

UNIVERSITÉ DE BORDEAUX

MÉMOIRE DE STAGE

Conception de scénarios temporels de spatialisation

Présenté par **Lucas BROUET**

Enseignant-référent :
Pierre BENARD

Tuteur en entreprise :
Joseph LARRALDE

Stage effectué au SCRIME Du **1 avril** au **30 septembre**

Année universitaire 2024-2025

Table des matières

1	Introduction	3
2	Le SCRIME	4
2.1	Les équipements	4
2.2	Les événements	5
3	Contexte du stage	6
4	Objectif de stage	7
5	Planning	9
5.1	Planning initial	9
5.2	Déroulement effectif	9
6	Outils utilisés	10
6.1	Max/MSP	10
6.2	Reaper	11
7	Réalisation	11
7.1	Données en entrée	13
7.2	Analyse et Visualisation	14
7.2.1	Analyse de l'audio	15
7.2.2	Analyse de la spatialisation	17
7.2.3	Interprétation et intuition	22
7.3	Entraînement d'un modèle	23
7.3.1	Sélection du jeu de donnée	23
7.3.2	Sélection du modèle	25
8	Résultats	27
8.1	Entraînement du modèle	27
8.1.1	Sélection d'extrait musical	27
8.1.2	Sélection et formatage des données	28
8.2	Prédiction du modèle	29
8.2.1	Observation	29
8.2.2	Analyse et discussion	31
8.2.3	Améliorations possibles et perspectives	32
9	Bilan / Conclusion	34
10	Glossaire	35
	Glossaire	35

Remerciements

Je tiens à exprimer ma sincère gratitude à toutes les personnes qui m'ont accompagné au cours de ce stage

Je remercie tout particulièrement mes tuteurs, **[Joseph Larralde]**, **[Louis Bigo]** et **[Pierrick Legrand]** pour leur encadrement attentif, leurs précieux conseils et leur soutien tout au long de ce projet. Leur expertise et leur disponibilité ont été déterminantes pour le bon déroulement de mon travail et m'ont permis de progresser tant sur le plan scientifique que personnel.

Je souhaite également remercier **[Annick Mersier]**, pour son accueil tout à fait chaleureux, ainsi que son accompagnement des stagiaires tout au long du stage.

J'aimerais de plus remercier mes collègues stagiaires, **[Noémie Kegl–Tourneix]**, **[Florent Donat-Bouillud]** et **[Victor Benitez Ramirez]** pour leur compagnie agréable ces derniers mois.

Enfin, j'aimerais remercier **[Pierre Bénard]** pour son accompagnement sur la durée de ce stage.

[Lucas BROUET]

1 Introduction

Depuis longtemps, les êtres humains se sont intéressés à la perception du son dans l'espace et ont constaté à de nombreuses reprises que cette expérience pouvait varier considérablement. Modifier l'acoustique d'un lieu ou déplacer la source sonore par rapport à l'auditeur transforme radicalement la manière dont le son est perçu. Dans le domaine musical, cet effet est encore plus marqué : la recherche d'un rendu précis implique de porter une attention particulière aux conditions d'écoute, qui influencent directement l'expérience artistique. Cette préoccupation n'est pas nouvelle, puisqu'elle était déjà présente dès l'Antiquité dans la construction des théâtres grecs et romains, où l'acoustique était pensée de manière à ce que chaque spectateur puisse entendre distinctement les acteurs. Plus tard, au Moyen Âge, les grandes cathédrales furent conçues de manière à magnifier les chants liturgiques et à leur donner une dimension spirituelle plus forte. Le recours à des volumes immenses, à des voûtes et à des coupoles visait à prolonger la réverbération et à exalter la puissance des chœurs, renforçant ainsi l'impression d'élévation vers le divin.

Avec l'arrivée de l'enregistrement sonore, la question de la spatialisation a pris une nouvelle ampleur. L'un des dispositifs les plus marquants de la fin du XIX^e siècle est le **Théâtrophone**, installé à l'Opéra Garnier. Grâce à plusieurs microphones disposés de part et d'autre de la scène, deux canaux distincts étaient transmis via des lignes téléphoniques, permettant à des auditeurs, parfois situés à plusieurs kilomètres, d'écouter un opéra à distance. Ce procédé, bien qu'encore rudimentaire, ouvrait déjà la voie à une perception spatialisée du son. Au XX^e siècle, Alan Blumlein, ingénieur britannique, a profondément marqué l'histoire de l'audio en développant le principe de la stéréophonie. À l'époque, cette approche était désignée sous le terme de **son binaural**, même si ce mot ne recouvre plus aujourd'hui la même signification : ce que l'on appelle actuellement binaural désigne une technique distincte, basée sur l'écoute au casque et la simulation des indices de localisation propres à l'oreille humaine.

À l'époque contemporaine, la spatialisation a connu une diversification considérable avec l'émergence de nouveaux formats et technologies. Les systèmes multicanaux tels que le Dolby Surround ou le Dolby Atmos permettent une immersion complète en disposant des enceintes tout autour de l'auditeur, tandis que l'ambisonie et les techniques binaurales modernes cherchent à recréer une perception tridimensionnelle du son, parfois adaptée à la réalité virtuelle et augmentée. Ces innovations témoignent de la place croissante accordée à la spatialisation dans les expériences artistiques et technologiques.

Aujourd'hui, on la retrouve dans la musique, le cinéma, ou encore les installations muséales, où la position des sources sonores enrichit et transforme l'expérience d'écoute. Elle peut être considérée comme une composante à part entière du langage artistique, au même titre que le rythme, la hauteur ou l'intensité.

2 Le SCRIME

Le stage se déroule au SCRIME (Studio de Création et Recherche en Informatique Musicale et Expérimentale).

Le SCRIME est un Groupement d'Intérêt Scientifique et Artistique (GIS) administré par le Laboratoire Bordelais de Recherche en Informatique (LaBRI) et rattaché à son département Image et Son.

Il s'agit également et principalement d'une équipe qui mène des projets de recherche scientifiques dans le domaine de la musique. Cela inclut des domaines de compétences et métiers variés comme le traitement du signal, l'informatique, mais aussi l'intelligence artificielle. Au delà de l'équipe de recherche, on y retrouve aussi des membres collaborateurs comme d'autres chercheurs, ou bien des compositeurs associés qui mènent régulièrement leurs projets en relation avec le SCRIME.

2.1 Les équipements

Naturellement, on retrouve au sein du SCRIME des outils et espaces permettant la réalisation d'expériences liées au son.

Hémicylia

Il s'agit d'une salle de conférence comprenant 25 places, munie d'un écran de projection et d'un acousmonium 19 voies. Hémicylia peut accueillir des résidences, des présentations et des concerts acousmatique et instrumentaux (1).



FIGURE 1 – Salle Hémicylia

Studio Élise

Il s'agit d'un studio spacieux et modulable pour un travail de studio en multi-canal, pour la prise de sons, la composition et le mixage. Il permet également les répétitions en petit ensemble (2).



FIGURE 2 – Studio Élise

Dôme 3D

Dans cette salle on retrouve un studio conçu pour le traitement du son 3D, composé de 25 enceintes disposées en dôme sur cinq niveaux. Le Dôme 3D permet des travaux portant sur la spatialisation, l'écoute ambisonique et la réalité virtuelle. Il est également équipé d'une Quadriphonie pour des projets de spatialisation plus modestes (3).



FIGURE 3 – Dôme 3D

2.2 Les événements

Le SCRIME prend part à l'organisation de multiples événements et projet liés à la recherche et à la création, dont on peut citer quelques exemples.

Projets en collaboration

Régulièrement, on retrouve des collaborations avec d'autres organismes dans le cadres de projets. On peut citer l'exemple de la "nuit des escaliers", organisé par l'Opéra National de Bordeaux. La participation du SCRIME concerne la transformation d'un de ces escaliers en instrument de musique "intelligent". La prochaine édition en cette année se produira le 20 septembre 2025 dans le cadre des journées du patrimoine.

Séminaires

Les activités de recherche du SCRIME incluent l'organisation et l'accueil de séminaires. On peut en retrouver un exemple sur le site internet du SCRIME sur les motifs musicaux basée sur la morphologie mathématique

Concerts

Il arrive que le SCRIME organise des concerts, en sa qualité de GIS. Plusieurs artistes effectuent des projets en collaboration avec ce dernier. Certains de ces concerts possèdent un intérêt scientifique, grâce à une expérimentation. C'est le cas du dernier en date "*Praetorian around space*".

3 Contexte du stage

L'idée même du stage étant un prolongement d'une expérience réalisée pendant un concert, ce dernier sera présenté au préalable.

Around space est un concert organisé par le groupe de métal PRAETORIAN, dans le cadre duquel a été réalisée une expérience scientifique consistant à spatialiser chaque instrument du groupe en **quadriphonie**. Cette expérience a notamment été rendue possible grâce aux résultats d'une thèse de recherche réalisée dans le cadre du SCRIME et présentée au public en amont du concert, permettant une simulation fidèle des amplificateurs à lampe habituellement présents sur scène, et éliminant ainsi dans la salle toutes les source sonores autres que le son spatialisé. À l'issue du concert les impressions du public ont été récoltées dans le but d'orienter les futures recherches.

Méthode de spatialisation

L'idée de cette expérience est de spatialiser **individuellement** chaque *canal d'instrument*. Ainsi, on pourra avoir la sensation du mouvement des sources dans l'espace tout autour du public. Cette expérience a été réalisée en direct pendant le concert de musique, en sachant que les trajectoires de spatialisation ont été capturées préalablement en studio, et cela grâce à un *périphérique Gametrak* avec 3 degrés de liberté. La spatialisation a été réalisée en se basant sur les pistes de l'album *Furialis*. L'implémentation du déplacements de ces sources sonores ont été réalisées grâce à un patch Max/MSP (4). Il s'agit d'un logiciel de programmation visuelle, spécialisé dans l'audio et créé par l'IRCAM. Son utilité est particulièrement démontrée dans la réalisation de prototypages rapides. Nous reviendrons dessus dans la section 6.1 [4]

position d'un instrument dans l'espace au fil du morceau. Pour cela, plusieurs axes de travail sont envisagés :

- Analyser une ou plusieurs pistes audio, et éventuellement leur partition, afin d'en extraire des caractéristiques pertinentes susceptibles de guider la spatialisation ;
- Concevoir un algorithme de spatialisation capable de transformer ces informations musicales en trajectoires cohérentes et expressives ;
- Explorer l'apport des méthodes d'apprentissage automatique, dans la mesure où il s'agit de comprendre et de modéliser les principes implicites de la spatialisation réalisée par un musicien, puis de restituer un résultat de qualité comparable.

Enfin, une dimension visuelle sera intégrée au projet : chaque piste instrumentale se verra associée à une couleur lumineuse, dont les déplacements dans l'espace scénique seront synchronisés avec ceux de la spatialisation sonore. L'ambition est ainsi de proposer une expérience artistique enrichie, alliant spatialisation auditive et mise en scène lumineuse.

Ce projet s'inscrit également dans la réflexion plus large abordée durant ma formation sur l'intelligence artificielle et la durabilité. L'utilisation d'algorithmes de *machine learning* implique en effet une attention particulière à la consommation de ressources, afin de limiter l'empreinte écologique des traitements. Par ailleurs, la valeur sociétale du projet réside dans la complémentarité entre automatisation et créativité humaine : l'IA ne vise pas à remplacer l'intuition musicale, mais à l'enrichir et à l'accompagner. Il s'agira d'observer le comportement humain concernant les choix musicaux, et de les reproduire, les automatiser.

État de l'art de la spatialisation sonore

La spatialisation sonore est un domaine vaste et multidisciplinaire, couvrant des aspects acoustiques, psychoacoustiques et technologiques. Un certain nombre de projets et d'outils ont été développés pour explorer et manipuler la perception de l'espace sonore, mais à notre connaissance, peu d'expériences ont cherché à spatialiser individuellement des sources instrumentales autour du public.

Parmi les projets et installations majeurs, on peut citer :

- **The Sphere Las Vegas** [3], une salle immersive conçue pour offrir une expérience audio et visuelle enveloppante.
- Concert Spatialiser, Odyssey Cercle [1], un projet explorant la spatialisation des sons dans des concerts immersifs, permettant de placer des sources sonores autour du public de manière dynamique.

Plusieurs thèses et articles scientifiques abordent la dimension spatiale en audio :

- Spatialisation sonore : approche et méthodes [?] et Étude de la spatialisation immersive dans la musique [?] discutent des méthodes de spatialisation pour la musique et l'audio immersif.
- Perception et modélisation de l'espace sonore [?] traite de la perception de la localisation sonore et des techniques associées pour l'audio 3D.

- Analyse approfondie sur la synthèse spatiale et le rendu d’ambiance sonore [?] propose une analyse détaillée sur la synthèse spatiale et les méthodes de rendu d’ambiance sonore.

Parmi les outils logiciels permettant la spatialisation :

- **Sound Particles** [2], un logiciel spécialisé dans la création d’environnements sonores immersifs et la manipulation de sources multiples dans un espace 3D.

Ces travaux et outils illustrent la richesse et la diversité des approches de la spatialisation sonore, allant des installations immersives aux logiciels professionnels. Cependant, le domaine reste ouvert, notamment pour la spatialisation précise de sources individuelles d’instruments autour d’un public, ce qui constitue l’un des axes explorés dans le cadre de ce projet.

5 Planning

5.1 Planning initial

Le stage avait été pensé en plusieurs étapes complémentaires. Il devait tout d’abord s’ouvrir sur une phase de recherche bibliographique consacrée aux méthodes de spatialisation sonore. L’objectif était de mieux comprendre les approches existantes, d’en saisir les fondements théoriques et les applications concrètes, et de repérer celles qui pourraient, à terme, nourrir le projet.

En parallèle, il était prévu de se former aux outils logiciels indispensables, en particulier **Max/MSP**, utilisé pour la manipulation des trajectoires spatiales. *Reaper* devait également intervenir, principalement comme support d’écoute et pour la séparation des canaux instrumentaux.

Une fois ces bases acquises, l’attention devait se porter sur les données fournies afin de se les approprier. Cela passait par leur **visualisation**, sous différentes formes, pour comprendre concrètement les logiques de spatialisation déjà mises en œuvre.

Ce travail technique devait s’accompagner d’une démarche continue d’**écoute musicale attentive**, destinée à analyser les choix de spatialisation réalisés. Cette approche sensible devait permettre d’identifier les passages « clés » des morceaux où la spatialisation joue un rôle déterminant.

Enfin, le projet devait culminer avec l’entraînement d’un modèle sur les données. Cette étape supposait une réflexion préalable sur les paramètres à retenir, suivie d’une mise en œuvre expérimentale et d’une évaluation à la fois objective (données produites) et subjective (perception musicale et intuitive).

5.2 Déroulement effectif

Dans les faits, le déroulement du stage est resté globalement fidèle au cadre prévu, avec quelques ajustements. La phase de documentation s’est révélée particulièrement riche et a occupé une place importante en début de parcours. Elle a permis d’approfondir la compréhension des enjeux, mais a aussi légèrement repoussé la période consacrée à l’implémentation pratique.

Du côté des outils, *Reaper* a été finalement utilisé de manière ponctuelle, essentiellement pour l’écoute et l’exploration des pistes instrumentales. Une idée s’est imposée pendant le stage, comme quoi *reaper* pourrait servir d’écriture à la spatialisation grâce à des courbes d’automation génériques. En effet, l’idée était d’écrire via *reaper* le déplacement

de l'instrument tout au long du morceau. On pourrait ensuite envoyer les informations vers Max/MSP, grâce au protocole OSC (Open Sound Control) basé sur l'envoi de paquets UDP. Dans les faits, Max/MSP s'est avéré suffisant. L'intégration de coordonnées spatiales générée grâce au *machine learning*, ainsi que son interprétation directe au sein du logiciel s'avère tout à fait satisfaisante pour l'évaluation de ces trajectoires.

La suite du stage a suivi le fil initial : visualisation des données, écoute attentive et analyse musicale ont constitué un socle continu du travail. L'entraînement de modèles a pu être entamé dans les dernières phases, cette étape nécessitant impérativement le traitement des données effectué précédemment.

6 Outils utilisés

6.1 Max/MSP

Max/MSP est un environnement de programmation visuelle largement utilisé dans le domaine de la création sonore et interactive. Il permet de concevoir des *patches* modulaires pour le traitement du signal, la synthèse et la spatialisation. Sa grande flexibilité en fait un outil privilégié pour expérimenter et effectuer des prototypes rapides. Dans le cadre de ce stage, Max/MSP joue un rôle clé en tant que moteur de spatialisation : il reçoit des données de contrôles externes, génère les trajectoires des sources et les applique en temps réel à la diffusion sonore.

Dans le cadre de mon stage, j'ai du utiliser l'interface de spatialisation 5 réalisée dans le cadre du projet *Praetorian around space*. Sa prise en main fut nécessaire afin de visualiser des trajectoires de spatialisation, tout en écoutant l'audio. L'export en binaural proposé permet également d'obtenir un résultat direct de ces trajectoires. Cette interface à l'avantage de proposer d'ajouter ses propres trajectoires rapidement. Il suffit de sélectionner un dossier, contenant les fichiers sur lesquels sont inscrits les coordonnées spatiales. Grâce à cette fonctionnalité, tester les résultats générés est une étape rapide et facilement réalisable.

L'outil utilisé pour spatialiser concrètement le son est la bibliothèque [Spat](#) développé par l'IRCAM. Intégrée à Max/MSP, elle fournit des outils spécialisés pour la spatialisation multicanal et permet de contrôler des paramètres avec son interface tels que la position, la distance ou la réverbération des sources sonores 6. Son installation est nécessaire pour faire marcher la spatialisation avec le patch mais je n'ai pas eu besoin de prendre en main la bibliothèque.

Format des données Max/MSP

Afin de pouvoir lire les coordonnées, Max/MSP requiert un format particulier. Chaque instrument spatialisé possède un fichier de séquence, dans lequel sont enregistrées les coordonnées de toutes les *tracks* 1. Il est donc nécessaire d'extraire la séquence correspondant au bon instrument et au bon morceau.

De plus, les *timecodes* présents dans ces séquences sont normalisés : ils sont exprimés en entiers relatifs par rapport à la durée totale du morceau, avec des valeurs comprises entre 0 et 1. Cette représentation nécessite une conversion systématique, car les données utilisés

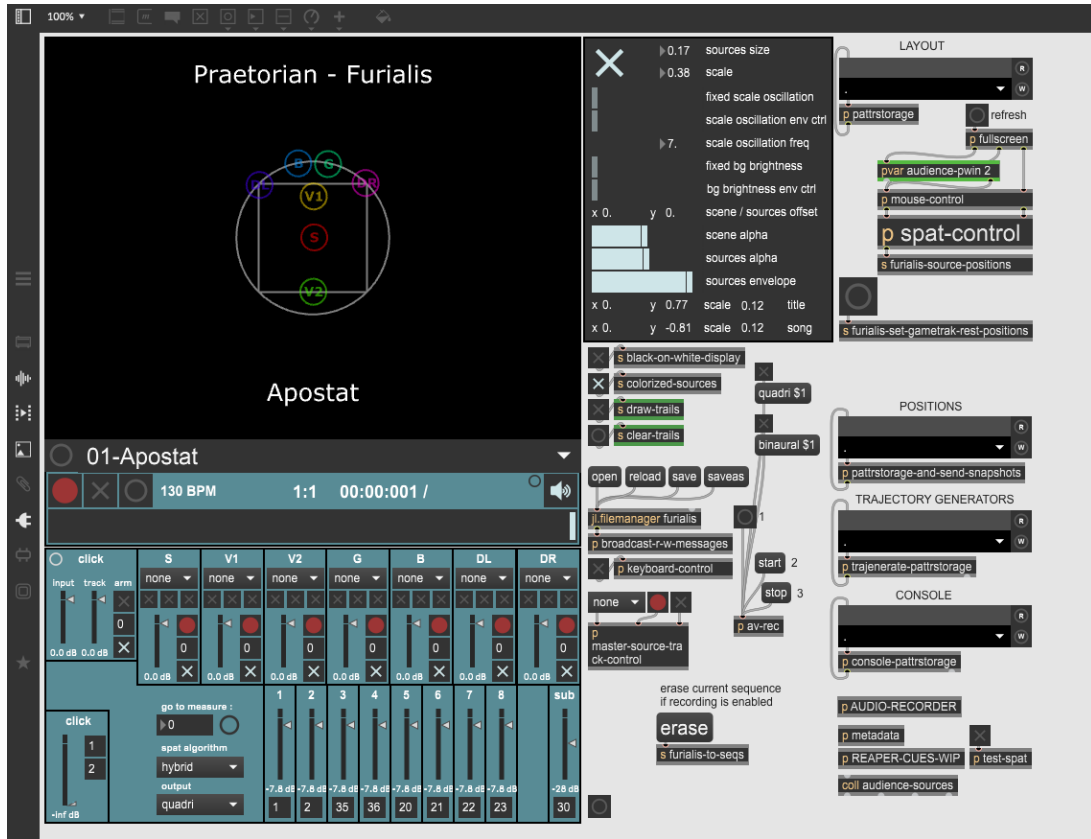


FIGURE 5 – Interface de Max/MSP

pour l'apprentissage des trajectoires par le modèle sont, eux, exprimés en échantillons par seconde, afin de disposer d'une représentation concrète du temps pour les trajectoires.

Toutes ces contraintes ont constitué un travail chronophage, qui s'est fait ressentir tout au long du stage.

6.2 Reaper

Reaper est une station audionumérique (ou *Digital Audio Workstation*) conçue pour le traitement de piste audio et *MIDI*. Le choix de Reaper se justifie simplement par sa facilité d'accès et rapidité de prise en main. Le logiciel s'avère particulièrement utile pour l'écoute analytique des morceaux de *Praetorian*. En effet, le format contenant les audios de tous les morceaux, que l'on peut appeler **contenu musical**, a été fourni sous forme de piste *multi track*. Il est donc nécessaire de décomposer l'ensemble de cette piste afin de séparer les différents *tracks*, constituant l'audio de chaque instrument. La décomposition permettra ainsi l'écoute de chacun de ces derniers individuellement 7.

7 Réalisation

Dans le cadre de ce projet, plusieurs ensembles de données ont été mis à disposition afin de permettre l'analyse et l'expérimentation. Il convient de distinguer, d'une part, les versions finales des morceaux spatialisés, et d'autre part, les données qui constituent le matériau principal pour une étude scientifique, dans le cadre d'une approche par Machine Learning.

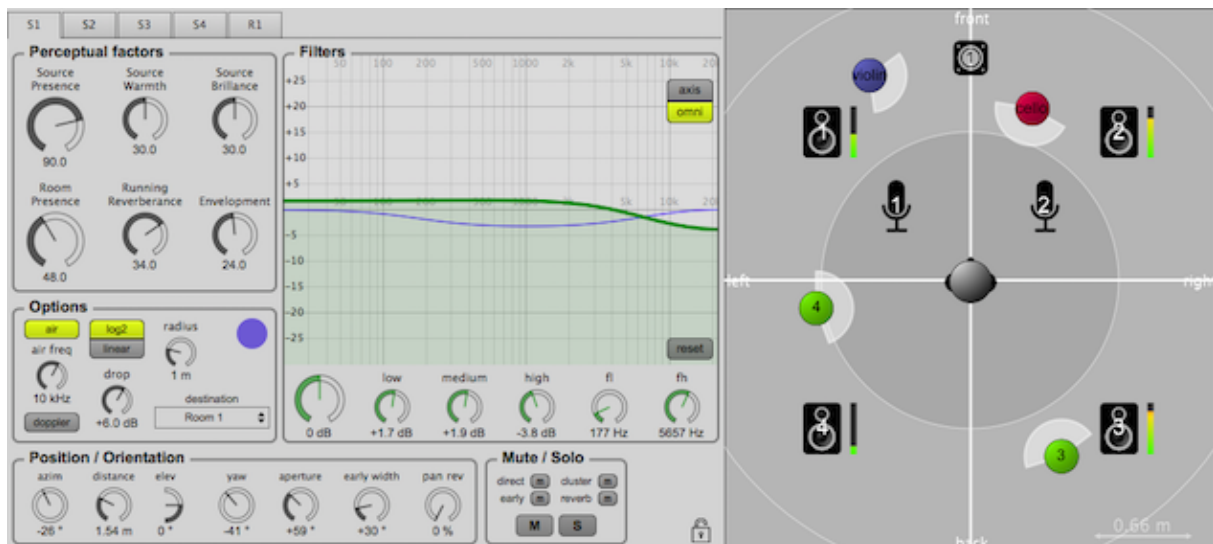


FIGURE 6 – Interface de la bibliothèque Spat

seq	Temps normalisé	x	y
0	id 08-Ballade-entre-les-mines		
1	0.000007	0.035984	-0.833232
2	0.000007	0.035984	-0.833232
3	id 05-Communion		
4	0.000020	-0.018094	-0.732866
5	0.000003	-0.018094	-0.732866
6	id 06-Face-aux-geants		
7	0.128503	0.006292	-0.703418
8	0.134527	0.006292	-0.703418
9	0.135881	0.006292	-0.711134
10	0.135881	0.014408	-0.711134
11	0.136122	0.021724	-0.711134
12	0.136481	0.029424	-0.711134
13	0.137055	0.037156	-0.711134
14	0.137327	0.044872	-0.711134
15	...	...	...

TABLE 1 – Extrait de Voc 2.txt

Les morceaux finalisés jouent avant tout un rôle de référence auditive : ils me permettent d’effectuer une écoute active et critique, en cherchant à comprendre par moi-même les mécanismes d’une spatialisation réussie. L’objectif de ces écoutes est d’identifier les éléments clés qui induisent le mouvement des sources sonores et d’analyser dans quelles conditions la spatialisation produit un effet convaincant.

Cependant, l’essentiel de la démarche repose sur les données analytiques fournies séparément. Celles-ci sont peu nombreuses, ce qui s’explique par le caractère encore récent et expérimental de l’expérience menée. Elles n’en demeurent pas moins essentielles pour adopter une approche structurée et moins intuitive, fondée sur l’analyse automatique des trajectoires et des paramètres audio grâce à des modèles de Machine Learning. Cette approche permet d’objectiver l’étude de la spatialisation et de traiter les données de manière

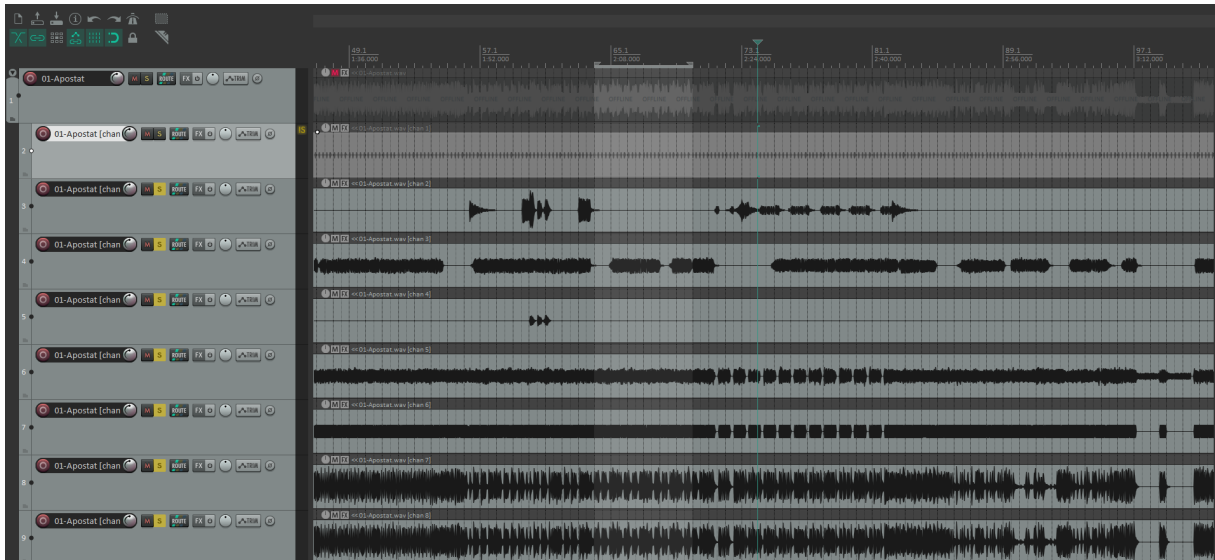


FIGURE 7 – Interface de Reaper

systématique, au-delà de l’appréciation purement subjective.

7.1 Données en entrée

Pistes audio

Neuf **pistes audio** (ou *tracks*) ont été mises à disposition pour le stage. Elles correspondent à la quasi-totalité des morceaux de l’album *Furialis*, réalisé par le groupe *Praetorian*. Chaque piste contient sept canaux distincts, représentant les différents instruments et effets. On y retrouve notamment :

- Un canal dédié au **clic**,
- Un canal dédié aux **samples**,
- Un canal pour la **voix principale**,
- Un canal pour les **chœurs**,
- Un canal pour la **guitare**,
- Un canal pour la **basse**,
- Deux canaux distincts pour la **batterie**, répartis respectivement à gauche et à droite.

Cette organisation permet d’analyser séparément le comportement de chaque instrument et d’observer plus précisément leur rôle spécifique dans la construction spatiale du morceau. On peut noter que la piste du clic ne représente pas directement les instruments, mais se révèle fort utile pour repérer le tempo 7.1.

Trajectoires des instruments

Les trajectoires de spatialisation associées à chaque instrument ont été fournies indépendamment des pistes audio. Elles sont représentées sous la forme de listes de coordonnées temporelles, permettant de suivre le déplacement d’un instrument dans l’espace tout au long du morceau. Ces données ont été générées dans **Max/MSP**, et leur format est

donc adapté à cet environnement. Chaque fichier correspond à un instrument particulier, et contient en réalité l'ensemble des coordonnées de tous les morceaux. Afin de distinguer des derniers les uns des autres, chaque séquence est introduite par une ligne précisant la piste concernée (voir table 1).

Marqueurs

En complément des trajectoires, des marqueurs ont été fournis pour segmenter les morceaux en différentes régions musicales. Définis directement par le groupe, ces marqueurs correspondent aux parties logiques de la structure d'un titre (couplets, refrains, ponts, breaks, etc. voir table 2). Ils constituent des repères temporels précieux pour l'analyse, car ils coïncident souvent avec des variations d'intensité, de dynamique ou de texture sonore. Ces changements, par leur impact perceptif, influencent directement la manière dont la spatialisation est perçue et justifient une étude approfondie de leur rôle.

#	Name	Start
M1	Start	7.126
M2	Punch	7.359
M3	Punch	10.592
M4	Punch	11.059
M5	Punch	14.292
M6	Punch	14.759
M7	Punch	17.992
M8	Punch	18.459
M9	Build Up	20.292
M10	Break	21.659
M11	Intro	22.126
M12	Break / Pre Verse	36.892
...	...	...

TABLE 2 – Découpage temporel des sections

Tempo

Enfin, le tempo (battements par minute ou BPM) a également été communiqué avec les pistes. Bien que les morceaux de l'album *Furialis* présente un tempo constant et qu'il aurait donc été possible d'estimer le tempo par une simple analyse rythmique de la piste de clic, la disponibilité directe de cette information simplifie considérablement l'alignement temporel des données. Le BPM constitue une donnée essentielle pour la synchronisation dans l'analyse des trajectoires, mais aussi pour indiquer un changement au sein de la musique. Cette donnée s'avère ainsi pertinente pour l'entraînement d'un modèle.

7.2 Analyse et Visualisation

Un aspect central du stage concerne la mise en place d'outils de **visualisation** des spatialisations déjà réalisées. La représentation graphique des trajectoires constitue en effet un élément plus que nécessaire pour comprendre les mécaniques de trajectoire d'instrument.

Elle permet non seulement d'illustrer de manière claire et accessible les déplacements effectués par chaque instrument, mais également d'identifier les déclencheurs et paramètres qui influence leur évolution dans le temps.

La visualisation a ainsi une double finalité : d'une part, elle facilite l'analyse des morceaux en offrant un support objectif pour l'étude des données, et d'autre part, elle joue un rôle démonstratif essentiel en rendant perceptibles des phénomènes qui demeurent souvent abstraits à la seule écoute. Elle constitue donc un outil incontournable pour à la fois **comprendre**, **expliquer** et **communiquer** les mécanismes à l'origine de la spatialisation.

Nous pouvons décomposer l'analyse en plusieurs parties.

7.2.1 Analyse de l'audio

Plusieurs options peuvent être envisageables pour analyser une piste audio. L'enjeu ici est de trouver les bons **descripteurs audios**, ceux qui sont cohérents avec ce que l'on cherche. Dans notre cas, on s'intéresse premièrement à un seul concept : l'instrument joue-t-il ? Avec cette information pour chaque instrument sur une même ligne temporelle, on peut afficher leurs différentes combinaisons d'état (joue ou ne joue pas). Ainsi on pourra faire un lien avec la spatialisation.

Le choix évident pour cette tâche est l'analyse de RMS (Root Mean Square Value), ou moyenne quadratique. L'idée est la suivante : l'audio est une liste de valeurs, dont la taille s'élève à la fréquence d'échantillonnage multipliée par le nombre de secondes. Cela représente un grand nombre de valeurs (10^7 pour un morceau de 4 minutes). Le RMS qui constitue l'intensité du signal segmente ces valeurs, et pour chaque segment, réalise la moyenne quadratique $\sqrt{\frac{1}{N} \sum_{n=1}^N x[n]^2}$ avec N étant le nombre de valeur au sein de la fenêtre de temps calculée et x une famille de valeur. Il suffit de choisir la taille des segments pour obtenir une suite de valeur décrivant l'intensité du signal. Dans notre cas, on fera avec 10 ms. Les valeurs RMS étant très variables, on fera le choix de les normaliser de 0 à 10. Pour cela on repère la valeur RMS maximum, et pour chaque valeur on effectue sa division par la valeur max, que l'on multiplie à l'échelle (10 dans notre cas).

C'est à partir de ce signal, que l'on va pouvoir déterminer si l'instrument joue ou non. Cela se traduit par transformer les valeurs entières (normalisées) en valeurs booléennes. A nouveau, plusieurs paramètres sont nécessaires pour cela. On considérera qu'un instrument joue si la valeur RMS est au dessus du seuil choisi (ici 1 après normalisation), **et** que l'instrument joue plus de 2 secondes au dessus de cette intensité. Ces valeurs ont été choisies de manière arbitraire et ne constitue pas une vérité fixe.

Afin d'avoir des points de références sur la visualisation, on rajoute aussi les marqueurs temporels. (8) (??)

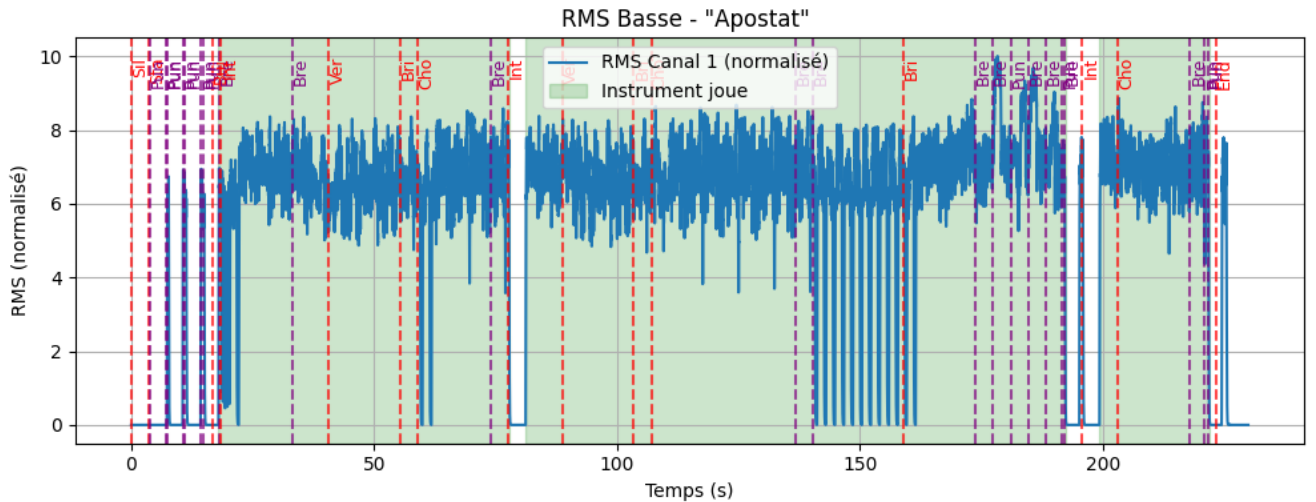


FIGURE 8 – RMS de la basse pour le morceaux "Apostat"

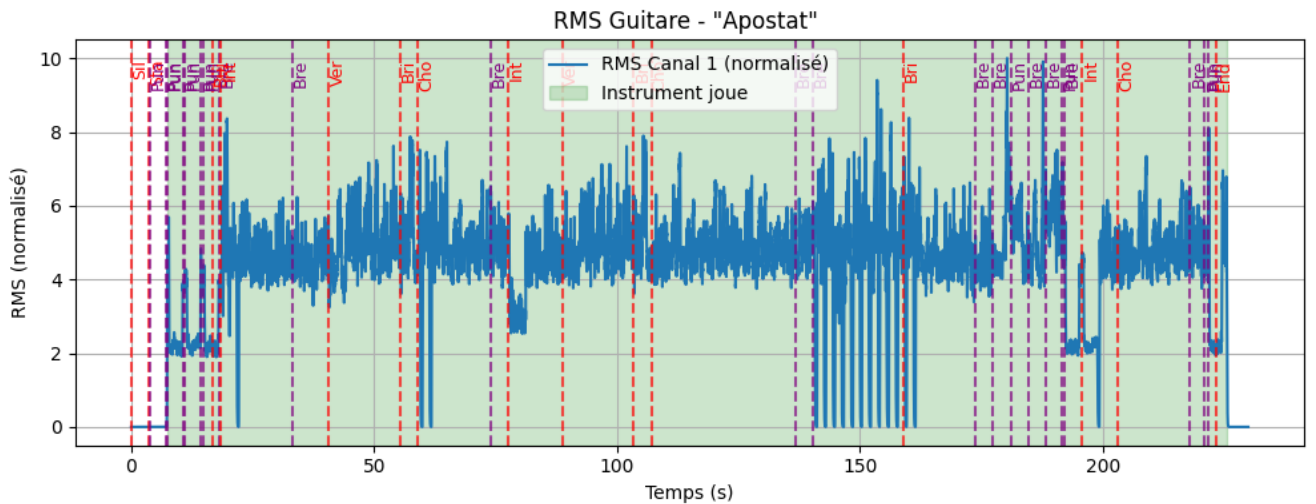


FIGURE 9 – RMS de la guitare pour le morceaux "Apostat"

On observe d’ores et déjà des corrélations apparentes entre certains marqueurs et des variations significatives du RMS. C’est notamment le cas pour la piste de basse, dont l’énergie fluctue de manière marquée au niveau des différentes introductions (*int*) ainsi que lors de certains breaks (*bre*).

On peut également observer l'exemple du canal de sample. Son rôle dans les morceaux est d'appuyer des passages fort en intensité, afin d'en augmenter le ressenti. Cet appui se produit généralement en plein milieu d'une région audio. 10)

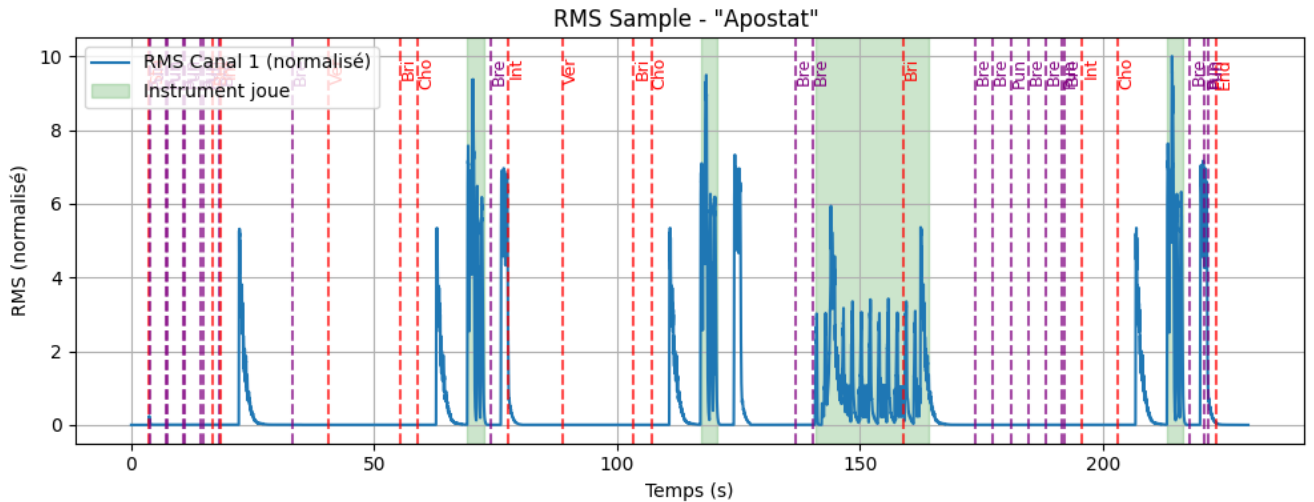


FIGURE 10 – RMS du sample pour le morceaux "Apostat"

7.2.2 Analyse de la spatialisation

Le terme de *spatialisation* recouvrant un champ large, il convient d'en préciser plusieurs aspects que l'on retiendra comme particulièrement pertinents dans le cadre de cette étude. D'abord, il faut s'intéresser au terrain de jeu de la spatialisation réalisée, afin de mieux comprendre les déplacements des instruments. Concernant le déplacement de ces derniers, différents axes d'analyse peuvent être dégagés. D'abord la succession des positions occupées au cours du morceau, qui permet de restituer la trajectoire suivie. Par la suite, la vitesse de déplacement, qui renseigne sur la dynamique et l'intensité du mouvement dans l'espace sonore. Enfin, la position moyenne ou globale de chaque instrument sur l'ensemble de la pièce afin d'observer la présence globale d'un instrument sur tout le morceau.

Terrain

Concernant le terrain de jeu de spatialisation, il s’agit d’un cercle unitaire. La raison à cela est du au choix de la **quadriphonie** comme type de diffusion. Quatres hauts parleurs sont placés dans les coins, ce qui permet de diffuser la même intensité partout autour du public. Ce cercle entoure un carré qui représente la scène, endroit dans lequel le public est placé. La source d’un instrument peut théoriquement dépasser le cadre du cercle unitaire, dans ce cas la spatialisation simule, grâce à l’utilisation d’indices acoustiques tels que l’atténuation des hautes fréquences et la réverbération, un éloignement encore plus grand.

(11) [4]

Positions

L’affichage des positions pour un instrument donné permettra de représenter son déplacement, tel qu’il est défini par le format de spatialisation de *Max/MSP*. Dans ce format, les points ne sont enregistrés et affichés que lorsqu’un mouvement est effectivement présent, ce qui explique que la visualisation corresponde principalement aux phases de

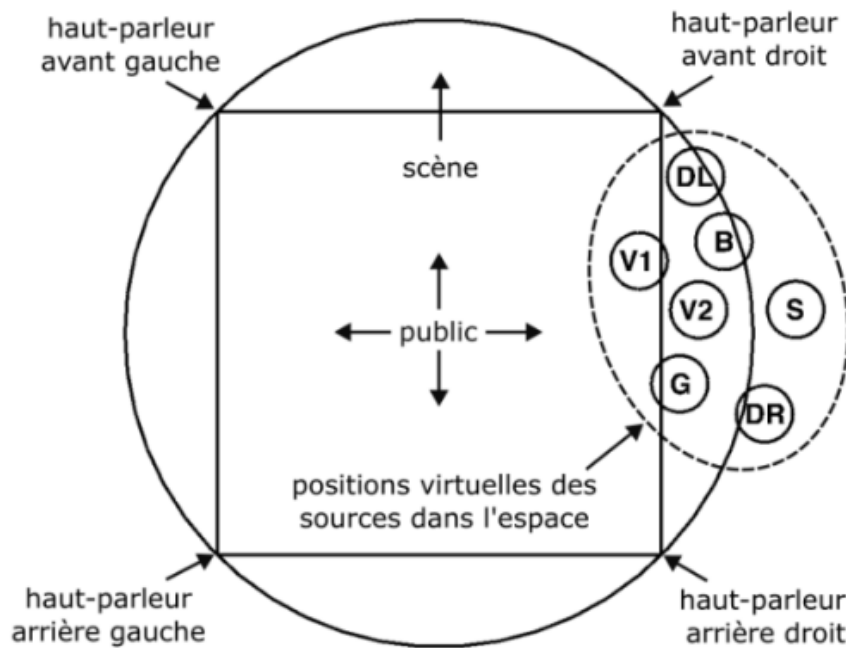


FIGURE 11 – Description de l’interface graphique représentant l’espace d’écoute et les positions virtuelles des sources sonores.

déplacement. Ces points successifs permettent ainsi de reconstituer la trajectoire globale de l’instrument. Afin d’éviter une surcharge visuelle liée à l’accumulation d’informations, la représentation sera segmentée par régions, ce qui en facilite la lecture et l’interprétation (12). Pour cela, on utilise les marqueurs fournis au préalable (voir table 2).

Dans l’exemple de la figure 9, les instruments ont la caractéristique de recevoir la même spatialisation mais en symétrie. L’origine qui représente le centre de la scène est situé aux coordonnées (0,0). On distingue un motif circulaire, répété à de nombreuses reprises au centre, donnant lieu à une forte concentration de points. À l’inverse, un mouvement plus ample n’apparaît qu’une seule fois. L’espacement relativement réduit entre les points de cette grande trajectoire traduit une vitesse d’exécution assez lente.

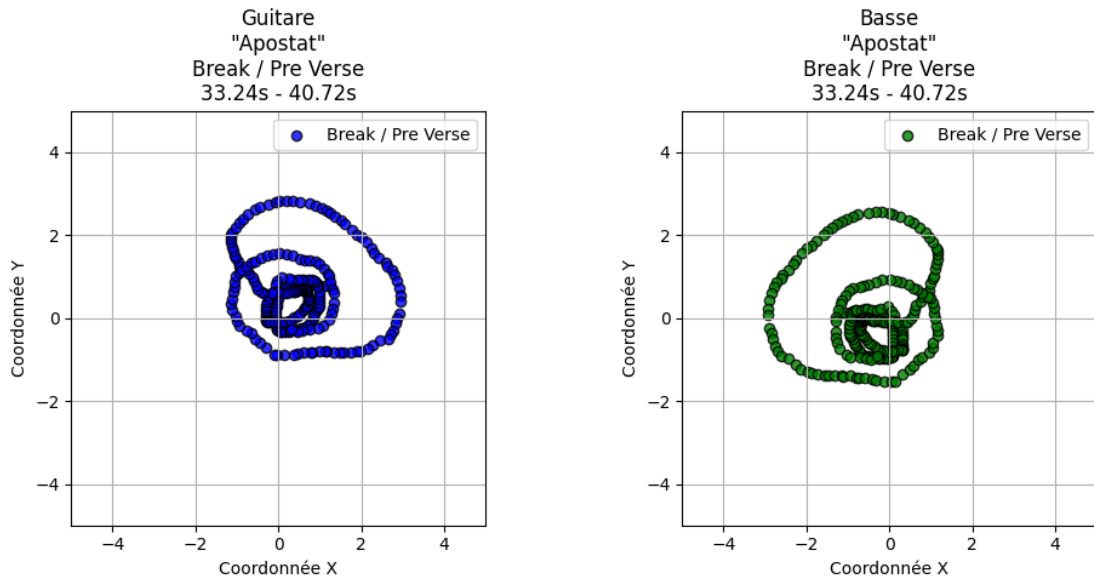


FIGURE 12 – Positions de la basse et de la guitare lors du Pre Verse

En comparaison, l'exemple de la Figure 10 illustre un déplacement particulièrement lent, rendu visible par le très faible espacement entre les points. La densité accrue, traduite par une teinte plus sombre, témoigne de cette concentration. (13)

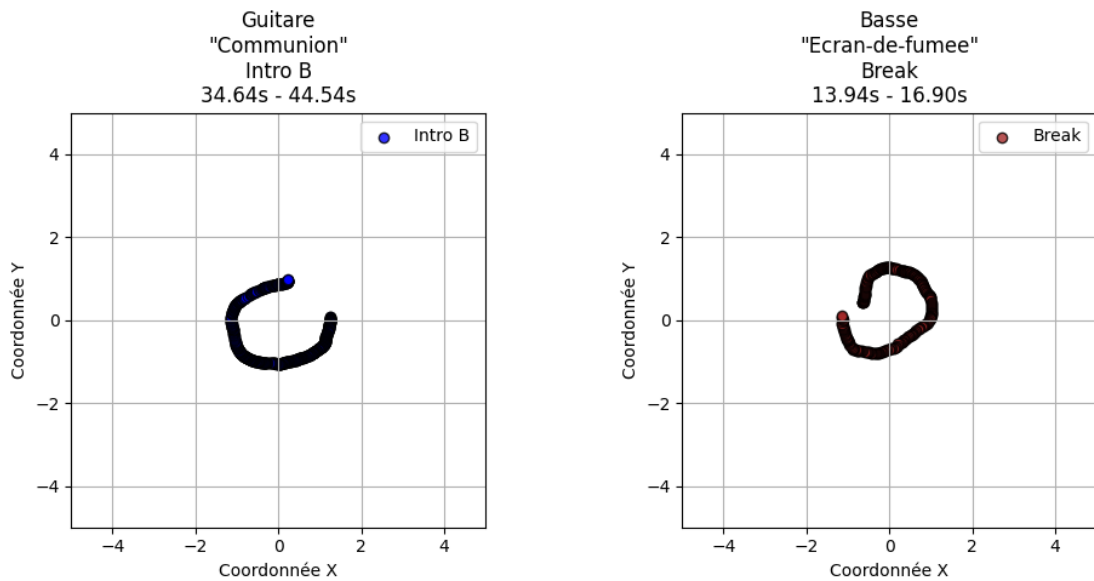


FIGURE 13 – Déplacement lente de la guitare dans "Communion" et de la basse dans "Ecran de fumée"

À l'inverse, comme illustré en Figure 11, des mouvements beaucoup plus rapides se manifestent par des points nettement plus espacés, indiquant une progression plus vive dans la trajectoire. (14)

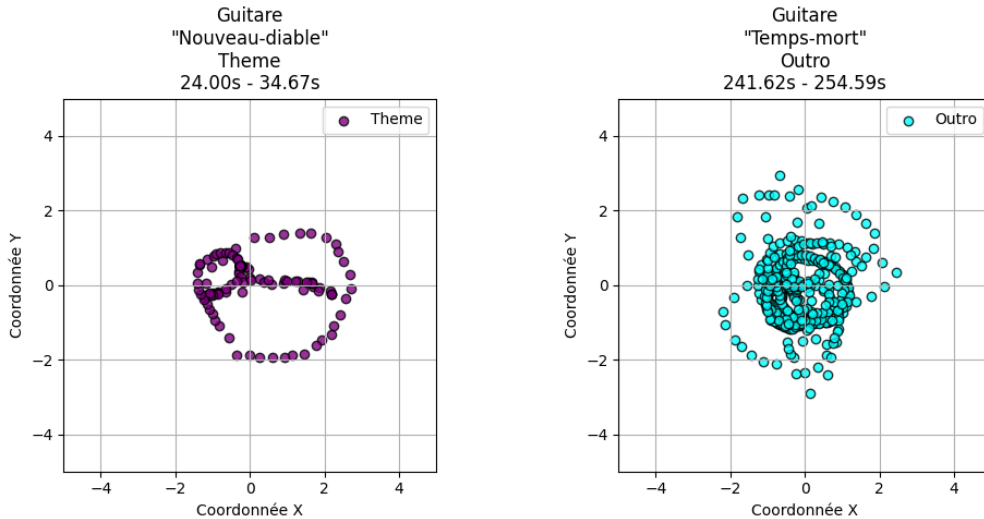


FIGURE 14 – Point de position pour la guitare dans "Nouveau diable et "temps-mort"

Vitesse

La vitesse de déplacement d'un instrument constitue un indicateur pertinent de l'intensité de la spatialisation au cours du morceau. Elle peut être représentée sous forme de diagramme et mise en relation avec les différentes régions audio, ce qui permet d'établir des parallèles entre les variations de mouvement et la structuration musicale. (15) (16)

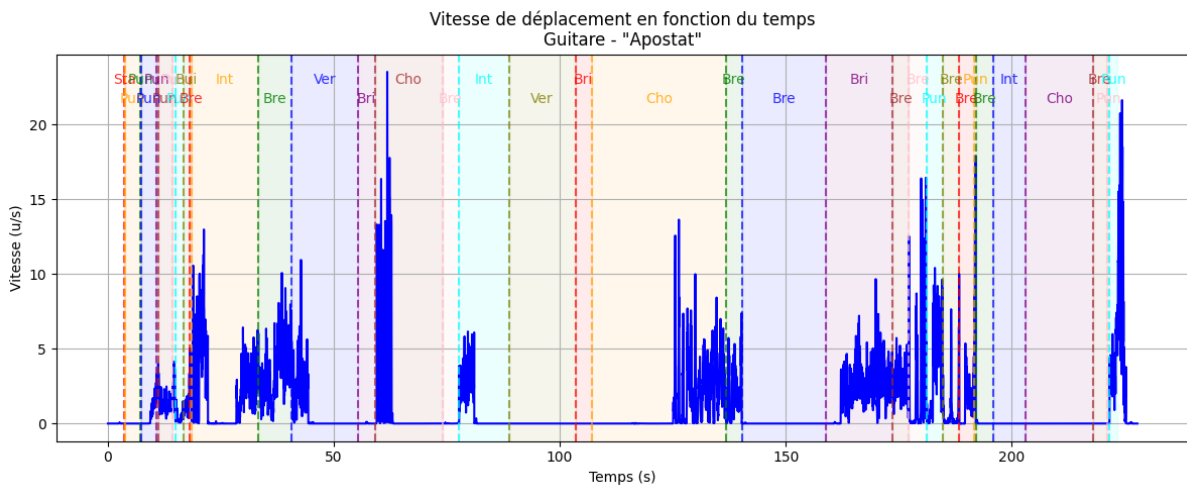


FIGURE 15 – Vitesse de spatialisation pour la guitare sur le morceau "Apostat"

Ces deux exemples des figures 12 et 13 offrent une vision d'ensemble de ce que l'on pourrait qualifier de "fréquence d'activation de la spatialisation", ainsi que de son intensité. On observe ici une corrélation notable avec les régions du morceau, comparable à celle déjà mise en évidence avec les valeurs de RMS. Par ailleurs les spatialisations proposées sont en fait parfaitement coordonnées dans cet exemple entre la basse et la guitare (dans le cas ou elle n'est pas inexistante). C'est un phénomène que l'on retrouve fréquemment parmi tous les morceaux. Cela est en fait cohérent puisque ces derniers possèdent une structure précise à laquelle tous les instruments se réfèrent (couplet, refrain etc.). Dans les faits, la plupart des exemples montrent la cohérence de déplacement des sources tout au long des

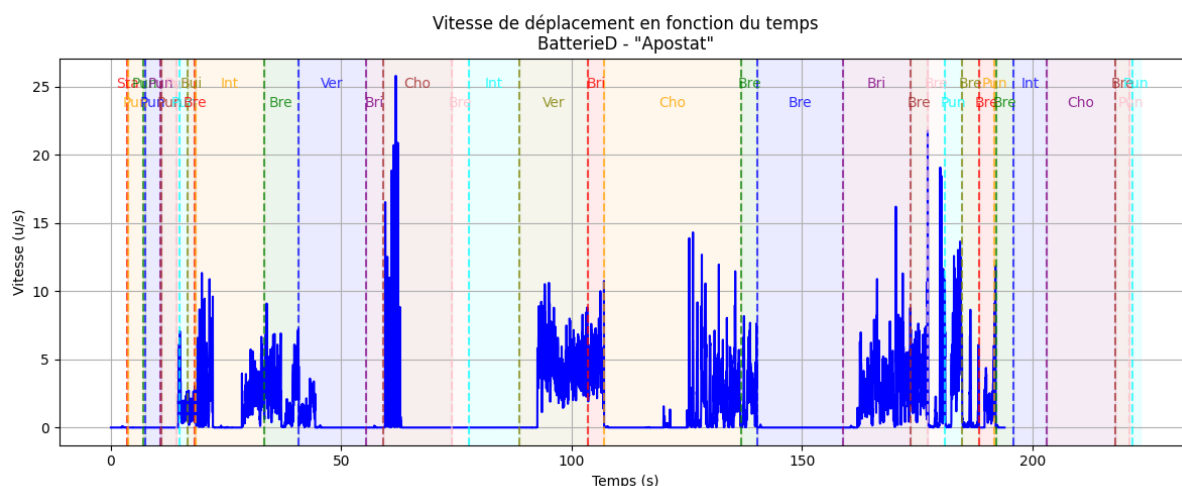


FIGURE 16 – Vitesse de spatialisation pour le canal droit de la batterie sur le morceau "Apostat"

morceaux. Il arrive toutefois que certains instruments soient spatialisés individuellement durant un laps de temps.

Heatmap

Le 3eme axe intéressant concerne la répartition de la position des instruments tout au long du morceaux. Pour représenter cela, on peut utiliser une ou carte de chaleur. Pour la réaliser, on choisit de sélectionner l'ensemble des instruments, afin de connaître les endroits plus ou moins occupés sonoremment. L'idée est d'observer si des zones sont plus favorisées que d'autre tout au long du morceau de manière générale.

Pour obtenir ces données de positions, on récupère les positions originelles. Ces dernières étant présentes uniquement lorsque l'instrument effectue un déplacement, il nous faudra **ré-échantillonner** afin d'indiquer si ce dernier *stationne* longtemps à un endroit. Les données utilisées pour le calcul de la densité proviennent uniquement des positions ré-échantillonnées. Chaque instrument est associé à une liste de coordonnées, ajustée de sorte que toutes ces listes possèdent la même longueur. Ce procédé garantit que l'évolution temporelle des positions reste comparable entre instruments, quelle que soit la fréquence réelle de leurs déplacements. En combinant l'ensemble de ces listes de coordonnées, on obtient la heatmap, qui reflète les zones de présence les plus fréquentes. Le repère spatial utilisé correspond directement à celui des positions initiales. Cette discrétisation a été choisie en raison de la forme de la scène afin de pouvoir situer concrètement la position "favorite" l'ensemble sonore formé par les sources instrumentales. (17) (18)

On distingue plus ou moins nettement la forme carrée de la scène selon les morceaux analysés. Cette forme de heatmap correspond au schéma de l'espace occupé par le public, dit "espace d'écoute". On observe que la répartition spatiale se concentre majoritairement sur la zone située à l'avant du public, à proximité de la scène. Ce constat renvoie directement à une problématique de cohérence entre le rendu sonore et l'expérience visuelle. En effet, lors de l'expérience menée précédemment, plusieurs retours faisaient état d'une gêne perceptive liée au décalage entre la position réelle des instrumentistes sur scène et la localisation virtuelle du son, parfois projeté derrière le public. Une telle dissociation entre le visuel et l'auditif tend à perturber le cerveau, qui recherche naturellement une

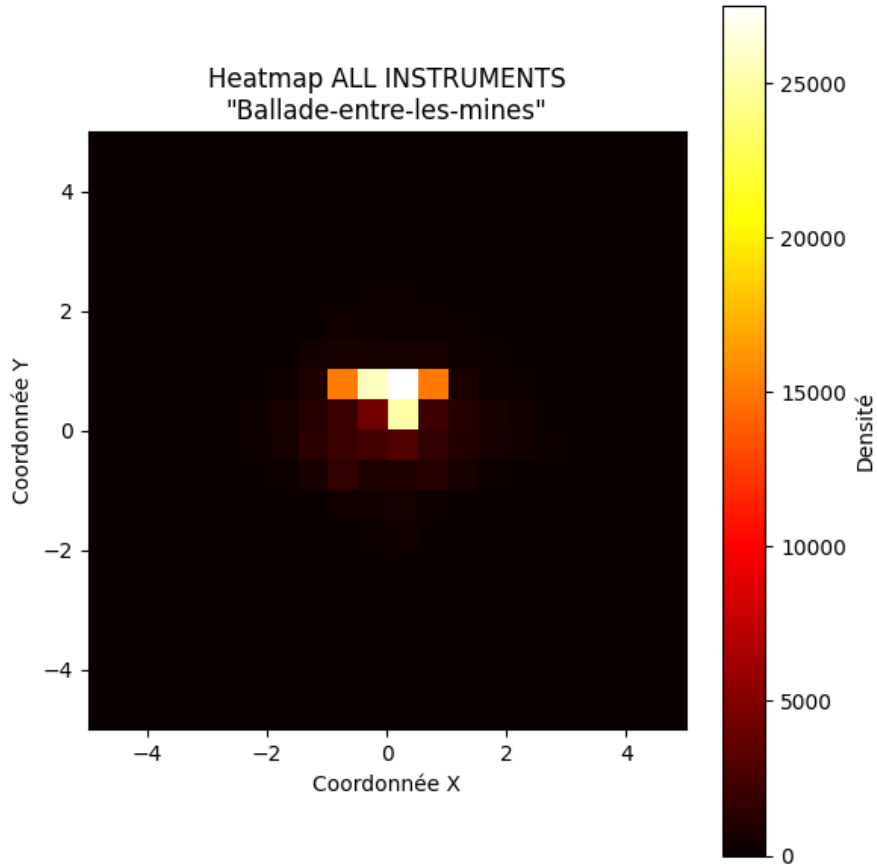


FIGURE 17 – *Heatmap* pour le morceau "Ballade entre les mines"

correspondance entre ce qu'il voit et ce qu'il entend.

Ainsi, dans la plupart des cas, les sons ont été projetés préférentiellement vers l'avant, traduisant une tendance instinctive à placer les sources virtuelles à l'endroit où se trouvent effectivement les musiciens. Certains morceaux se distinguent toutefois de cette logique, comme *Écran de fumée*, qui adopte une organisation bien plus symétrique dans la répartition des positions spatiales. (19)

7.2.3 Interprétation et intuition

L'analyse montre que la spatialisation peut être envisagée comme un véritable paramètre musical, au même titre que l'intensité, la hauteur, le rythme ou le timbre. Elle ne se limite pas à une technique de diffusion, mais joue un rôle actif dans l'écriture et dans l'expression. Chaque morceau observé présente ainsi une organisation spatiale spécifique, qui peut être assimilée à un véritable "style de spatialisation". Ce constat implique que tout travail d'entraînement d'un modèle ou de modélisation devra prendre en compte cette diversité, et qu'il n'existe pas qu'une seule manière "correcte" de spatialiser.

On constate aussi que la spatialisation est étroitement liée à la structure des morceaux. Les passages où la musique change de fonction — couplet, refrain, break — s'accompagnent souvent d'une évolution claire de la distribution spatiale. L'espace devient ainsi

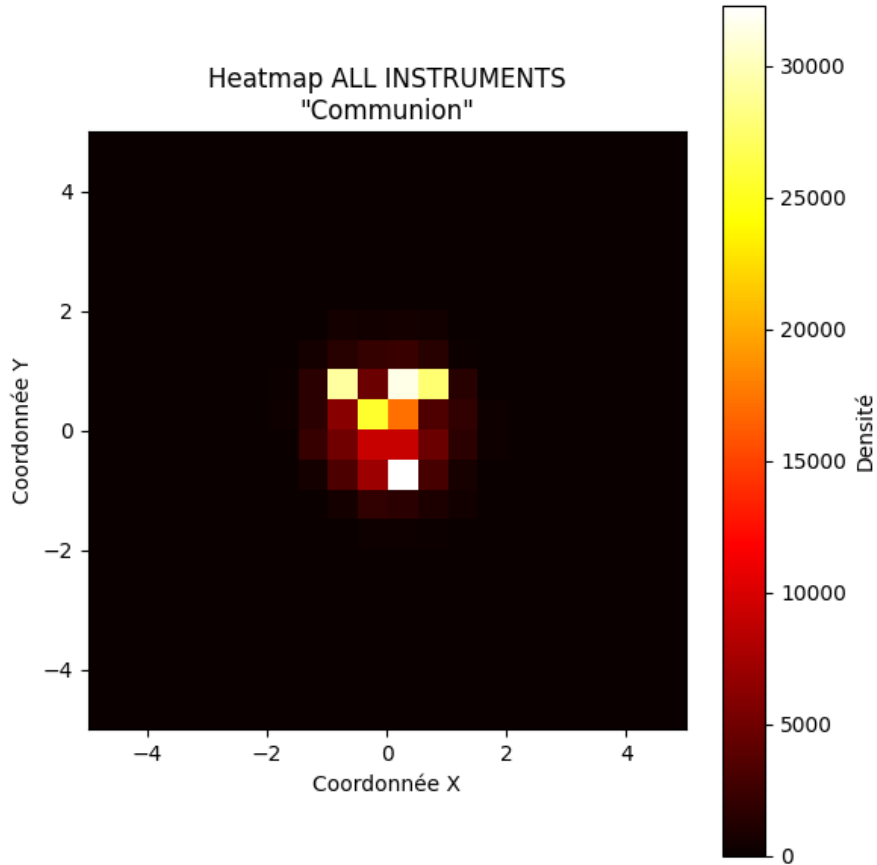


FIGURE 18 – *Heatmap* pour le morceau "Communion"

une extension de la forme musicale, qui accentue ou prolonge son effet.

L'expérience d'écoute souligne un autre point essentiel : la perception d'une spatialisation n'est pas la même selon qu'on écoute au casque ou en concert. Le lien entre ce que l'on voit (la position des instrumentistes) et ce que l'on entend (la projection sonore) est décisif dans le ressenti. Pour obtenir un rendu satisfaisant, il paraît donc nécessaire de confronter les analyses aux deux contextes d'écoute, afin de garder une cohérence entre l'intention artistique et la perception réelle.

7.3 Entraînement d'un modèle

L'objectif final consiste à exploiter l'ensemble des données précédemment décrites afin d'entraîner un modèle de **Machine Learning** capable de reproduire une spatialisation. La première étape de ce processus repose sur la constitution d'un jeu de données adapté, regroupant les paramètres jugés pertinents. Une fois ce dataset établi, il s'agira de sélectionner une architecture de modèle appropriée, puis d'évaluer les résultats obtenus.

7.3.1 Sélection du jeu de donnée

Il est essentiel de s'interroger sur les paramètres à retenir pour l'entraînement, car ce choix conditionne directement la qualité et la pertinence du modèle. Une première évidence consiste à utiliser les données déjà exploitées et visualisées précédemment, puisqu'elles constituent une base solide et directement liée à la spatialisation. Cependant, il semble

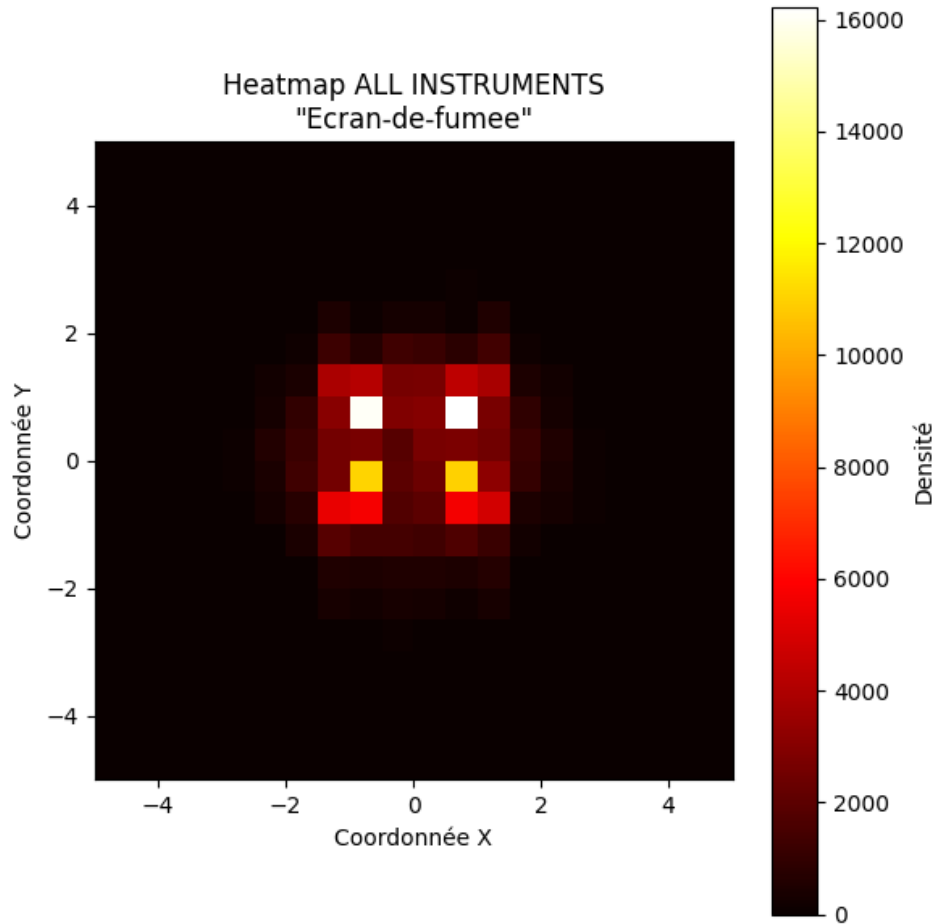


FIGURE 19 – *Heatmap* pour le morceau "Écran de fumée"

également pertinent d'élargir cette sélection à d'autres variables qui, selon notre expérience et notre analyse des morceaux, exercent une influence notable sur le rendu spatial.

Données existantes

Le jeu de données réalisé repose dans un premier temps sur les informations déjà disponibles et jugées pertinentes pour l'entraînement du modèle. On retrouve notamment :

Les **régions**, permettant de segmenter le morceau en différentes parties logiques (intro, couplet, refrain, break, etc.).

Les **RMS**, représentés sous la forme d'un vecteur binaire (0 ou 1), indiquant si l'instrument est en train de jouer ou non.

Les **coordonnées spatiales** (x, y), qui constituent les valeurs cibles de l'entraînement, correspondant directement aux trajectoires instrumentales.

Données complémentaires

À ces données de base, on peut donc ajouter des informations complémentaires, jugées pertinentes pour enrichir le dataset. Elles concernent en particulier la dimension temporelle du morceau, afin de mieux rendre compte de son évolution et de ses repères structurels.

La **valeur de temps** en cours, permettant de relier chaque point de la spatialisation à un repère temporel précis.

La **mesure musicale** correspondante, qui structure le morceau et fournit un cadre rythmique clair. Dans les musiques de *praetorian*, une mesure à une signature rythmique de 4/4 ce qui signifie qu'il y a 4 temps dans une mesure.

L'avantage des morceaux étudiés (issus de *Praetorian*) est qu'ils possèdent tous un **tempo constant**. Cette caractéristique simplifie considérablement l'alignement temporel. En effet, le *timecode* est calibré dès le début grâce à la pulsation des clics (deux clics d'intro avant l'entrée des instruments). Il est ainsi aisé de faire correspondre chaque événement à son temps exact. Concernant les mesures, la structure régulière en 4/4 permet d'automatiser le repérage : un changement de mesure intervient systématiquement tous les quatre temps.

Normalisation temporelle

Afin de rendre les données comparables et exploitables par le modèle, il est nécessaire de procéder à un **ré-échantillonnage**. Concrètement, le dataset est découpé en trames régulières, de sorte que chaque ligne du tableau de données corresponde à un instant précis de **20 ms**. À chaque pas temporel, l'ensemble des informations est mis à jour : coordonnées spatiales (x, y), RMS, régions et repères rythmiques. Cette approche permet d'obtenir une *normalisation temporelle*, en plaçant toutes les données sur une grille commune, ce qui simplifie l'analyse et garantit une cohérence dans le traitement par le modèle. (20)

7.3.2 Selection du modèle

Il nous faut maintenant choisir un modèle pour entraîner les données détaillées précédemment.

Au tout début de nos essais, nous avons exploré des approches simples basées sur des modèles linéaires de la bibliothèque Scikit-learn. Bien que ces derniers permettent d'obtenir une première approximation, ils se révèlent rapidement insuffisants pour capter la dynamique réelle de nos données, notamment en raison de l'absence de prise en compte de la dimension temporelle. Un aspect caractéristique de nos données est en effet la notion du **temps**. Comme pour toutes les données audios, la notion temporelle joue un rôle déterminant dans les résultats et ne peut être négligé. Dans ce cadre là, il est donc intéressant pour nous de se pencher sur un modèle utilisant des **chaines de Markov** et plus particulièrement vers des **modèles de Markov cachés** (HMM, Hidden Markov Models). Un HMM est un modèle probabiliste génératif utilisé pour représenter des séquences temporelles. L'idée est qu'une suite d'états cachés, non observables directement, évolue selon une chaîne de Markov et génère une suite de variables observées. Les probabilités de transitions entre états décrivent la dynamique temporelle, tandis que les lois d'émission définissent la distribution des observations en fonction de l'état caché.

Bibliothèque utilisée

Afin de prendre un modèle simple qui répond à nos attentes, j'ai choisi de me tourner vers **hmmlearn**. Il s'agit d'une librairie qui implémente des modèles de markov cachés

```

time,rms,region,temps,mesures,x,y
0.00,0,Silence,1,1,0.199425,0.965794
0.02,0,Silence,1,1,0.199425,0.965794
0.04,0,Silence,1,1,0.199425,0.965794
0.06,0,Silence,1,1,0.199425,0.965794
0.08,0,Silence,1,1,0.199425,0.965794
0.10,0,Silence,1,1,0.199425,0.965794
0.12,0,Silence,1,1,0.199425,0.965794
0.14,0,Silence,1,1,0.199425,0.965794
0.16,0,Silence,1,1,0.199425,0.965794
0.18,0,Silence,1,1,0.199425,0.965794
0.20,0,Silence,1,1,0.199425,0.965794
0.22,0,Silence,1,1,0.199425,0.965794
0.24,0,Silence,1,1,0.199425,0.965794
0.26,0,Silence,1,1,0.199425,0.965794
0.28,0,Silence,1,1,0.199425,0.965794
0.30,0,Silence,2,1,0.199425,0.965794
0.32,0,Silence,2,1,0.199425,0.965794
0.34,0,Silence,2,1,0.199425,0.965794

```

FIGURE 20 – Exemple de dataset avec un temps normalisé

(Hidden Markov Model ou HMM).

Modèle utilisé

Pour implementer un modèle de markov, je me suis tourné vers une solution simple dans un premier temps, en utilisant le *GaussianHMM*. Ce dernier est un HMM où les observations sont des nombres réels modélisés par des lois normales. Il est très utilisé quand on observe des données continues comme des températures, des signaux, de l'audio, ou dans notre cas des **trajectoires**.

Paramètres du modèle

Le GaussianHMM de *hmmlearn* dispose de plusieurs paramètres qui contribuent au résultat. Parmi ces derniers, il y a ceux qui définissent le modèle, en pouvant être saisi manuellement ou, entraîné à partir de données existante. les paramètres **startProb_** et **transmat_** définissent la chaîne de markov contenu dans le modèle. Le premier définit la probabilité initiale de commencer dans chaque état caché, le second constitue la matrice de transition entre états. Les deux autres paramètres **means_** et **covars_** définissent plutôt la distribution gaussienne. Le premier étant la moyenne de la gaussienne associée à chaque état et le deuxième la matrice de covariance associée à chaque gaussienne.

Ces paramètres étant entraînés automatiquement lors d'un `model.fit()`, nous ne les expliciterons pas ici.

- **n_components** : définit le nombre d'états cachés du modèles. Plus ce nombre est grand, plus le modèle peut capturer des comportements variés, mais au risque de sur-apprentissage.
- **covariance_type** : spécifie la forme de la matrice de covariance des gaussiennes. Différentes options sont possibles (*full*, *diag*, *spherical*, *tied*), permettant de s'adapter à la complexité des données.
- **n_iter** : nombre maximal d'itérations de de l'algorithme d'optimisation.
- **random_state** : permet de fixer la graine aléatoire utilisée pour l'initialisation. Ce paramètre est important pour assurer la reproductibilité des résultats.

Ainsi, ces paramètres contrôlent à la fois la complexité du modèle (nombre d'états et type de covariance) et la stabilité de l'apprentissage (nombre d'itérations, initialisation). A noté que dans notre cas, nous ne faisons varier majoritairement que le nombre d'état caché, en raison sa grande influence sur les premiers résultats. Le reste des paramètres variables ont été choisis de manière arbitraire et ne sont pas modifiés au cours des expériences. Il en convient que ce sera probablement le cas à l'avenir afin d'affiner les résultats.

8 Résultats

Une des idées principales était de repérer des passages dans les différents morceaux de furialis qui sont caractéristiques d'une spatialisation intuitive. L'objectif est de faire apprendre au modèle un comportement le poussant à spatialiser de manière *logique* pour notre cerveaux. Plusieurs aspect peuvent définir cette notion de logique, la plus évidente, à mon sens, est celle du rythme. Il est facilement remarquable visuellement et auditivement de percevoir l'instruments se déplacer en même temps que les **temps** qui sont marqués. D'autre notion pouvant être moins évidente serait le changement de rythme de la batterie ou encore l'arrêt de jeu d'un instrument comme la basse ou la guitare. C'est dans cet enjeux qu'une approche de machine learning prend tout son sens.

8.1 Entraînement du modele

Afin de pouvoir entrainer le model, la première tache à réaliser au préalable est la selection d'extrait musical ou la spatialisation à un role majeur selon moi. Entraîner des extraits de manière précise pourra nous apporter des premières **indications de comportement** du model. En effet, dans le contexte de ce rapport, il ne sera possible de présenter que les débuts des résultats obtenus, qui constitueront donc des approches et non des réelles trajectoires dessinées.

8.1.1 Selection d'extrait musical

Une partie de mon temps était consacrée uniquement à l'écoute active des morceaux de Praetorian sur Max MSP, afin d'en relever les extraits intéressants pour la spatialisation.

(3)

Tous ces extraits se révèlent intéressants à étudier, et il peuvent constituer à mon sens, de bon support pour introduire des règles de spatialisation pour le modèle. Cet échantillon

Instrument	Timecode	Description de la spatialisation	Origine probable de la spatialisation
Basse guitare	2'05 – 2'20	Écartement des sources (centre puis droite ou gauche)	Changement rythme Batterie + Le chant s'arrête
Batterie	2'57 – 3'01	Trajectoire ligne triangle	Temps fort guitare puis gros appui sur chaque syllabe : <u>CO</u> <u>MU</u> <u>NAU</u> <u>TE</u>
Écran de Fumée (Instruments)	1'21 – 1'48	Échange de position entre Batterie et Basse / Guitare	Changement de région
Sample	début – 1'31	Se déplace partout sur la scène sans former une trajectoire stable	Bruit continu d'introduction un peu terrifiant qui cherche à atteindre les spectateurs
V1 V2 sample	1'49 – 1'52	Un coup par syllabe autour de la scène (<u>A PO CA LYPSE</u>)	Déplacement des sources autour du public (gauche, droite, devant, derrière). Le spectateur entend le chœur apparaître partout autour de lui

TABLE 3 – Exemples de séquences de spatialisation pertinentes

est d'ailleurs plutôt éclectique, dans le sens où il constitue une bonne représentation de ce que l'on peut retrouver en terme de **style de spatialisation**.

Malgré cette panopli d'exemple (non exhaustif), j'ai dû me concentrer sur une fraction de ces derniers pour avoir un comportement intéressant, pour le temps qu'il m'était accordé. J'ai donc dû faire un choix de 2 ou 3 exemples maximum sur lesquels je pourrai me concentrer. Mon choix s'est porté sur l'exemple de la ligne cinq et dernière du tableau, en raison de son évidence et la clareté de ses mouvements. En écoutant l'extrait, on comprend facilement l'intention musical, ainsi que la **netteté** du déplacement de la source sonore. Il s'agit donc d'un bon choix de départ des données sur lesquels on peut diriger notre entraînement.

8.1.2 Sélection et formatage des données

Les données utilisées pour l'entraînement du modèle sont organisées sous forme de séquences temporelles contenant les paramètres numériques définis précédemment. Chaque séquence contient les caractéristiques audio correspondant au contenu musical ainsi que les coordonnées x et y de l'instrument choisi. La stratégie consiste à montrer au modèle une seule trajectoire par séquence, c'est-à-dire une seule séquence de coordonnées associée aux paramètres influençant cette trajectoire. Dans le cadre des expériences réalisées, seule la source V1 a été traitée afin de se concentrer sur une trajectoire unique.

Pour préparer les fichiers à prédire, on constitue un dataset appelé `_no_coord.csv` en suffixe, dans lequel les colonnes de coordonnées sont initialement vides. Pour rendre les données exploitables par le modèle, toutes les valeurs manquantes, y compris les coordonnées, sont temporairement remplacées par 0, créant ainsi un tableau complet sans NaN. Cette étape ne fournit pas de valeurs réelles pour les coordonnées mais permet au HMM et au *StandardScaler* de traiter correctement les observations. Le *StandardScaler* est cru-

cial car il met toutes les caractéristiques sur la même échelle, en centrant et réduisant les valeurs selon les moyennes et écarts-types calculés lors de l'entraînement. Cela garantit que le HMM compare correctement les différentes dimensions des données et prédit des états cohérents.

Toutes les séquences issues des différents fichiers sont ensuite concaténées tout en conservant leur longueur, de sorte que le modèle connaisse le début et la fin de chaque séquence. Après normalisation, le modèle de Markov caché gaussien prédit pour chaque instant l'état caché le plus probable en fonction des observations audio. Chaque état correspond à une configuration typique apprise pendant l'entraînement et est associé aux coordonnées moyennes correspondantes, ce qui permet de reconstituer la trajectoire spatiale prédite. Le remplacement des coordonnées par 0 est donc une solution pragmatique pour gérer les fichiers `_no_coord.csv`, tandis que la normalisation assure que les données audio sont correctement mises en forme pour la prédiction.

8.2 Prédiction du modèle

Le modèle ressort une liste de coordonnées qui correspond à la période qu'on lui a donné. Etant dans la première phase de résultats, on va s'intéresser aux comportements des trajectoires produites selon le contexte et moment du morceau que l'on a passé au modèle. L'évaluation de ces résultats se fera donc visuellement, en analysant les directions, intensités, précisions et bruitage du déplacement de l'instrument.

8.2.1 Observation

La première expérience réalisée correspond à l'extrait dernière ligne du tableau. Il s'agit de la voix (V1) qui se déplace en même temps que la prononciation du mot *apocalypse* en quatre syllabes distinctes. Cela permet de projeter la source dans toutes les directions cardinales, produisant un effet de *surround*. Un élément marquant de cette trajectoire est que la source est déplacée lors de courts silences entre les syllabes, ce qui donne parfois l'impression d'une "téléportation". Un autre aspect important est que les RMS de la voix 1, de la voix 2 et du sample sont utilisés comme paramètres d'entrée. Ces trois sources ayant des comportements similaires dans l'extrait, elles apportent une cohérence utile pour l'apprentissage.

Le paramètre `n_component` est fixé à 25 pour cette première expérience. L'idée est de conserver un nombre d'états suffisant pour capturer les trajectoires globales tout en évitant une sur-spécialisation. En effet, si ce nombre est trop faible, les trajectoires produites deviennent trop lissées et génériques. À l'inverse, un nombre trop élevé permet de modéliser finement les variations locales mais au risque de capter aussi du bruit ou des détails non pertinents.

Après prédiction, on obtient le résultat suivant. (21)

On observe déjà un résultat satisfaisant : le modèle reproduit bien l'effet rythmique et les directions de déplacement, même si seules trois directions sont effectivement utilisées au lieu de quatre comme dans l'extrait original. Le déplacement présente un léger bruit, ce qui peut s'expliquer par la nature même de la prédiction : les colonnes de coordonnées étant absentes dans les fichiers `no_coord`, elles sont remplacées par 0, et le modèle déduit ensuite leur valeur uniquement via les états cachés. Ce mode de reconstruction indirect peut accentuer les petites irrégularités.

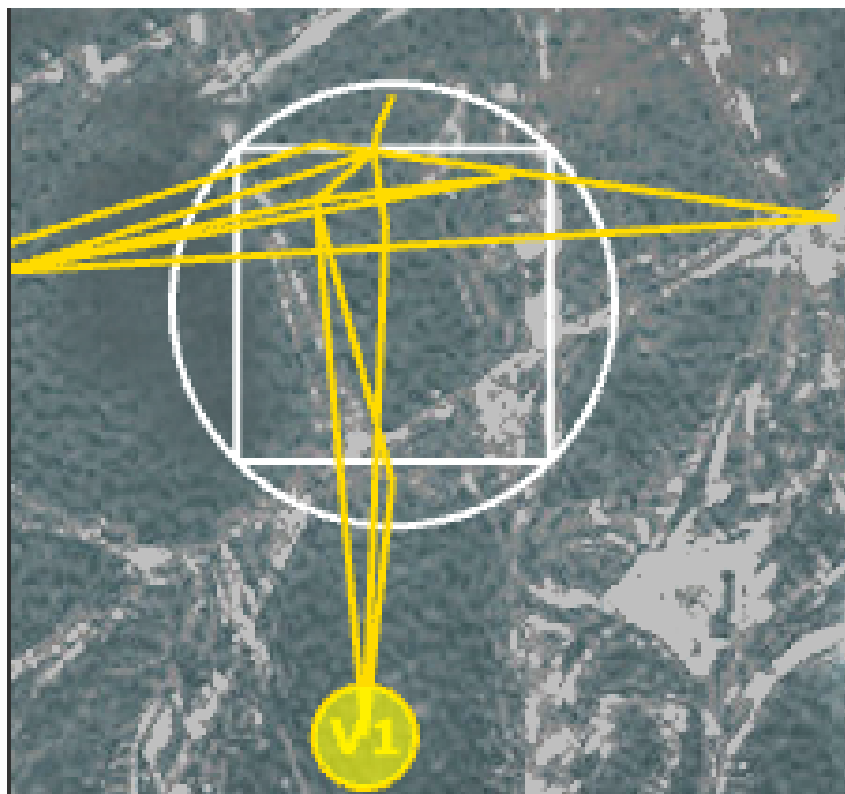


FIGURE 21 – Résultat d'entraînement 1

Nous ajoutons ensuite une autre séquence d'entraînement, correspondant à un déplacement plus lent et décrivant une trajectoire plus détaillée.

Avec cette séquence supplémentaire, le résultat devient moins évident à prédire. Le déplacement est restreint à une ligne et n'explore plus que deux directions. Afin de mieux capturer cette diversité, le paramètre *n_component* est augmenté à 40. De manière générale, plus la quantité et la variété des données d'entraînement augmentent, plus il est pertinent d'accroître le nombre d'états afin de permettre au modèle de décrire correctement la richesse observée, tout en réduisant le risque d'overfitting.

Ici (22), la source explore une direction supplémentaire (vers le haut), tout en restant proche de sa position initiale. Le rythme est respecté, mais les trajectoires sont plus bruitées. Cela est cohérent avec l'augmentation du nombre d'états, qui permet de mieux suivre les variations locales mais rend le tracé moins stable.

Pour approfondir, nous modifions les paramètres utilisés : cette fois-ci, nous ajoutons brutalement tous les RMS restants (guitare, basse et deux canaux de batterie), en plus des précédents.

Les trajectoires obtenues (23) se limitent à un plus petit ensemble de directions, ce qui semble logique puisque ces instruments sont relativement statiques pendant la séquence. En revanche, on observe une diminution nette du bruit : l'instrument devient plus stable et ses déplacements sont moins erratiques. On peut donc en conclure que l'ajout de variables "peu dynamiques" contraint le modèle à produire des trajectoires plus calmes, renforçant la cohérence globale.

Enfin, nous sélectionnons une séquence où guitare, basse et batterie bougent simultanément, mais pas la voix. Cette complexité accrue justifie l'augmentation de *n_component* à 50.

Le résultat reste pertinent (24), bien que la trajectoire dérive légèrement par rapport

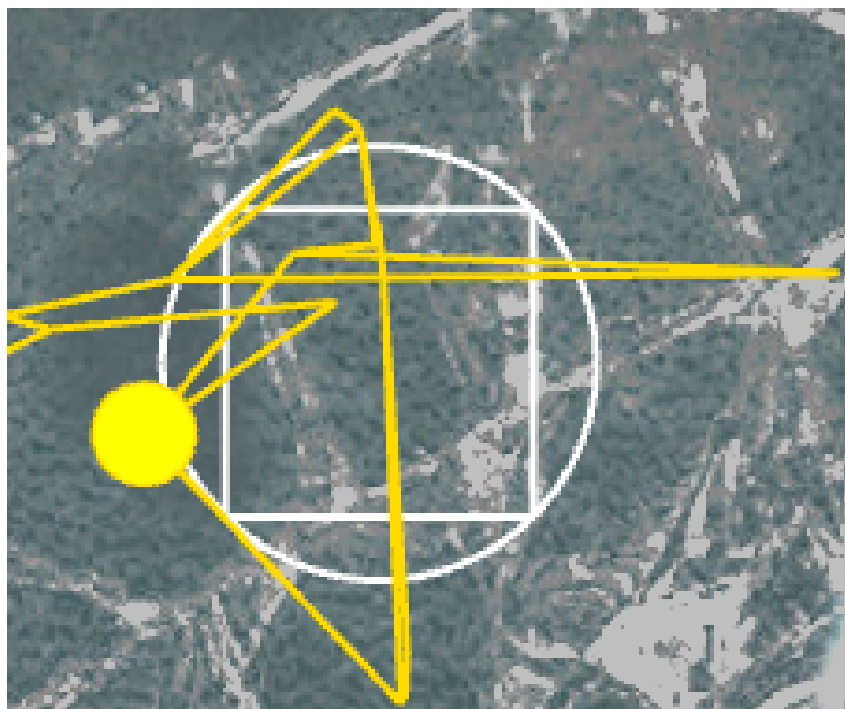


FIGURE 22 – Résultat d’entraînement 2

aux tracés plus stables observés auparavant. Ce comportement peut s’expliquer par la nature des déplacements instrumentaux qui apportent des variations moins prévisibles. Néanmoins, le rythme est respecté, et le bruit reste contenu.

Ces expériences montrent donc que :

- Le choix de $n_component$ influence directement le compromis entre stabilité et finesse locale.
- Les RMS sélectionnés déterminent fortement la dynamique des trajectoires (plus de sources statiques = moins de bruit).
- La diversité des séquences d’entraînement conditionne la capacité du modèle à généraliser au-delà de simples répétitions.

8.2.2 Analyse et discussion

Il est important de préciser que les résultats présentés ne constituent pas des conclusions définitives, mais bien une première série d’expériences exploratoires. L’objectif n’est pas encore d’obtenir des trajectoires finales exploitables telles quelles, mais plutôt d’observer comment différents paramètres influencent le comportement du modèle. Ces essais servent donc de base pour comprendre les mécanismes à l’œuvre et guider les ajustements futurs.

Les expériences montrent notamment que le nombre d’états cachés ($n_component$) joue un rôle central dans la qualité des trajectoires. Un faible nombre d’états conduit à des mouvements plus stables mais trop génériques, tandis qu’un nombre élevé permet de mieux suivre les détails locaux, au prix d’une plus grande variabilité et parfois d’un sur-apprentissage.

Le choix des descripteurs audio utilisés comme variables d’entrée, et en particulier des RMS, apparaît également déterminant. Lorsque plusieurs sources dynamiques sont prises en compte, le modèle tend à produire des trajectoires plus riches, mais aussi plus

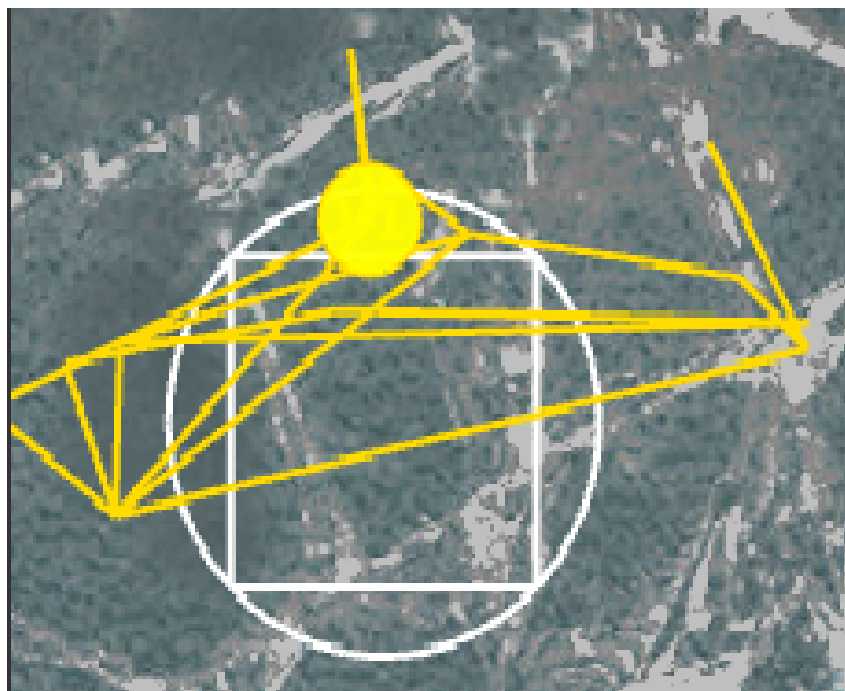


FIGURE 23 – Résultat d’entraînement 3

bruitées. À l’inverse, l’ajout de sources plus statiques contraint le système à générer des mouvements plus sobres et réguliers, ce qui peut être vu comme une forme de stabilisation.

Enfin, la diversité des séquences d’entraînement influence directement la capacité de généralisation. Plus les données couvrent de situations variées, plus le modèle peut éviter de simplement reproduire à l’identique des motifs déjà vus. Cette généralisation reste toutefois limitée dans les expériences actuelles, car les séquences choisies concernent une seule source (la voix V1), et sur des extraits relativement restreints.

Ces observations constituent donc des premiers jalons : elles permettent de mieux comprendre les leviers d’action du modèle, mais devront être consolidées par un corpus d’expériences plus large et plus varié.

8.2.3 Améliorations possibles et perspectives

Les résultats présentés dans cette étude constituent une première exploration du sujet et doivent être considérés comme des résultats préliminaires. De ce fait, de nombreux aspects restent à approfondir afin d’obtenir des conclusions plus solides et exploitables. Les différentes manipulations de paramètres audio ont permis de se faire une première idée de leurs effets sur la spatialisation et le rendu sonore, mais il est nécessaire d’étendre le nombre d’exemples et de situations testées. Cela permettra de confirmer si les tendances observées se maintiennent dans des contextes variés et d’augmenter la fiabilité des observations.

Pour les prochaines étapes, il serait pertinent d’intégrer d’autres paramètres audio, tels que le H1 ou les coefficients MFCC, qui fournissent une information plus fine sur le timbre et la couleur sonore. L’ajout de ces mesures permettrait d’enrichir l’analyse et de mieux caractériser les sons au-delà de leurs propriétés spatiales, ce qui pourrait révéler des nuances importantes pour la qualité de la spatialisation et pour la perception par l’auditeur.

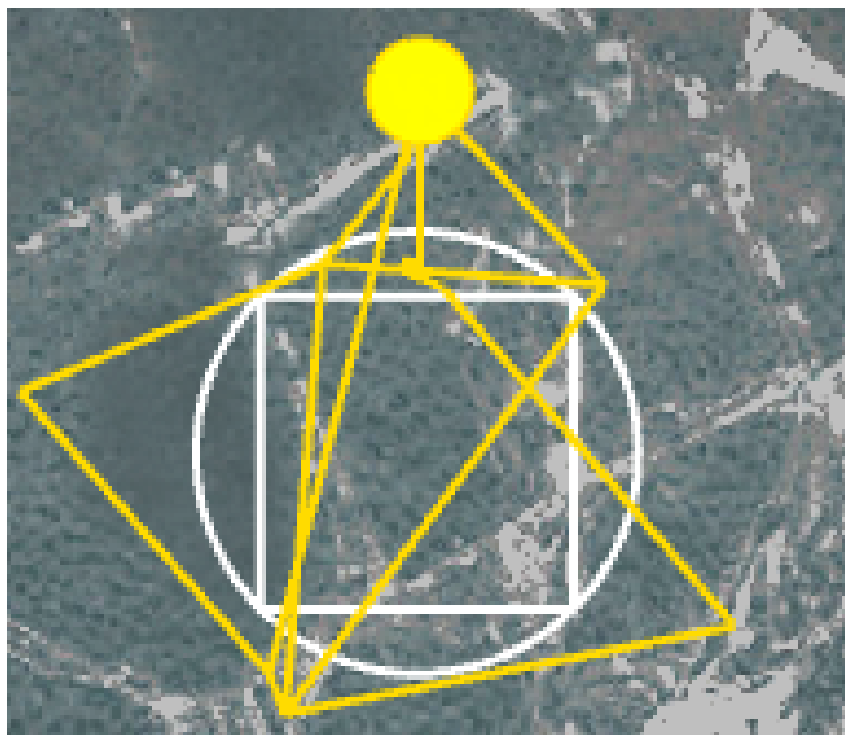


FIGURE 24 – Résultat d’entraînement 4

Un autre point essentiel concerne l’analyse des résultats. Actuellement, celle-ci repose essentiellement sur des impressions subjectives et des évaluations humaines directes. Bien que ces observations soient utiles pour un premier repérage, elles restent limitées et difficiles à généraliser. Il sera donc nécessaire de mettre en place des méthodes plus quantitatives et automatisées, comme des mesures statistiques ou des indicateurs perceptifs standardisés, afin d’obtenir une évaluation objective et reproductible des effets des différents paramètres audio.

Il serait également possible de choisir une autre mesure que le RMS en considérant des unités de mesure basées sur la perception humaine, telles que les dBFS ou, de manière plus pertinente, les LUFS. Cette option permettrait d’adapter l’analyse de l’intensité à une échelle perceptive, ce qui pourrait enrichir l’interprétation des résultats et améliorer la pertinence des modèles de Machine Learning.

Enfin, une perspective future du stage serait une dimension visuelle qui pourra être intégrée au projet : chaque piste instrumentale se verra associée à une couleur lumineuse, dont les déplacements dans l’espace scénique seront synchronisés avec ceux de la spatialisation sonore. L’ambition serait ainsi de proposer une expérience artistique enrichie, alliant spatialisation auditive et mise en scène lumineuse.

9 Bilan / Conclusion

Le sujet de mon stage a été particulièrement passionnant et stimulant. Dès le début, j'ai été confronté à des problématiques variées qui m'ont permis d'approfondir mes connaissances théoriques tout en les appliquant à des situations concrètes. Travailler sur ce projet m'a offert l'opportunité de découvrir des aspects techniques que je n'avais encore jamais abordés et de mieux comprendre les enjeux liés à la spatialisation et à la modélisation des trajectoires.

L'équipe du SCRIME a été d'un grand soutien tout au long de cette expérience. Leur accueil chaleureux, leur disponibilité et leur capacité à partager leur expertise ont rendu ce stage très enrichissant. Les échanges avec les membres de l'équipe ont été autant de sources d'apprentissage, me permettant de progresser rapidement et de mieux comprendre les attentes du projet. Je suis particulièrement reconnaissant d'avoir eu l'opportunité de travailler dans un environnement aussi collaboratif et stimulant.

Les résultats obtenus au cours de ce stage sont prometteurs, même si beaucoup reste à explorer et à développer. Certaines parties du projet, comme la documentation sur la spatialisation, ont été particulièrement vastes et m'ont demandé un investissement important en temps. Cela a quelque peu retardé le moment où j'ai pu commencer à coder et à travailler sur les aspects les plus concrets du projet. Cependant, cette phase de documentation a été essentielle pour acquérir une compréhension globale du sujet et pour mieux structurer mon approche lors des phases pratiques.

En parallèle, l'expérience m'a permis de mieux appréhender le compromis entre théorie et pratique, ainsi que l'importance de la planification et de l'organisation dans un projet complexe. Chaque étape, même celle qui semblait moins concrète, a contribué à ma progression et m'a donné une vision plus complète des enjeux techniques et scientifiques du stage.

Enfin, je tiens à remercier chaleureusement toute l'équipe du SCRIME pour leur encadrement, leurs conseils et leur soutien tout au long de cette expérience. Ce stage a été pour moi une occasion unique de développer mes compétences, de découvrir de nouvelles méthodes de travail et de me confronter à des problématiques réelles dans un environnement professionnel stimulant. Il constitue une base solide sur le domaine de la spatialisation et plus généralement sur l'audio, domaine sur lequel je n'ai pas eu beaucoup l'occasion de travailler pendant mes études.

10 Glossaire

Glossaire

acousmatique Se dit d'un son dont on entend la diffusion sans voir la source. Par extension, désigne un courant musical (initié par Pierre Schaeffer) où la perception du son est indépendante de son origine instrumentale ou causale.. 4

acousmonium Dispositif de diffusion sonore constitué d'un ensemble de haut-parleurs de tailles et de caractéristiques variées, conçu à l'origine par le GRM (Groupe de Recherches Musicales) pour interpréter des œuvres électroacoustiques en concert.. 4

ambisonie Méthode de spatialisation sonore se concentrant sur la directionnalité des ordres. Il en existe différents ordres allant de 1 à théoriquement l'infini. L'ordre de l'ambisonie désigne le nombre de direction et donc sa précision et richesse sonore.. 3

BPM Beats per Minute, tempo du morceau. 14

dBFS Decibels Full Scale, mesure de niveau audio. 33

GaussianHMM HMM avec observations continues modélisées par des lois normales. 26

HMM Hidden Markov Model, modèle de Markov caché. 25

LUFS Loudness Units Full Scale, mesure perceptive de l'intensité sonore. 33

Max/MSP Environnement de programmation pour le traitement audio et visuel. 6

MFCC Mel-Frequency Cepstral Coefficients, descripteurs du timbre sonore. 32

Quadriphonie Diffusion sonore sur quatre canaux/haut-parleurs. 5

RMS Root Mean Square, moyenne quadratique du signal audio. 15

spatialisation Placement des sources sonores dans un espace virtuel ou réel. 3

StandardScaler Méthode de normalisation des données en Machine Learning. 28

trajectoires Succession de positions d'une source sonore dans le temps. 6

Références

- [1] Concert Spatialiser, Odyssey Cercle. <https://odyssey.cercle.io/>.
- [2] Sound Particles : logiciel de spatialisation. <https://soundparticles.com/>.
- [3] The Sphere Las Vegas. <https://www.thesphere.com/>. Consulté le 15 septembre 2025.
- [4] Joseph Larralde and Pierrick Legrand. Spatialisation d'un concert de musiques actuelles amplifiées : étude de cas avec le groupe Praetorian. In Proceedings of the 32th Journées d'Informatique Musicale, Lyon, France, June 2025. GRAME and Inria. PDF disponible à <https://hal.science/hal-05102329v2/file/9.pdf>.