiques. Nous présenterons dans cet article une analyse de ces deux corpus. Le corpus étalon fera l'objet de représentations schématiques des placements instrumentaux dans les espaces stéréophoniques avec des *Sound-Boxes*. Le corpus expérimental sera classé à partir

d'une grille d'évaluation prenant en compte les symétries instrumentales dans l'espace stéréophonique des morceaux. Nous tenterons ensuite de confronter ces analyses avec les résultats obtenus à l'aide des descripteurs audio.

1. INTRODUCTION

1.1. Contexte

« *Les pistes de rock en particulier ne sont plus considérées comme des «enregistrements» d'une performance live, mais plutôt comme une mise en forme spécifique d'«événements sonores» mélangés pour produire une expérience* »¹ [1]. Cette réflexion sur le statut de l'enregistrement sonore dans les musiques populaires est récurrente chez les musicologues spécialistes de la question. Theodor Gracyk parle de « performance virtuelle » et souligne que les enregistrements dépassent la représentation du réalisme [2]. Cette idée est également partagée par Albin J. Zak III [3].

Ce changement de considération de l'objet phonographique² a eu pour conséquence une diversification des espaces internes dans les productions musicales populaires. Les espaces sonores représentés se sont éloignés progressivement de la représentation de l'espace scénique vers la représentation de tous nouveaux espaces virtuels.

Ce passage s'est opéré dans les enregistrements monophoniques dès le début des années 1950 avec l'invention du multipiste et de l'*overdubbing* par Les Paul. Cette innovation a permis de fixer sur un même morceau plusieurs espaces avec leurs acoustiques propres [4].

Dans le cas des productions stéréophoniques, Allan F. Moore et Ruth Dockwray relèvent ce changement progressif entre 1966 et 1972 dans les productions rock [1].

¹ « *Rock tracks in particular are viewed no longer as 'records' of a live performance, but rather as a specific shaping of 'sonic events, mixed to produce an experience'.* »

² C'est-à-dire de celui d'une reproduction d'une performance *live* pour un nouveau type de performance propre à l'enregistrement

L'analyse de cet espace virtuel dans le champ des musiques populaires enregistrées est un terrain d'étude encore à défricher mais a déjà fait l'objet de tentatives de représentations telles que la *Sound-Box* de Moore et Dockwray³ [5]. Ce modèle de représentation est actuellement le plus employé lors d'analyses d'espaces internes dans les œuvres phonographiques.

Si l'on souhaite cependant dresser une typologie des différents espaces internes existant dans les productions stéréophoniques, l'analyste se retrouvera vite confronté au volume extrêmement important du corpus en question. L'étude comparative des tendances de pratique de larges corpus se révélerait donc difficile.

Couramment employés pour étudier le corpus des musiques contemporaines, les descripteurs audio peuvent relever une ou plusieurs caractéristiques du signal sonore en les convertissant en données. C'est cette possibilité que nous tenterons d'exploiter avec la méthode d'analyse présentée ici.

1.2. État de l'art

Les descripteurs audio sont des paramètres qui mettent en exergue une ou plusieurs dimensions du signal sonore. Ils sont utilisés dans de nombreux domaines d'applications allant de l'étude musicologique à la reconnaissance vocale et timbrique. Leur nombre ne cesse de croître et ils permettent aujourd'hui de relever l'ensemble des dimensions du signal sonore.

La musicologie des musiques savantes emploie couramment les descripteurs audio. On prendra comme exemple les travaux de Mikhail Malt [6-7] dont il représente les « masses sonores » à l'aide du BStD⁴ chez Xenakis⁵.

Dans le cas des musiques populaires enregistrées, les études utilisant l'AMAO⁶ sont encore peu nombreuses. Celles qui emploient les descripteurs audio à des fins analytiques le sont encore moins. En guise d'exemples, on peut citer David Brackett [8] et Philippe Gonin [9] qui utilisent le sonagramme pour analyser respectivement *Phill and Soaps* d'Elvis Costello et l'album *Atom Heart Mother* des Pink Floyd afin de démontrer certaines réalités musicales.

Philippe Lalitte présente les possibilités analytiques des descripteurs audio pour les musiques populaires en

illustrant avec une analyse de *Heroes* de David Bowie [10].

Après avoir exposé les besoins de l'étude, nous présenterons ladite méthode qui aura pour vocation d'y répondre. Nous testerons ensuite cette méthode à travers les deux corpus.

Le corpus étalon, est constitué de sept versions d'une composition créée pour l'occasion. Elle est composée de samples libres de droit et chacune des versions présente des placements instrumentaux différents dans la stéréophonie.

Le corpus expérimental, contient cinquante morceaux d'époques et de styles différents. Ils ont été composés, enregistrés et interprétés par des artistes de renom, anglo-saxons pour la plupart.

Nous concluons cet article par une réflexion autour des résultats obtenus et nous exposerons les prochains axes d'étude pour le développement de la méthode.

2. ÉLABORATION DE LA MÉTHODE

2.1. Description des besoins

La pratique de la stéréophonie fait appel à de nombreuses techniques de mixage audio, doublures⁷, élargissement du spectre, déplacements continus ou ponctuels. Il existe aujourd'hui bon nombre d'outils, de techniques de mixages et de mastering qui jouent sur les différences timbriques, temporelles ou encore spectrales entre le canal gauche et le canal droit⁸.

Ces pratiques répondent soit à des besoins palliatifs tels que l'intelligibilité de l'instrumentation et la reproduction d'un espace scénique *live*, ou bien créatif en répondant à des besoins spécifiques au discours musical et artistique [4].

La méthode d'analyse que nous essayons de développer ici pour objectif d'analyser de grands corpus en tirant des informations essentielles relatives aux placements instrumentaux dans la stéréophonie à l'aide des descripteurs audio et de valeurs statistiques.

Cette méthode s'inscrit dans une étude plus large qui tente de proposer une classification des espaces sonores dans les musiques populaires enregistrées. Elle classera ces musiques en fonction du degré d'abstraction de leurs espaces sonores à partir de

³ La *Sound-Box* est un modèle de représentation de l'espace stéréophonique interne d'un enregistrement musical. Elle se présente comme un pavé droit (qui représente la scène virtuelle). On y place les différentes sources sonores du titre en fonction de leur emplacement dans la stéréophonie.

⁴ Pour *Brightness-Standard Deviation*. C'est un outil de représentation conçue par Mikhail Malt et Emmanuel Jourdan. Il consiste en une représentation sur trois dimensions. Elle permet notamment de déterminer la brillance du spectre.

⁵ *Diamorphoses* (1957), *Concret PH* (1958), la version concert d'*Orient-Occident* (1960) et *Bohor* (1962).

⁶ Analyse Musicale Assistée par Ordinateur.

⁷ Pratique de studio qui consiste à enregistrer deux fois, ou plus, la même ligne instrumentale. Les différentes pistes obtenues sont ensuite réparties dans la stéréophonie. Cette technique donne l'illusion d'un seul instrument joué mais que les micro-différences (temporels et timbriques) liées aux différentes prises participent à mettre en avant l'instrument et à accentuer sa dynamique.

⁸ Pour des raisons de praticité, nous remplacerons les termes "canal gauche" et "canal droit" respectivement par *L* et *R* pour la suite de cet article.

l'utilisation des techniques de spatialisation⁹, des facteurs extramusicaux qui gravitent autour de ces musiques et de leurs tendances stylistiques.

Une fois mise en place, cette classification servira également de base de réflexion et d'usage à l'emploi des techniques de mixages et d'enregistrements audio 3D à destination des musiques populaires enregistrées.

2.2. Descripteurs audio employés et méthode d'obtention des valeurs comparatives

En s'inspirant des textes de Mikhail Malt [6] et de Eric Scheirer et Malcolm Slaney [11], nous nous sommes pour le moment concentrés sur l'utilisation de deux descripteurs audio, le spectral centroid (ou barycentre du spectre, ou *SC*) et le spectral rolloff (ou *SRF*)¹⁰.

Glenn Hall [12] définit le *SC* comme le barycentre spectral. Il quantifie la brillance du son et la sensation de netteté. Cette dernière est accrue d'autant que la valeur du *SC* est élevée. Elle partage le spectre en deux parties d'égales énergies. Enfin, il encode une partie de notre perception du timbre [6] et permet une représentation synthétique du spectre liée à la perception des sons instrumentaux [13].

Le *SRF* est défini par Pierre Couprie [14] comme la fréquence en dessous de laquelle se concentre un certain pourcentage de l'énergie spectrale (environ 95%). Tout comme le *SC*, le *SRF* est un bon indicatif de la brillance spectrale. Il est utilisé également pour la discrimination parole / musique [11].

Comme le signalent Malt ou encore Schreier et Slaney, ces descripteurs sont souvent corrélés. Le fait de les utiliser ensemble nous permettra de mettre en avant à quels moments il y a corrélation ou pas dans un contexte comme celui de notre cas d'étude.

Nous posons l'hypothèse qu'en comparant les valeurs issues de calculs statistiques - comme les médianes et les quartiles 1 et 3¹¹ - des descripteurs audio¹² des canaux *L* et *R* d'un même titre, nous pourrions spéculer sur la similarité des deux canaux et des placements instrumentaux dans la panoramique.

Ainsi, plus les valeurs obtenues entre *L* et *R* sont différentes, plus la dissymétrie instrumentale sera importante. L'écoute des deux canaux serait indispensable pour l'appréciation globale du titre et de son espace sonore.

À l'inverse, plus les valeurs obtenues entre *L* et *R* sont proches, plus le mixage sera symétrique, centré, peu

ouvert et/ou avec des placements instrumentaux similaires.

À partir des deux corpus, dont nous avons au préalable analysé, classé et/ou représenté les espaces sonores des morceaux les constituants, nous allons tenter de répondre à cette hypothèse.

3. EXPÉRIMENTATION

3.1. A partir du « corpus étalon »

Le corpus étalon présente différentes versions d'une même composition écrite pour l'occasion par nous-même¹³. Cette composition, qui est une boucle audio d'une durée de 17s, est composée de quatre samples issus de la bibliothèque de *loops* du séquenceur Logic Pro X¹⁴. L'instrumentation, de type pop, est composée d'une batterie, d'une basse, d'un piano et d'une guitare¹⁵ sur une tonalité de *la#*.

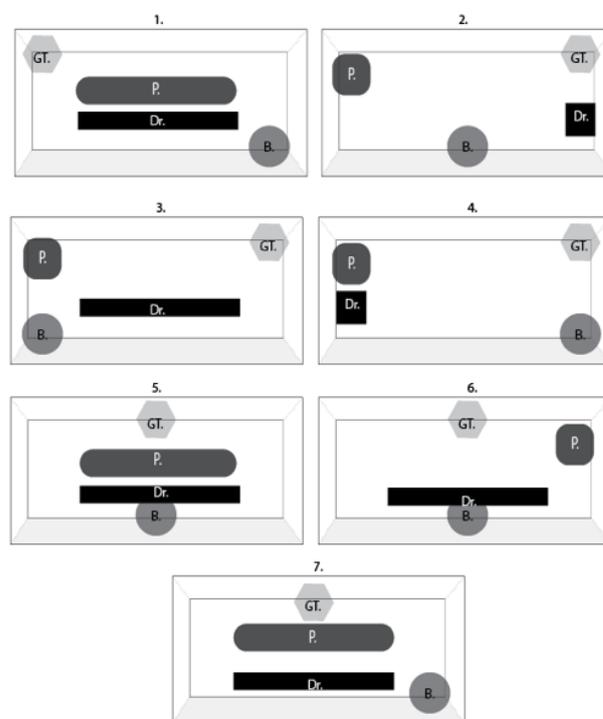


Figure 1. Sound-Boxes des différentes versions de l'extrait constituant le corpus étalon¹⁶.

⁹ Nous entendons par ici les placements instrumentaux dans la stéréophonie, l'emploi des effets de réverbération et d'écho et les techniques d'enregistrements et de mixages relatives aux idées de proximité et de distance avec l'auditeur.

¹⁰ Les calculs des descripteurs ainsi ont été effectués à partir du logiciel Sonic Visualiser avec les plug-ins QM VAMP du *Center for digital Music* du *Queen Mary University of London*. Liens internet : www.sonicvisualiser.org ; www.vamp-plugins.org. Respectivement V.3.1.1.1. et V.1.7.1.

¹¹ Médiane et Quartile 1 et 3 que nous appellerons respectivement M, Q1 et Q3.

¹² *SC* et *SRF* dans le cas de cet article.

¹³ Ce corpus est écoutable vers ce lien : jlevin.bandcamp.com.

¹⁴ Version 10.4.1, édité par Apple.

¹⁵ Dans Logic Pro X, La piste de batterie est issue du jeu *Half-pipe* du batteur virtuel *Kyle* sur le kit batterie Logic Pro *SoCal*. Les autres pistes sont respectivement nommées *Straight Ringer Bass*, *Motorway Guitar* et *Disco Delight Piano*.

¹⁶ GT. Pour guitare, P. pour piano, Dr. Pour batterie et B. pour basse.

Version	Descripteurs Audio	Fréquences			Fréquences			Fréquences		
		Quartiles 1		Différence	Médianes		Différence	Quartiles 3		Différence
		L	R		L / R	L		R	L / R	
1	CS	1086	766	320	1469	1247	222	2146	2369	223
	SRF	2541	1766	775	3187	2799	388	5211	8958	3747
2	CS	294	855	561	544	1307	763	760	2076	1316
	SRF	431	2196	1765	1507	2972	1465	1938	6589	4651
3	CS	747	1184	437	1022	1694	672	2164	2574	410
	SRF	1593	2670	1077	2455	3618	1163	7149	7924	775
4	CS	1414	509	905	2261	693	1568	3473	1059	2414
	SRF	3230	1249	1981	6891	2110	4781	10853	2713	8140
5	CS	852	886	34	1178	1242	64	1787	1960	173
	SRF	2153	2196	43	2756	2842	86	4522	5469	947
6	CS	813	886	73	1193	1253	60	1853	1927	74
	SRF	2153	2196	43	2842	2885	43	5082	5469	387
7	CS	1083	888	195	1451	1251	200	2137	1905	232
	SRF	2541	2196	345	3187	2842	345	5125	5297	172

Table 1. Différences des quartiles 1 et 3 et des médianes des *CS* et *SRF* entre *L* et *R* dans les sept versions du corpus étalon.

Sept versions ont été mixées. Chacune présente une disposition différente des instruments dans la stéréophonie. La figure 1 montre ces placements en suivant le modèle de la *Sound-Box* de Moore et Dockwray.

Il est à noter que la piste batterie et piano sont stéréophoniques. Les canaux gauche et droit sont donc différents. Ces instruments sont représentés par des barres horizontales quand ils sont placés au centre du mix pour illustrer cela.

Après avoir effectué la différence des médianes et des quartiles 1 et 3 des *SC* et des *SRF* des canaux *L* et *R* des différents mixages on obtient les résultats présentés dans la table 1.¹⁷

À partir de ce corpus étalon, nous pouvons relever des tendances au niveau des résultats obtenus. Sur la table 1, nous constatons que les versions 5, 6 et 7 présentent les plus faibles différences entre *L* et *R*. Nous remarquons sur la figure 1, que les instruments sont principalement placés au centre du mix dans ces versions¹⁸.

Nous avons ensuite les versions 1 et 3 avec la batterie au centre. Le piano tient la même position dans la version 1. La guitare et la basse sont à l'opposé dans les deux cas. Ce sont les versions 2 et 4 qui présentent les plus grands écarts entre *L* et *R*. Dans ces deux versions les instruments stéréophoniques (piano et batterie) sont aux extrêmes (et par conséquent monophoniques). La version 2 présente

une basse au centre tandis que la version 4 a *L* et *R* qui sont totalement indépendants. C'est d'ailleurs la version 4 qui présente les différences entre les Q1, M et Q3 les plus importantes.

Nous pouvons donc supposer une corrélation entre les différences instrumentales entre *L* et *R* dans le mix et la différence des résultats statistiques des *SC* et *SRF* entre les deux canaux de la stéréophonie.

Les résultats obtenus à partir des *SRF* semblent être les plus révélateurs. Plus les différences entre *L* et *R* sont importantes et audibles dans le mixage, plus les différences entre les médianes et quartiles 1 et 3 des *SC* et *SRF* entre *L* et *R* sont importantes. Ces valeurs pourraient donc être des indices suggérant la diversité de traitements entre les deux canaux de la stéréophonie.

¹⁷ Les résultats présentés ici sont les valeurs absolues arrondies à l'unité.

¹⁸ Il y en a donc, au moins, trois instruments au centre du mix si on suit la figure 1.

Registre instrumental		A	AB	BA	B	BC	CB	C
Totem	Dissymétrie	+++	++	+	+			
	Symétrie	+	+++	+++	+++	+++	+++	+++
Mélodique	Dissymétrie	+++	++	+				
	Symétrie		+	++	+++	+++	+++	+++
Harmonique homorythmique	Dissymétrie	+++	+++	++	+			
	Symétrie		+	+	++	+++	+++	+++
Harmonique polyrythmique	Dissymétrie	+++	+++	+++	++	+		
	Symétrie		+	+	+	++	+++	+++
Ornemental	Dissymétrie	+++	+++	++	+			
	Symétrie	+	+	+	++	+++	+++	+++
Déplacements cycliques		+	+	+	+	+		
Déplacements ponctuelles		+	+	+				
Largeur du mix		+++	++	+	+	+	--	---

Table 2. Grille d'évaluation du placement des instruments dans la stéréophonie.

3.2. À partir du « corpus expérimental »

Afin de confirmer les résultats, la méthode a ensuite été appliquée au corpus expérimental. Ce corpus est constitué de 50 titres provenant de diverses tendances stylistiques et époques. La plupart sont des productions anglo-saxonnes¹⁹. Certains de ces titres ont déjà fait l'objet d'études sur la constitution de leurs espaces sonores²⁰. Les autres productions ont été sélectionnées à partir d'une discographie de 1125 titres²¹. La sélection de ces titres s'est faite à l'aide de la fonction « lecture aléatoire » du lecteur audio iTunes.

3.2.1. Classification du corpus

Après avoir effectué une analyse auditive et construit la *Sound-Box* de ces titres, nous les avons classés en sept catégories autour de trois pôles (A, B et C) en fonction de la diversité entre *L* et *R* et de l'emploi de la stéréophonie.

Le pôle A regroupe les titres dont *L* et *R* sont complémentaires et indispensables à l'appréciation et l'œuvre dans sa globalité. Nous entendons par ici :

- Le placement des instruments rythmiques, percussifs, harmoniques et mélodiques n'est pas symétrique par rapport au centre du mix.
- La présence de déplacements cycliques ou ponctuels d'une ou plusieurs sources.
- Un empan²² stéréophonique important.

Le pôle B comporte les titres ayant :

- Un rapport quasi-frontal avec l'auditeur et la scène virtuelle.

- Des placements instrumentaux se rapprochant au maximum de placements scéniques et des standards de mixage. À savoir : voix lead, basse et batterie au centre et les autres instruments à *L* ou à *R*.
- Aucun ou très peu de déplacements instrumentaux.
- Un spectre stéréophonique plus étroit.

Enfin le pôle C comporte les titres ayant :

- Le spectre stéréophonique et les placements instrumentaux qui sont étroits.
- La quasi-totalité des sources se situent au centre du mix. Il n'y a aucune différence notable entre *L* et *R*. Le titre est (ou, presque) monophonique.

Les sept catégories en question sont les suivantes : A, AB, BA, B, BC, CB et C. La première lettre correspond au pôle auquel le titre se rattache et la seconde souligne la présence d'éléments appartenant au pôle en question. Pour obtenir cette classification qui se veut la plus objective possible, nous avons fait l'analyse auditive à partir d'une grille d'évaluation que nous avons établie. Celle-ci prend en compte plusieurs critères dans le mixage²³. La table 2 montre cette grille de classification.

Le Totem comprend l'assise rythmique du titre soit l'ensemble *kick*, *snare* et basse. Le registre mélodique comprend les lignes mélodiques faites par le chant, les mélodies instrumentales et les soli. Les accompagnements harmoniques en homorythmie concernent les doublures instrumentales réparties dans la stéréophonie. Elles sont généralement à l'unisson. Les

¹⁹ 39 productions anglo-saxonnes, 7 françaises, 2 jamaïcaines, 1 allemande et 1 japonaise

²⁰ *1'000'000* (Nine Inch Nails) par Sébastien Saint-André [15], *Born Free* (Matt Monro), *Lola* (The Kinks), *Purple Haze* (Jimi Hendrix Experience) et *Dick-a-Dum-Dum* (Des O'Connor) par Allan F. Moore [5].

²¹ Cette discographie comporte des titres de production stéréophonique provenant de plusieurs styles musicaux depuis 1960. Environ 75% de

cette discographie comporte des productions d'artistes anglo-saxons et 50% sont d'un registre rock.

²² Terme issu du registre visuel. L'empan visuel correspond à la largeur de texte où le lecteur est en possibilité de tirer des informations utiles. Dans le cas de la perception auditive, l'empan correspondrait à la largeur du spectre stéréophonique.

²³ Les écoutes ont été effectuées au casque. Cette classification garde une part de subjectivité et d'interprétation.

harmoniques polyrythmiques concernent les jeux rythmico-harmoniques entre, au moins, deux instruments dont chacun est placé à différents endroits dans la stéréophonie. Enfin le registre ornemental concerne les instruments et les sources sonores ponctuelles présents à des fins de ponctuations et d'accompagnement rythmique. Les percussions, hors « totem », entrent également dans ce registre instrumental.

Les trois autres critères d'évaluation concernent la présence de déplacements cycliques et ponctuels et l'empan stéréophonique.

L'évaluation de la symétrie ou dissymétrie des registres instrumentaux est faite sur quatre niveaux (+++, +, +, et " "). Chacun de ces niveaux indique la probabilité de symétrie ou de dissymétrie du registre instrumental dans la catégorie en question, "+++" voulant dire "très probablement présent" et " " pour "absent dans cette catégorie".

Pour la largeur du spectre stéréophonique, nous avons également rajouté les notations "--" et "---" pour faire une distinction entre la catégorie C et CB.

Il est à noter qu'il y a une certaine part de subjectivité dans cette classification. Nous en retiendrons principalement l'attraction autour d'un des pôles A, B ou C décrits précédemment.

Suite à cette classification, notre corpus expérimental est composé de :

- 6 titres en catégorie A.
- 15 titres en catégorie AB.
- 9 titres en catégorie BA.
- 11 titres en catégorie B.
- 6 titres en catégorie BC.
- 2 titres en catégorie CB.
- 1 titre en catégorie C.

Soit 20 titres autour du pôle A, 27 titres autour du pôle B et 3 titres autour du pôle C.

3.2.2. Résultats du corpus expérimental

Après avoir effectué les mêmes calculs que pour le corpus étalon, on obtient les résultats compilés dans la table 3. Chaque résultat correspond à la valeur absolue de la différence (en Hz) entre les Q1, M et Q3 des descripteurs audio correspondants des canaux gauches et droits des titres du corpus.

Après avoir effectué les mêmes calculs que pour le corpus étalon, on obtient les résultats compilés dans la table 3. Chaque résultat correspond à la valeur absolue de la différence (en Hz) entre les Q1, M et Q3 des

descripteurs audio correspondants des canaux gauches et droits des titres du corpus.

Pour une meilleure lecture, nous avons triés les résultats en quatre catégories :

- La première concerne une différence de plus de 1000Hz entre *L* et *R* (cellules grises foncées).
- La deuxième rassemble les résultats compris entre 500 et 999Hz (cellules grises).
- La troisième entre 100 et 499Hz (cellules grises claires).
- La quatrième entre 0 et 99Hz (cellules blanches).

Les titres sont triés en fonction de la catégorie où ils ont été placés suite à l'analyse auditive précédente et aux résultats obtenus (du plus grand nombre de cellules foncées, par catégories, au plus faible).

Au regard des résultats obtenus sur la table 3 entre la classification faite et les résultats statistiques, on peut y déceler un lien entre ces deux types de résultats. Plus le morceau est proche de la catégorie A (ou en fait partie), plus les différences entre les Q1, M et Q3 des canaux gauches et droits sont élevées. Et inversement, plus nous allons vers la catégorie C, plus ces différences s'amenuisent.

Ces résultats doivent être nuancés. Dans la table 3, nous avons surligné en gris les titres dont la classification faite et les résultats obtenus ne correspondent pas. Ils sont au nombre de treize (soit environ 25% du corpus expérimental). Ces différences peuvent s'expliquer de diverses manières.

Certains titres ont une charge instrumentale importante et/ou ont un fort taux de compression quand on regarde leur spectre d'amplitude²⁴. Ces deux raisons peuvent brouiller l'analyse auditive ou l'étude à partir de certains descripteurs audio.

Cela peut venir également des descripteurs audio utilisés et du fait de relever des valeurs statistiques qui, certes, révèlent une différence de diffusion dans la stéréophonie mais ne mettent pas en avant les déplacements ponctuels d'instruments tel que les chœurs dans *Bohemian Rhapsody* de Queen.

La dernière raison qui pourrait expliquer ces différences viendrait de la classification faite au préalable des morceaux du corpus expérimental (de A à C). Cette classification établie à partir de la grille d'évaluation relève une part de subjectivité. Il serait bon à l'avenir de proposer la classification de ces titres, avec la grille d'évaluation en question, à un panel d'auditeurs afin de rendre cette classification plus objective. Ce projet de crowdsourcing²⁵ est actuellement en cours de réflexion autour des différentes plateformes tel que Zooniverse²⁶ ou encore Google Forms.

²⁴ Ces fortes compressions sont à mettre dans le contexte de la *Loudness War*, soit "la guerre du volume". Elle fait référence à une pratique de l'industrie du disque qui consiste à augmenter le sentiment de volume sonore en augmentant la compression et donc en réduisant la plage dynamique du morceau. Cette pratique a pris davantage d'ampleur ces

dernières décennies malgré les nombreuses critiques dont elle est sujette. Voir Robert Levine [16].

²⁵ Production participative.

²⁶ www.zooniverse.org

L'autre point qui appelle à la prudence vient du corpus expérimental et de la manière dont il s'est constitué. En effet, la base de données sources de 1125 morceaux d'où provient ce corpus n'est pas représentative de l'ensemble des productions musicales. De plus, le déséquilibre dans le nombre de titres entre les pôles A, B et C appelle à garder un certain recul et à effectuer d'autres relevés en élargissant ce corpus pour avoir des résultats plus objectifs à partir d'un corpus équilibré.

Malgré ces remarques, nous pouvons relever une corrélation entre la diversité stéréophonique et les résultats obtenus à partir des valeurs statistiques des descripteurs audio utilisés. Et ce, aussi bien à partir du corpus étalon qu'à partir du corpus expérimental. Ces résultats vont dans le sens de l'hypothèse qui a été faite.

Titre - Artiste - Année	Catégorie	CS			SRF		
		Q1	M	Q3	Q1	M	Q3
<i>The Wind Cries Mary</i> – Jimi Hendrix Experience – 1967	A	702	764	914	2713	2799	2842
<i>Stairway To Heaven</i> – Led Zeppelin – 1971	A	325	532	718	1163	1637	3101
<i>War Pigs</i> – Black Sabbath – 1970 (ed. 2009)	A	159	470	795	1249	2239	2756
<i>Josephine</i> - Tori Amos – 1995	A	485	596	221	2670	3618	1464
<i>Computer Love</i> – Kraftwerk – 1981	A	191	521	725	345	3058	2110
<i>Space Oddity</i> – David Bowie – 1969	A	44	81	942	474	689	991
<i>One Step Beyond</i> – Madness – 1979	AB	356	900	1481	775	1637	3575
<i>Bongo Bongo</i> – Manu Chao – 1998	AB	320	378	485	689	1120	2369
<i>Don't Know Why</i> – Norah Jones – 2002	AB	137	256	335	603	1206	1507
<i>Born Free</i> – Matt Monro – 1966	AB	164	244	348	560	1120	1593
<i>Everything In Its Right Place</i> – Radiohead – 2000	AB	274	122	79	1120	345	86
<i>Africa</i> – Toto – 1982	AB	334	303	390	947	689	732
<i>Purr</i> – Sonic Youth – 1992	AB	152	125	198	991	517	635
<i>Roundabout</i> – Yes - 1971	AB	65	216	237	86	818	904
<i>Everything Counts</i> – Depeche Mode – 1983	AB	138	207	398	646	431	517
<i>Golden Brown</i> – The Stranglers – 1982	AB	72	379	420	560	301	172
<i>FloriDada</i> – Animal Collective – 2016	AB	115	113	159	431	345	517
<i>Coffee Shop</i> – Red Hot Chili Peppers – 1995	AB	206	205	240	345	258	172
<i>Me Gustas Tu</i> – Manu Chao – 2001	AB	19	29	78	0	0	517
<i>Knights of Cydonia</i> – Muse – 2006	AB	93	52	24	129	86	0
<i>Bohemian Rhapsody</i> – Queen – 1975	AB	4	1	22	86	43	86
<i>Technopolis</i> – Yellow Magic Orchestra – 1979	BA	118	61	897	258	517	3273
<i>Come As You Are</i> – Nirvana – 1991	BA	240	243	261	861	646	775
<i>Men In Black</i> – Frank Black – 1996	BA	182	23	92	86	517	775
<i>Buffalo Soldier</i> – The Wailers – 1983	BA	21	60	95	86	560	517
<i>Sabrosa</i> – Beastie Boys – 1994	BA	99	202	313	388	560	431
<i>Dirty Water</i> – Foo Fighters – 2017	BA	132	125	224	388	431	560
<i>Dani California</i> – Red hot Chili Peppers – 2006	BA	119	135	236	215	258	301
<i>l'000'000'</i> - Nine Inch Nails – 2008	BA	115	21	13	301	129	86
<i>Purple Haze</i> – Jimi Hendrix Experience – 1967	BA	16	46	27	43	129	43
<i>Black Betty</i> – Ram Jam – 1977	B	90	212	538	258	947	1292
<i>Virtual Insanity</i> – Jamiroquai – 1996	B	66	93	346	86	215	1163
<i>Around The World</i> – Daft Punk – 1997	B	174	309	423	732	646	258
<i>Dick-a-Dum-Dum</i> – Des O'Connor – 1969	B	17	30	197	43	86	646
<i>Master Of Puppets</i> – Metallica – 1986	B	139	142	127	388	258	301
<i>Rocket Man</i> – Elton John – 1972	B	98	142	197	431	431	301
<i>The Beautiful Game</i> – Vulfpeck – 2016	B	36	106	141	301	345	301
<i>Faut Pas Pleurer Comme Ça</i> – D. Guichard – 1994	B	114	92	79	172	86	43
<i>Atlas</i> – Battles – 2007	B	71	123	46	86	258	0
<i>Before I Forget</i> – Slipknot – 2004	B	80	65	22	43	86	345
<i>No Tears To Left To Cry</i> – Ariana Grande – 2018	B	7	19	42	86	0	43
<i>The Stepping Dub</i> – King Tubby – 1975	BC	257	430	690	1120	1120	689
<i>Spiders</i> – System Of A Down – 1988	BC	46	216	396	86	904	1163
<i>God Save The Queen</i> – Sex Pistols – 1977	BC	201	78	148	215	0	646
<i>Aerials</i> – System Of A Down – 2001	BC	136	0	4	345	517	345
<i>Starlight</i> – The Supermen Lovers – 2001	BC	48	18	177	431	215	86
<i>Lola</i> – The Kinks – 1971	BC	78	64	11	172	0	129
<i>Retour Aux Pyramides</i> – Les X – 1997	CB	41	47	81	258	129	86
<i>Nightcall</i> – Kavinsky – 2011	CB	10	11	52	43	86	0
<i>Sum Of Us</i> – Jurassic 5 – 2002	C	4	16	21	43	43	43

Table 3. Résultats du corpus expérimental en fonction de la classification des titres.

4. CONCLUSION ET SUITE DU TRAVAIL

Le but premier de cet article est d'entrevoir une nouvelle méthode d'analyse de l'espace sonore et de la spatialisation dans les musiques populaires enregistrées. Elle cherche à obtenir des valeurs objectives pour l'analyse comparée de très grands corpus en tirant les informations relatives à la diversité stéréophonique à partir de valeurs statistiques de descripteurs audio. Cette méthode d'analyse s'inscrit dans un projet plus vaste qui vise à classifier les tendances d'usage de l'espace sonore et de la spatialisation en fonction des tendances stylistiques et des époques sur une échelle d'abstraction. Cette classification nous aidera à comprendre davantage la place que peut porter l'espace sonore et la spatialisation dans le discours musical et servira de guide de réflexion et d'usage pour intégrer les pratiques de l'audio 3D dans les musiques populaires enregistrées²⁷.

L'étude du corpus étalon, établi à partir de sept versions différentes d'une même composition, a permis de corrélérer diversité stéréophonique et différences des valeurs statistiques des *SC* et *SRF* des canaux *L* et *R*. Le corpus expérimental souligne un phénomène similaire en comparant la classification faite avec les résultats statistiques.

Ces résultats vont dans le sens de l'hypothèse faite au début de cet article, à savoir que les valeurs statistiques des descripteurs audio peuvent être des indicateurs pour relever des différences timbriques entre les deux canaux d'un même morceau. Elles nous permettraient ainsi de spéculer sur les placements instrumentaux.

Il est encore trop tôt pour conclure les faits. Cette étude doit être approfondie pour faire les rapprochements entre les résultats statistiques et les placements de tel ou tel registre instrumental dans la stéréophonie. De nombreuses autres pistes d'études à partir des descripteurs audio sont en cours. Une de ces approches consistera à séparer par bandes de fréquences les titres à étudier (basses, médiums-bas, médiums et aigues) et à les analyser de manières différentes afin de relever plus précisément les placements instrumentaux. D'autres descripteurs audio seront expérimentés à partir de cette méthode et sur ce corpus afin d'affiner nos résultats. Le spectral kurtosis, l'inharmonicité ou encore les tristimulus serviront de révélateurs timbriques tandis que le *zero crossing rate* et le *noisiness* nous révéleront la quantité de bruit dans le signal. Enfin la courbe de rugosité (ou *roughness*) nous indiquera le taux de dissonance.

Ce nouvel outil d'analyse en développement et le projet de classification dans lequel il s'inscrit ont pour objectif une compréhension plus approfondie des espaces

sonores et de la pratique de la spatialisation dans les musiques populaires enregistrées. Ces informations serviront ensuite de guide d'usages de l'audio 3D adaptées à ces musiques.

5. REFERENCES

- [1] Moore, A. F. Dockwray, R. "The Establishment of The Virtual Performance Space in Rock", *Twentieth-Century Music*, 5/2, Cambridge University Press, Cambridge, UK, 2008.
- [2] Gracyk, T. *Rhythm of Noise*, I. B. Tauris, London, UK, 1996.
- [3] Zak, A. J. III. *The Poetics of Rock: Cutting Tracks, Making Records*, University of California Press, Berkeley, USA, 2001.
- [4] Doyle, P. *Echo & Reverb. Fabricating Space Of Popular Music. 1900-1960*. Wesleyan University Press, Middletown, USA, 2005.
- [5] Moore, A. F. Dockwray, R. "Configuring The Sound-Box 1965-1972", *Popular Music*, 29/2, Cambridge University Press, Cambridge, UK, 2010.
- [6] Malt, M. "Une proposition pour l'analyse des musiques électroacoustiques de Xenakis à partir de l'utilisation de descripteurs audio", *Iannis Xenakis, La musique électroacoustique*, L'Harmattan, Paris, p. 159-198, 2015.
- [7] Malt, M. Jourdan, E. (2009), "La « BSTD » – Une représentation graphique de la brillance et de l'écart type spectral, comme possible représentation de l'évolution du timbre sonore", textes du colloque international *L'Analyse Musicale Aujourd'hui, Crise ou (r)évolution ?*, Université de Strasbourg/SFAM, novembre 2009.
- [8] Brackett, D. *Interpreting Popular Music*, University of California Press, Berkeley, USA, 2000.
- [9] Gonin, P. *Pink Floyd, Atom Heart Mother*, Scérén/CNDP, Paris, 2011.
- [10] Lalitte P. "Outils informatiques et méthodes pour l'analyse des musiques actuelles", *Focus sur le rock en France*, Éditions Delatour France, Sampzon, p. 191-213, 2014.
- [11] Scheirer, E. Slaney. M., "Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator", *Proc. ICASSP '97*, Munich, Allemagne, p. 1331-1334, 1997.
- [12] Hall, G. *Identification d'instruments de musique à l'aide de méthodes statistiques et d'algorithmes d'intelligence artificiel*, Mémoire de maîtrise, Université du Québec, Chicoutimi, 2013.

²⁷ Mixages et enregistrements binauraux en premier lieu.

- [13] Tardieu, D. *Modèles d'instruments pour l'aide à l'orchestration*, thèse de doctorat, dir. Rodet, X., Université Pierre et Marie Curie, 2008.
- [14] Couprie, P. "Analyse et représentation du fantastique dans les musiques anecdotiques de Luc Ferrari", *Le fantastique dans les musiques des XXe et XXIe siècles*, Delatour, p. 94-108, 2017.
- [15] Saint-André, S. "L'analyse du discours musical dans le rock : l'exemple de « 1,000,000 » de Nine Inch Nails", *Focus sur le rock en France*, Éditions Delatour France, Sampzon, p. 173-190, 2014.
- [16] Levine, R., "The Death of High Fidelity", *Rolling Stone Magazine*, 27 décembre 2007: http://www.electriccity.be/Images/The%20Death%20of%20High%20Fidelity%20_%20Rolling%20Stone.pdf (dernière consultation le 10/04/2019).