

CAHIER  
LOUIS-LUMIÈRE

no.

15

# La perception du son spatialisé :

## Un son propre pour un sens figuré

sous la direction de  
Corsin Vogel et Sylvain Lambinet

CAMILLE AUBRIOT  
VALENTIN BAUER  
ALAN BLUM  
LÉA CHEVRIER  
HUGO CORBEL  
DANIÈLE DUBOIS  
FRANÇOIS-XAVIER FÉRON  
CYRIL HOLTZ  
JEAN-PASCAL JULLIEN  
SYLVAIN LAMBINET  
CLÉMENCE LAVIGNE  
ALEXIS LING  
JEAN-MARC LYZWA  
JEAN-CHRISTOPHE MESSONNIER  
LAURENT MILLOT  
BENOÎT TURQUETY  
CORSIN VOGEL





# La perception du son spatialisé : un son propre pour un sens figuré

sous la direction scientifique de Corsin Vogel et Sylvain Lambinet

**P.3**  
**Avant-propos**  
Vincent Lowy

**P.5 À 7**  
**Introduction.**  
Sylvain Lambinet  
et Corsin Vogel

## 1. Histoires et concepts

**P. 10 À 30**  
**Espaces résonants : le son  
comme art du lieu**  
Benoît Turquetly

**P. 31 À 47**  
**L'immersion sonore : réalités  
physique et virtuelle,  
réalités psychologiques et  
culturelles**  
Corsin Vogel et  
Danièle Dubois

**P. 48 À 69**  
**Influence du réalisme,  
de l'agentivité et des  
représentations  
sur l'immersion sonore en  
réalité augmentée**  
Valentin Bauer

**P. 70 À 92**  
**Composer (avec) l'espace :  
une pratique musicale  
polysémique**  
François-Xavier Féron

## 2. Pratiques et technologies

**P. 96 À 113**  
**Enjeux de l'immersion dans  
une acoustique virtuelle pour  
les systèmes de spatialisation  
auditive**  
Jean-Pascal Jullien

**P. 114 À 125**  
**Méthodes de prise de son en  
multicanal / Objets**  
Jean-Christophe Messonnier,  
Jean-Marc Lyzwa et  
Alexis Ling

**P. 126 À 133**  
**Raconter en son binaural :  
entretien avec Léa Chevrier**  
Alan Blum

**P. 134 À 143**  
**Le multicanal étendu au cinéma :  
entretien avec Cyril Holtz**  
Sylvain Lambinet

## Varia

**P. 146 À 164**  
**Alternative aux analyses de  
Fourier globale et glissante pour  
les signaux temporels : Analyse-  
Resynthèse par Intégration de  
Densité Spectrale**  
Laurent Millot

## Notes de recherche

**P. 168 À 177**  
**Les enjeux de la transition dans  
la composition des paysages  
sonores**  
Hugo Corbel, Master Son 2021,  
ENS Louis-Lumière

**Filmer la brume, le brouillard  
et la fumée : enjeux artistiques  
et techniques**  
Camille Aubriot, Master Cinéma  
2020, ENS Louis-Lumière

**Reconstitution du procédé de  
photographie en relief  
de Louis Lumière :  
« la Photo-stéréo-synthèse »**  
Clémence Lavigne, Master  
Photographie 2021, ENS Louis-  
Lumière

## Notes de lecture

**P. 180 À 183**  
*Sensory Experiences: Exploring  
meaning and the senses*, Danièle  
Dubois, Caroline Cance, Matt Coler,  
Arthur Paté, Catherine Guastavino,  
John Benjamins Publishing, 2021  
Corsin Vogel

*Puissances de la parole. À l'écoute  
des films*, Mathias Lavin, Éditions  
Mimésis, 2021  
Corsin Vogel



## Avant-propos

Vincent Lowy

Cette quinzième livraison du Cahier Louis Lumière a été conçue et pilotée par deux enseignants permanents de l'école qui ne sont pas statutairement des enseignants-chercheurs. Et même si un jour ils rejoignent cette honnête corporation, ce que nous leur souhaitons, Corsin Vogel et Sylvain Lambinet ont réussi ce beau numéro avec leur seul talent, j'allais dire leur passion, nourris par leurs parcours singuliers, riches de leurs expériences respectives dans différents domaines liés au monde du son. Ils se sont emparés avec brio de la proposition un peu iconoclaste que je leur ai faite il y a un an et ont su démontrer qu'à l'ENS Louis-Lumière, la recherche est aussi portée en dehors des cadres souvent contraints et parfois formatés de la recherche académique. Ceci démontre une des qualités incontestables de notre école : sa capacité à questionner les concepts, les pratiques et même les disciplines en dehors des sentiers battus.

Nous tenons donc à remercier Corsin et Sylvain pour ce travail de fond, qui fait suite aux réflexions et recherches développées dans le numéro 13, qui abordait l'immersion audiovisuelle sous un angle privilégiant l'image, laissant ouvert pour la suite le champ de l'immersion sonore et de la spatialisation.

La genèse de ce numéro repose sur une approche à la fois perceptive et historique, technologique et praticienne du son spatialisé. Sans évidemment prétendre à l'exhaustivité, les concepteurs de ce dossier ont voulu brosser un portrait de l'immersion sonore et des évolutions des techniques de spatialisation du son, portrait qui nous semblait particulièrement

opportun au regard des recherches et développements récents que nous constatons dans ce domaine. Les technologies prometteuses au cinéma (Dolby Atmos), dans le spectacle vivant (Wavefield Synthesis) ou dans les concerts immersifs (Ambisonie) se confrontent ainsi aujourd'hui à notre expérience perceptive, expérience modelée par des usages culturels plus ou moins conscients. Ce nouvel environnement implique des changements de paradigmes, qu'il était urgent d'explorer dans le cadre de cette revue : les étudiant.e.s s'en emparent déjà et nous les voyons déjà se préparer à des usages et à des technologies qui mutent d'année en année. Ils en sont les futurs expérimentateurs.

Il s'agissait par conséquent de nous entourer de chercheurs.cheuses, de praticien.ne.s, de spécialistes, de pédagogues pour donner une forme à ce portrait du son spatialisé. Issus de centres de recherche reconnus (universités de Lausanne, Paris-Saclay, CNRS, Ircam), de grandes écoles (Conservatoire de Paris), explorant des pratiques ancrées dans le cinéma ou les balades sonores immersives, ces professionnel.le.s du sonore et de son histoire se sont attelés à cette lourde tâche de réaliser des contributions claires et synthétiques de leurs travaux. Et ce numéro de servir de guide de référence à toute personne, spécialiste ou étudiante, dans une approche globale de l'immersion sonore.

Au-delà et à titre plus personnel, je veux souligner qu'il s'agit du dernier numéro d'une série de cinq (plus un hors-série), que j'ai eu l'honneur de piloter comme directeur de l'ENS Louis-Lumière et donc comme directeur de cette publication. En cette fin de mandat, je forme le vœu que la revue continue de porter dans les années qui viennent les valeurs d'excellence scientifique et de pluralisme qui ont toujours été les siennes et dont nous avons tenté d'être dignes depuis cinq ans. Et surtout que les étudiant.e.s s'en emparent et puissent y contribuer toujours plus largement et librement. Vive l'avenir !



## Introduction

Corsin Vogel et Sylvain Lambinet

« Très court espace de temps à travers de très courts temps d'espace. Cinq, six : le nacheinander. Exactement, et voilà l'inéluctable modalité de l'ouïe. Ouvre les yeux. »<sup>1</sup>

Immersion sonore "totale", "hors du commun", "unique", "inégalée"... Les superlatifs ne manquent pas lorsqu'il s'agit de cette promesse de révolutionner l'expérience d'auditeur, d'enfin lui donner à entendre un panorama sonore jusqu'alors jamais atteint. Ces quinze dernières années furent particulièrement fécondes en annonces de la part de fabricants d'enceintes, casques, téléviseurs, processeurs et logiciels de signal audio, jusqu'aux plateformes de diffusion des contenus numériques musicaux et audiovisuels.

Quel est donc cet Eden auditif, enfin rendu accessible ? À quoi tient cette « immersion » sonore ? Il apparaît sans conteste que c'est toujours la *spatialité* du son qui est visée par ces termes. Voilà qui mérite examen. Le sonore n'a évidemment pas attendu pour être spatial : son déploiement depuis sa source jusqu'à l'auditeur, ses interactions avec son milieu, qui vient modifier et colorer son trajet et ses réflexions, la capacité de notre système auditif binaural à localiser dans les trois dimensions, tout ceci existe depuis que le son se propage et qu'il y a des paires d'oreilles sur son trajet. Il nous est impossible d'imaginer un son qui serait perçu indépendamment de ses caractéristiques spatiales.

Plus neuve, toutefois, est la question de la restitution de cette spatialité, qui se pose pour la première fois avec l'apparition des systèmes d'enregistrement et de reproduction sonores. La stéréophonie, dans son principe, offre déjà la possibilité d'une restitution de la latéralisation et de la profondeur. Nous rappelons ici que le grec στερεός (*stéréos*) signifie *ferme, solide* et contient donc l'idée de relief, et que l'intérêt initial pour la reproduction de la spatialité sonore ne fait pas d'hypothèses sur le nombre de canaux à employer. Si l'on peut faire remonter les toutes premières expérimentations à Clément Ader et son Théâtrophone en 1881, puis à Alan Blumlein avec le premier système à deux canaux en 1931, il faut attendre les années 1960 pour voir se démocratiser les enregistrements et s'installer les systèmes de diffusions dans les foyers, au point que « stéréophonique » devienne synonyme de « deux canaux ». Avec cette démocratisation, on assiste également à l'avènement d'un discours sur la fidélité<sup>2</sup> : celui-ci, qui prévaut toujours aujourd'hui, met en avant la finesse et la qualité de la perception spatiale, autrement dit, l'ambition de reproduire un champ sonore au plus proche de l'original. Paradoxalement, l'objectif en filigrane semble d'emblée être celui d'un effacement du système de reproduction. Parvenu à son plus haut degré de perfection, celui-ci offrirait, en toute transparence, un rapport immédiat à l'objet qu'il restitue, et permettrait à l'auditeur un accès sans filtre – « comme si vous y étiez ».

Les chercheurs, artistes, ingénieurs rivalisent depuis d'ingéniosité et d'audace. S'appuyant sur une compréhension toujours plus poussée des mécanismes de l'audition humaine et des ressources techniques, démultipliées depuis l'avènement du numérique, nombre de systèmes de spatialisation ►

<sup>1</sup> James Joyce, *Ulysse*, trad. A. Morel, revue par V. Larbaud, S. Gilbert et l'auteur, Paris, Gallimard, « Folio », 1996, épisode 3, p. 59 (éd. orig. *Ulysses*, Shakespeare and Compagny, Paris, 1922).

<sup>2</sup> Sur la question de l'histoire de la stéréophonie et de sa réception, voir *Living Stereo, Histories and Cultures of Multichannel Sound*, P. Théberge, K. Devine, T. Everett et al., Bloomsbury, 2015

sont désormais disponibles. Du principe de la stéréophonie étendue dans les dispositifs surrounds, à la synthèse de front d'onde se déployant sur un grand nombre de haut-parleurs alignés, de l'ambisonie répartie sur une sphère de haut-parleurs au binaural qui se limite frugalement à un casque audio, il est notable que chaque solution articule une proposition de spatialisation avec des moyens et un *lieu* de diffusion spécifiques. La recherche de l'espace à restituer est indissociable d'une réflexion sur l'espace de restitution. On pense bien sûr aux problématiques de couplage entre l'acoustique propre de l'enregistrement et celle de la salle où il est diffusé, mais au-delà se dessine également un des enjeux de l'immersion. L'auditeur étant nécessairement lui-même situé, l'orientation de sa posture d'écoute devient déterminante si l'on vise à le faire s'abstraire de son environnement pour lui en proposer un autre. Quelles sont les conditions nécessaires et suffisantes pour atteindre ce but ? Tout tient-il à la qualité et à la finesse du dispositif technique ? Salle de concert, cinéma, écoute solitaire au casque, déambulation binaurale, système de réalité virtuelle, salle d'exposition : la diversité des lieux de diffusion nous laisse anticiper qu'il n'y aura pas de réponse technique univoque.

Ces questions, au cœur de la foisonnante histoire du son spatialisé, en amènent nécessairement une autre : restituer l'œuvre dans sa spatialité, soit, mais laquelle ? Il faut distinguer l'enregistrement « témoin », rendant compte de l'évènement sonore (concert, *field recording*...), et l'œuvre conçue et pensée pour la spatialisation. Car, comme il arrive souvent, l'outil initialement destiné à répondre à un besoin technique devient lui-même objet d'expérimentation et origine d'une nouvelle écriture esthétique, tirant parti des possibilités qu'offre

la technique pour se déployer dans un espace neuf, faire entendre des acoustiques paradoxales, des trajets sonores défiant le réel, créant des perspectives inédites. L'efficacité de l'immersion ne peut alors plus résider dans la seule performance des systèmes de captation et de diffusion, et il faut admettre un ensemble de mécanismes reposant sur la possibilité de l'auditeur de fonder son engagement sensoriel et émotionnel sur autre chose qu'une « simple » reproduction, physiquement correcte, des caractéristiques du champ sonore original.

Il faut alors ouvrir ces questions, et y inviter les ressources à la fois d'une psychophysiologie plus large, nous renseignant sur la multimodalité de la perception, et celles de la psychologie cognitive et de la sémantique, pour dessiner les contours d'un rapport plus fin entre ce qui est perçu, ce qui est entendu, et ce qui est pensé. Le son d'un oiseau en vol n'est-il pas toujours en haut, et le bruissement des pas sur le gravier en bas ? L'antique déclaration selon laquelle « l'art imite la nature<sup>3</sup> » peut une fois de plus être remise à contribution, et nous servir de guide dans l'examen de ce que cette nature re-présentée a de singulier, de cette logique propre au produit de la technique, qui saisit du réel les ressorts profonds « de telle sorte qu'une prompte intelligence puisse, en les dépassant, saisir leur signification inexprimée<sup>4</sup> ».

Exprimer les significations de cette multiplicité de représentations des espaces sonores, expliciter notre immersion perceptive dans des environnements plurisensoriels, convoquer les diverses approches conceptuelles et méthodologiques dans un contexte pluridisciplinaire pour les confronter aux questions technologiques liées à l'immersion sonore, tels sont quelques uns de nos objectifs dans ce numéro du

<sup>3</sup> Aristote, *Physique II*, « η τέχνη μιμείται τὴν φύσιν ».

<sup>4</sup> James Joyce, *Carnets de Paris*.

*Cahier Louis-Lumière*. Tout au long de ce numéro nous allons questionner ces approches, dont la visée est finalement toujours la même : celle de placer l'humain percevant au centre d'un dispositif qui permettra de le projeter dans un espace sonore et/ou musical, voire multimodal, mentalement reconstruit. L'ensemble des contributions embrassera donc une diversité de points de vues distincts, complémentaires.

Les premières contributions s'inscrivent dans un courant intellectuel écologique et de cognition située, où le concept d'*ambiances*, perçue et subie par l'humain, est remplacé par celui d'*Umwelt*, traduit de l'allemand par le terme *milieu*<sup>6</sup>, dans lequel l'humain se situe et interagit. Ces approches se basent également sur les cartographies sonores, les *soundscape*<sup>7</sup>, les archives sonores et les nouvelles approches de réalité virtuelle et de réalité augmentée. Elles trouvent ainsi un écho dans le travail d'historien du son de Benoît Turquety sur l'art des espaces résonants (chapitre 1), mais également dans les descriptions multiples des réalités associées à la notion d'immersion sonore : réalités physique, virtuelles, artéfactuelles et culturelles, présentées par Corsin Vogel et Danièle Dubois (chapitre 2). L'aspect plus spécifique de la réalité virtuelle, et notamment dans le contexte de la réalité augmentée comme outil d'aide à la personne, est abordé par Valentin Bauer (chapitre 3). Le musicologue François-Xavier Féron complète cette partie par un regard historique sur les démarches compositionnelles de l'espace physique, sonore et musical (chapitre 4).

La seconde partie des contributions se focalise davantage sur les pratiques contemporaines du son spatialisé et les technologies sous-jacentes. Ainsi,

Jean-Pascal Jullien présente les grands principes de la diffusion en multicanal et les véritables enjeux d'une acoustique architecturale virtuelle (chapitre 5). Forts de ces connaissances liées à la restitution du son immersif, nous abordons avec Jean-Christophe Messonnier, Jean-Marc Lyzwa et Alexis Ling la pratique de la prise de son et du mixage multicanal orienté objet, appliquée à la captation de la musique de concert (chapitre 6). Nous entendons ensuite en écho les voix de certains acteurs professionnels de l'immersion sonore, en technique binaurale avec Léa Chevrier dans un entretien mené par Alan Blum (chapitre 7), et en Dolby Atmos et mixage orienté objet au cinéma avec Cyril Holtz dans un entretien avec Sylvain Lambinet (chapitre 8).

Bonne lecture, et ouvrons les yeux pour entendre.

**« Si on peut passer ses doigts à travers, c'est une grille, sinon, une porte.**

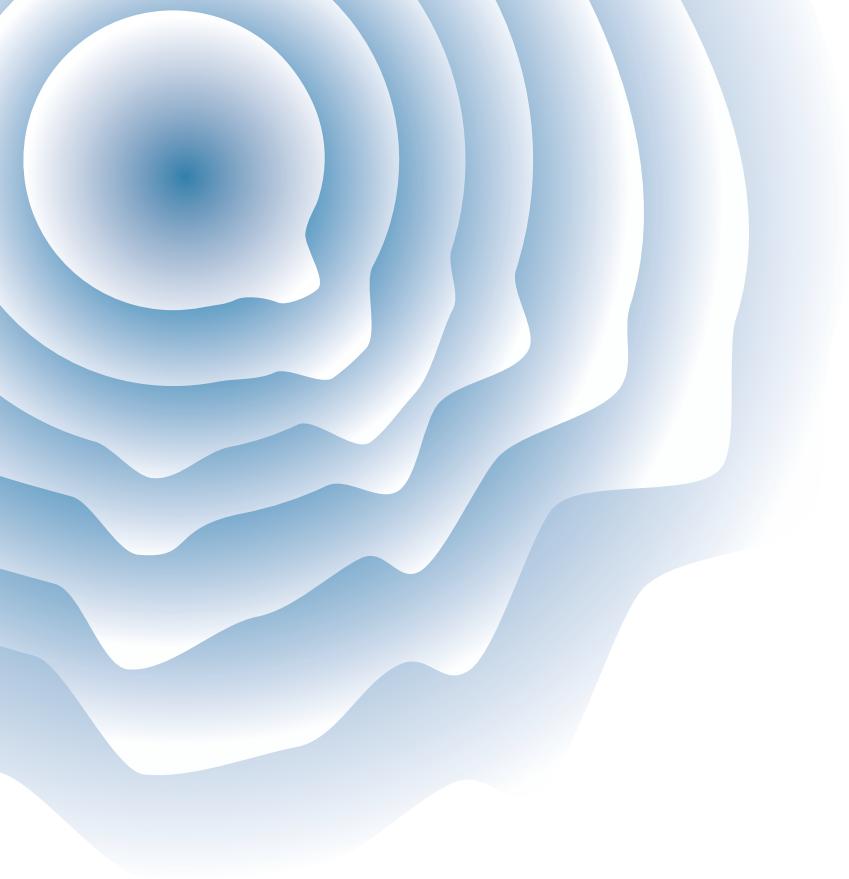
**Fermons les yeux pour voir. »**

<sup>5</sup> Augoyard, J.-F. (1995), *À l'écoute de l'environnement : répertoire des effets sonores*, Parenthèses, Marseille.

<sup>6</sup> Uexküll, J. von (1921), *Milieu animal et milieu humain*, Rivages, 2010 (éd. orig. *Umwelt und Innenwelt der Tiere*, Springer).

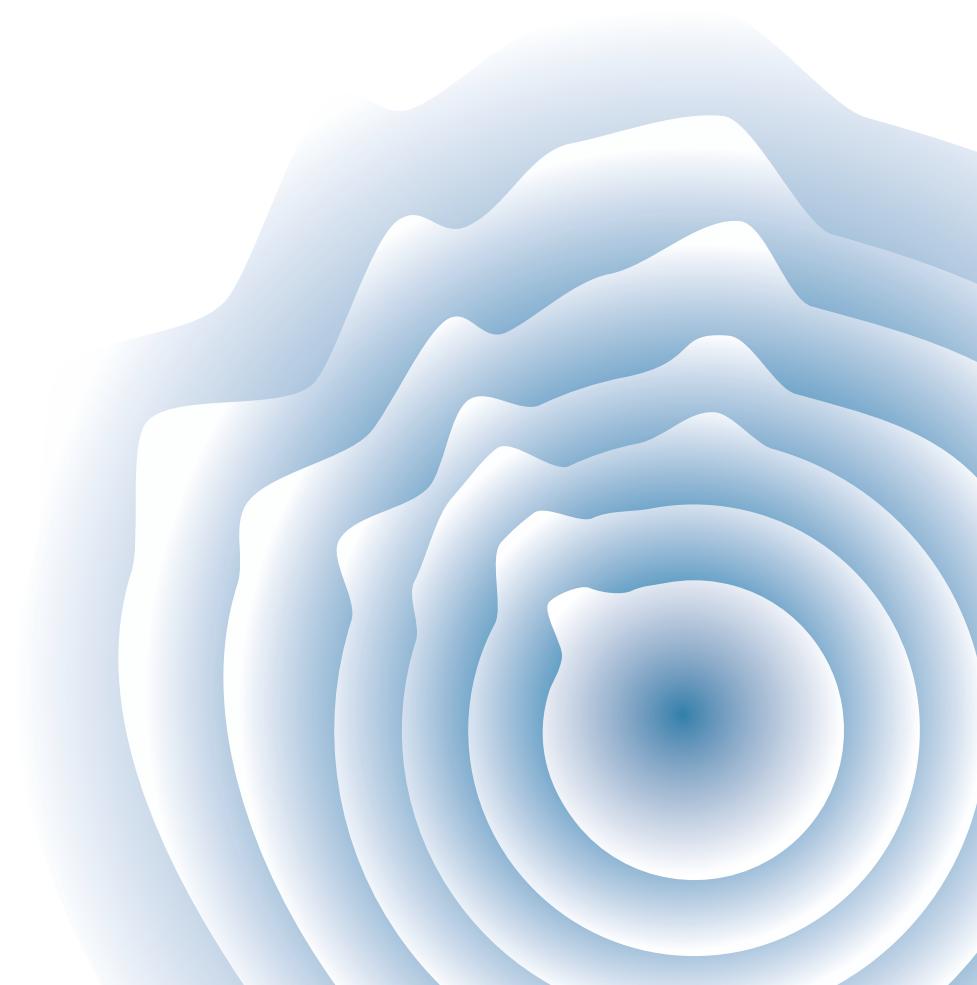
<sup>7</sup> Schafer, R. M. (1977), *Le paysage sonore, le monde comme musique*, Marseille, éditions Wild Project, 2010 (éd. orig. *The Tuning of the World*, Knopf, New-York).

<sup>8</sup> James Joyce, *Ulysse*, trad. A. Morel, revue par V. Larbaud, S. Gilbert et l'auteur, Paris, Gallimard, « Folio », 1996, épisode 3, p. 58 (éd. orig. *Ulysses*, Shakespeare and Compagny, Paris, 1922).



**1.**

# **Histoires et concepts**



## Espaces résonants : le son comme art du lieu

| Benoît Turquety

### Résumé

Inventé par R. Murray Schafer dans les années 1970, le concept de « paysage sonore » a joué un rôle fondamental dans l'émergence des *sound studies* et de ce que l'auteur nomma « écologie acoustique ». Face à l'élévation constante du niveau sonore des grandes villes occidentales, Schafer déploya une stratégie d'étude de l'environnement acoustique des lieux fondée à la fois sur l'écoute subjective et sur la mesure systématique, sur une critique de la pollution sonore et sur une reconsidération de la qualité musicale des bruits. Cette démarche fut déterminante pour le développement d'une réflexion sur les interactions entre son et espace. On en retrouve des traces aussi bien dans la tradition du *field recording*, conçu comme méthode d'investigation sonore et musicale des lieux, que dans les tactiques d'enregistrement mises en place par certains musiciens particulièrement intéressés par l'interaction entre instrument et espace – notamment dans le cadre de l'improvisation libre. Enfin, cette question du rapport son/espace se pose à nouveau dans le travail cinématographique, singulièrement en son direct. Ces pratiques montrent combien le son est toujours spatialisé, d'emblée et dans toutes ses dimensions.

### Abstract

*Invented by R. Murray Schafer in the seventies, concept of « Soundscape » played a fundamental role in the rise of sound studies, and in what the author called « acoustical ecology ». Witnessing a constant augmentation of sound level in occidental cities, Schafer deployed a strategy for studying the acoustical environment of places, based both on subjective listening and systematic measurement, as well as on a critic of sound pollution and a new look over sound quality of noises. This approach was decisive for developing a rethinking of interactions between sound and space. Traces of it can be found in the tradition of field recording, conceived as a sonic and musical investigation method of places, as well as in the recording strategies used by some musicians, particularly interested in the interactions between instruments and space – especially within free improvisation. Lastly, this question of the relation between sound and space appears once again in the cinematographic field, foremost in production sound. These practices show to what extent sound is always spatialized, straight away and in all its dimensions.*

En 1977, R. Murray Schafer publiait *The Tuning of the World* (1977b), « l'Accordage du monde », ouvrage fondateur des *sound studies* réédité ensuite sous le titre de *The Soundscape* (1994). Fait relativement rare, le livre a inauguré à la fois un champ d'études théoriques, universitaires, et un nouvel espace de pratiques artistiques, documentaires et musicales. L'une des innovations qui a le plus marqué est le néologisme *soundscape*. Basé sur le terme de *landscape*, « paysage », c'est-à-dire « étendue du pays que l'on voit d'un seul aspect » écrivait Littré, le *soundscape* désigne le « paysage sonore », c'est-à-dire l'ensemble de ce que l'on entend lorsqu'on se place en un certain lieu. Construire ce terme, c'était en fait construire le paysage sonore en objet, le faire entendre. Schafer y inventait ainsi une nouvelle compréhension du lien profond qui unit le son et l'espace, qui ancre les sons dans un lieu, et qui établit le lieu comme vaste configuration sonore.

## Portraits de lieux

Mais *The Tuning of the World* ne venait pas, comme on dit, de nulle part. R. Murray Schafer était compositeur, d'œuvres orchestrales et vocales, depuis le milieu des années 1950. Il était également professeur dans le département de communication de l'Université Simon Fraser, à Vancouver. Or dans le courant des années 1960, Schafer prend conscience des transformations qui affectent sa ville, et notamment de l'augmentation notable du niveau sonore dans tout l'espace urbain. Il va alors, dans ses cours et dans diverses publications, travailler d'une part à essayer de décrire ces évolutions, à les évaluer et à les mesurer, et d'autre part à réfléchir sur le concept de bruit et sur la compréhension du problème de la pollution sonore. C'est dans ce cadre qu'apparaît le concept de *soundscape*, dans une brochure de 1969

intitulée *The New Soundscape : A Handbook for the Modern Music Teacher* qui constitue à maints égards une première version de l'ouvrage de 1977, tandis qu'en 1970 Schafer publie *The Book of Noise*, dont le ton est beaucoup plus inquiet sinon alarmant. Ce « Livre du bruit » s'inscrit davantage au sein d'une réflexion institutionnelle et politique sur les effets négatifs des bruits modernes sur la vie humaine<sup>1</sup>, que dans une tradition musicale.

L'œuvre de Schafer restera marquée par cette ambivalence, voire cette contradiction. D'un côté, le Canadien l'affirme à plusieurs reprises dans *Le paysage sonore* : « Je considérerai le monde dans ce livre comme une immense composition musicale (Schafer, 2010, p. 25). » Il explicite ce postulat en faisant référence aux écrits et aux œuvres de John Cage, dans lesquelles les sons musicaux proviennent autant de la salle ou de la rue que de l'orchestre – jusqu'à cette œuvre emblématique que fut *4'33"*, interprétée pour la première fois par David Tudor en 1952, qui, pendant cette durée exacte, ne toucha le piano que pour ouvrir et refermer le clavier. La pièce n'était pas faite uniquement de silence : elle était tissée de tous ces « bruits » qui le parasitent, ou qui le *composent*. Cage amenait ainsi à repenser la hiérarchie si prégnante dans la culture occidentale entre *bruit* et *musique*, et à la considérer largement comme une construction idéologique.

Schafer adopte cette approche, qui est aussi celle d'une partie de la musique concrète, mais chez lui cela ne rend jamais caduc le concept de *bruit*, qui conserve une définition claire : « Est bruit tout son non désiré. Le bruit c'est le mauvais son au mauvais endroit<sup>2</sup>. » Cela fait certes du bruit « a relative term », reconnaît Schafer (*idem*), puisqu'un même son peut être, dans différents contextes ou pour différentes personnes, ►

<sup>1</sup> L'ouvrage fait notamment référence à des publications sur ce sujet de l'UNESCO (1967), du Committee on the Problem of Noise in Great Britain (1966) ou de la Task Force on Noise Control établie par la mairie de New York (1970) (Schafer, 1970, p. 30).

<sup>2</sup> « Noise is any undesired sound. Noise is the wrong sound in the wrong place. » (Schafer, 1970, p. 4, notre traduction)

du bruit, ou non. Mais abandonner le concept de bruit ferait perdre son sens à celui de « pollution sonore », et donc à toute l'entreprise. Schafer se retrouve ainsi au cœur d'une tension, bien rendue par la préface du *New Soundscape*, où il se décrit demandant « Oui, mais est-ce de la musique ?<sup>3</sup> » en entendant un avion à réaction passer au-dessus de lui, après que cette même question se soit fait entendre dans les loges aux premières de toutes les grandes œuvres de la modernité occidentale, la *Cinquième symphonie* de Beethoven, *Tristan und Isolde*, *Le sacre du printemps* et *Poème électronique* de Varèse : « Oui, mais est-ce de la musique ? »

De cette tension inaugurale, Schafer va basculer vers la question qui va devenir centrale dans la méthodologie des *sound studies* et dans les pratiques artistiques associées, notamment liées au *field recording* : celle de l'écoute. En ouverture du *Book of Noise*, il écrit :

**Ceci est un portrait de votre ville. Écoutez-le attentivement. Peut-être n'avez-vous jamais vraiment écouté auparavant. C'est un concert fascinant et exaspérant de sons. Écoutez... Klaxons, sirènes, motos, camions, marteaux-piqueurs, scies électriques et machineries de construction, hélicoptères et avions à réaction. Un auditeur attentif conclura que ce sont là les instruments dominants de l'orchestre<sup>4</sup>.**

La transition se joue ici vers ce qui constitue l'apport décisif de Schafer : de la tension entre bruit et musique, entre exaspération et fascination devant l'orchestre de la ville moderne, naît le désir de former une auditrice ou un auditeur attentif, quelqu'un qui écoute peut-être pour la première fois, et qui entend vraiment la ville comme une symphonie. Car ensuite, le basculement nodal s'opère : cet agencement de sons devient un *portrait du lieu*.

<sup>3</sup> « Yes, but is it music? » (Schafer, 1969, n.p., notre traduction)

<sup>4</sup> « This is a portrait of your city. Listen to it closely. Perhaps you have never really listened before. It is a fascinating and exasperating concert of sounds. Listen... Horns, sirens, motorcycles, trucks, jack hammers, power saws and construction machinery, helicopters and jets. Any attentive listener will conclude that these are the dominant instruments in the orchestra. » (Schafer, 1970, p. 2, notre traduction)

Le modèle de la recherche n'est alors plus vraiment une analyse de type musical ; celle-ci devient plutôt une méthode au profit d'un projet plus ambitieux, celui d'une *écologie acoustique*. L'écologie est, dans la définition de Schafer, « l'étude des rapports entre les êtres vivants et leur environnement » (Schafer, 2010, p. 382) ; l'écologie acoustique rendra donc compte de la manière dont la configuration sonore d'un espace donné interagit avec les vivants qui l'habitent.

## Ici et ailleurs

C'est dans cette perspective, et avec cette ambition, que Schafer mit en place, en 1971, le *World Soundscape Project*, projet de recherche collectif au sein de la Simon Fraser University voué à établir une *géographie* de l'évolution des paysages sonores. L'approche de Schafer est historique, par l'importance accordée à l'opposition entre paysages sonores pré- et post-industriels ; mais elle est systématiquement spatialisée : on étudie l'évolution des agencements sonores dans des lieux spécifiques géographiquement et socialement situés, on compare les transformations des lieux par l'analyse de leurs sons, et au sein de chaque lieu on décrit avec précision la répartition spatiale des sons.

La publication la plus marquante du World Soundscape Project a sans doute été *The Vancouver Soundscape*, édité en 1973 sous la forme hybride d'un double disque (Schafer *et al.*, 1973) et d'un livre (Schafer, 1973). L'ensemble formait vraiment un « portrait de la ville », qui prenait en compte à la fois son histoire – sous la forme d'un vaste montage de citations de *earwitness accounts* (témoignages auditifs) recueillis dans diverses sources littéraires remontant au XVIII<sup>e</sup> siècle et s'étalant sur une vingtaine de pages – et son actualité, à travers des études rassemblant enregistrements de terrain et

des mesures, analyses, tableaux et synthèses. L'ouvrage présentait des propositions méthodologiques fortes, par exemple ce que Schafer nomme « carte isobel », présentant des courbes de niveau sonore (exprimé en décibels), sur le modèle des lignes d'altitude sur les cartes topographiques ou des isobares sur les cartes météorologiques [fig. 1]. Une véritable exigence de scientificité sous-tendait la procédure :

**Cette cartographie du bruit était réalisée de manière systématique : les chercheurs prenaient des douzaines de mesures directes de niveau sonore en utilisant des sonomètres portatifs, notant les valeurs tout en marchant le long des sentiers du parc. Ils prenaient trois relevés à dix secondes d'intervalle, répétaient le processus tous les cent mètres, et faisaient la moyenne afin de déterminer les contours des lignes isobel<sup>5</sup>.**

C'est donc bien une représentation objective du lieu par le son que visait le dispositif. Le parc Stanley, sur les bords de la baie Burrard au centre de la ville, se trouvait ainsi décrit en zones d'intensité sonore et, là où les lignes s'accumulaient, en pentes descendant vers le silence ou montant vers le bruit, à l'approche notamment des routes et des carrefours. À la fin du livre, un autre plan, ressemblant à un dessin d'enfant, exhibait une « marche sonore » (« *soundwalk* »), autre méthode adoptée par les chercheuses et chercheurs du World Soundscape Project [fig. 2] :

**Quand vous emmenez vos oreilles dans une marche sonore, vous êtes à la fois public et *performeur* dans un concert de son qui se déroule continûment autour de vous. Grâce à la marche, vous êtes capable d'entrer en conversation avec le paysage<sup>6</sup>**

Le plan garde la trace de l'itinéraire et des points d'enregistrement, et s'accompagne du récit de l'auditeur ou de l'auditrice, de sa description de tous les sons entendus – à commencer par ses propres pas. Cette méthode des « *soundwalks*<sup>7</sup> » sera développée, par exemple dans le *European Sound Diary*, publié par Schafer en 1977 : les membres du World Soundscape Project voyagèrent entre février et juin 1975 à travers l'Europe, réalisant mesures et « marches sonores » d'Amsterdam en Écosse en passant par la Suède, l'Allemagne, l'Autriche, l'Italie, la France, l'Angleterre.



**Fig. 1 : Carte « isobel » de Stanley Park (© R. Murray Schafer, 1973)**

Schafer réalisa ensuite un long collage de leurs journaux de bord, collationnant en plus des études objectives leurs impressions sonores subjectives – par exemple, à Vienne le 19 mars Bruce Davis note (Schafer, 1977b, p. 33) : « The telephone in this restaurant (corner of Belvederegasse and Argentinierstrasse) rings quietly in its secret panel in the rear wall. It deserves a World Soundscape stamp of approval. » (Le téléphone dans ce restaurant (au coin de Belvederegasse et Argentinierstrasse) sonne doucement dans son placard secret dans le mur arrière. Il mérite un tampon *approuvé par le World Soundscape* [notre traduction].) ▶

<sup>5</sup> « This noise mapping was done systematically : researchers took dozens of measurements of sound levels using handheld noise meters, taking readings while walking along the park's footpaths. They took three readings every ten seconds, repeated this process every hundred yards, and averaged the values in order to determine the isobel contours. » (Ouzounian, 2020, p. 125, notre traduction)

<sup>6</sup> « When you take your ears for a soundwalk, you are both audience and performer in a concert of sound that occurs continually around you. By walking you are able to enter into a conversation with the landscape. » (Schafer, 1973, p. 71, notre traduction)

<sup>7</sup> Les « *soundwalks* » constitueront ensuite la base de projets artistiques, tels que *Go Your Gait* de l'artiste allemande katrinem (<https://www.katrinem.de/category/go-your-gait/>, consulté le 10 décembre 2021), dans lequel les « partitions » sont constituées par des plans de villes où sont dessinés des chemins de marche et d'écoute qu'elle appelle « *paths of awareness* » (chemins de conscience).

Toujours en 1977, le groupe mené par Schafer publiait, à partir du même voyage, *Five Village Soundscapes*, étude plus objective dans la forme, construisant des études détaillées des paysages sonores de cinq villages européens choisis pour leurs caractéristiques emblématiques : Skruv (sud-est de la Suède [fig. 3]), Bissingen (Allemagne), Cembra (Italie), Lesconil (Bretagne) et Dollar (Écosse). Les analyses comportaient des mesures d'intensité sonore dans divers lieux et à divers moments, mais aussi des descriptions de bruits entendus – cloches d'églises, moteurs, sifflements de trains ou d'usines, cornes de brume, canons, etc. – ou d'airs musicaux traditionnels. Elles incluaient des entretiens avec les habitantes et habitants, afin de définir les « attitudes des communautés vis-à-vis du paysage sonore » – avec toutes les difficultés inhérentes à ces recherches psycho-sociales. La démarche comparative et itinérante impliquait l'ambition de dessiner un ou des paysages sonores, une représentation à large échelle composée de cartes locales extrêmement denses, précises et multimodales. Le voyage européen entraîne aussi une nouveauté en décentrant la problématique du paysage sonore de la grande ville, et de la seule question de l'industrialisation ou de la pollution sonore. Ici, ce sont chaque fois des configurations très spécifiques qui sont en jeu : des histoires, des traditions et des cultures, mais aussi des faunes et des flores, des situations géographiques – bord de mer ou continent, proximité d'usines ou contexte rural, plaine ou montagne. La description par le son permet – et en fait *impose* – de prendre en compte ensemble l'intégralité de ces phénomènes et conditions [fig. 4]. Les faits sonores, produits par l'écoute attentive, mais singulièrement par l'écoute instrumentée et l'enregistrement, offrent une description précise du lieu dans toutes ses dimensions mêlées : le vivant (animaux, humains, insectes, arbres), le mécanique

(voitures, bateaux, usines), l'environnemental (vent, pluie), l'ensemble des cultures matérielles (matériaux des maisons, des outils, des routes) et immatérielles (cultures musicales), ainsi que la forme même de l'espace (immeubles, distances) qui en produit la résonance singulière.

La méthode s'accompagne de la production de sons, qui accompagnera les textes produits sous la forme de disques 33 tours ou de cassettes audio. Le matériel de prise de son de l'équipe du World Soundscape Project semble n'avoir pas évolué : quelques magnétophones stéréo Nagra IV-S et UHER 4200 ; des systèmes de micros AKG C-451E à condensateurs, équipés soit de

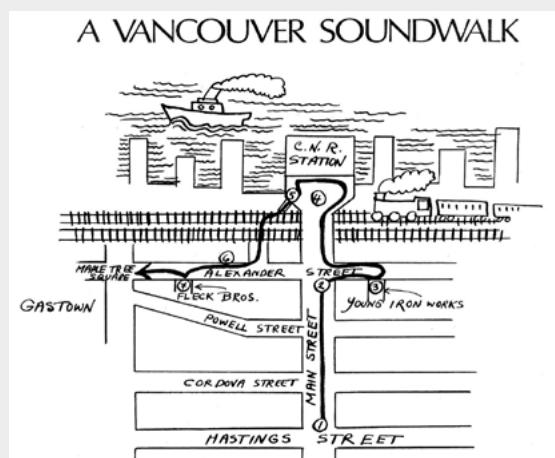


Fig. 2 : Une promenade sonore à Vancouver (© R. Murray Schafer, 1973)

capsules cardioïdes CK-1, soit de micros canons CK-9 ; des micros cardioïdes dynamiques AKG D202ES ; et des casques d'écoute<sup>8</sup>.

8

Il est listé en fin de *Five Village Soundscapes* en 1977, mais déjà à l'identique sur la couverture du disque *The Vancouver Soundscapes* en 1973. On peut s'étonner de ne pas voir de micros omnidirectionnels figurer dans cette liste.

## Sonographie

Cet équipement technique se double d'un bagage conceptuel. Schafer met en effet en place un vocabulaire de « notation des paysages sonores », ce qu'il nomme « sonographie » (Schafer, 2010, p. 29). La première distinction établie se situe entre la *tonalité* et les *signaux*. La tonalité d'un paysage sonore fait référence à la tonique d'une composition musicale :

**Dans l'étude du paysage sonore, la tonalité est donnée par un son que l'on entend en permanence, ou assez fréquemment pour constituer un fond sur lequel les autres sons seront perçus. Ainsi, le bruit de la mer dans une communauté maritime, ou celui du moteur à combustion interne dans une cité moderne. La tonalité n'est souvent perçue qu'inconsciemment, elle conditionne cependant la perception des autres sons, les signaux. (Schafer, 2010, p. 385)**

Par contraste, le signal est « [t]out son sur lequel l'attention se porte ». Finalement, « les signaux se distinguent de la tonalité, de la même manière que la figure se détache du fond dans la perception visuelle. » (Schafer, 2010, p. 385)

Cette opposition, issue des arts plastiques et de la psychologie de la forme, entraîne une catégorisation des paysages sonores en deux grands types : *hi-fi* et *lo-fi*.

**Dans l'environnement hi-fi, le rapport signal/bruit est satisfaisant. Le paysage sonore hi-fi est celui dans lequel chaque son est clairement perçu, en raison du faible niveau sonore ambiant. [...] Dans un paysage sonore lo-fi, les signaux**

**acoustiques individuels se perdent dans une surpopulation de sons. (Schafer, 2010, p. 77)**

L'écologie acoustique devient alors non seulement l'étude des relations des sons avec l'environnement naturel et humain, mais aussi celle de l'écosystème sonore, des interactions des sons entre eux. Dans le paysage hi-fi, les sons préservent leur existence pleine, leur vie propre. L'environnement lo-fi au contraire les agresse, les déforme, les appauvrit, les absorbe. Ces termes portent une hiérarchie plus ou moins explicite, qui est politique. « La campagne est généralement plus hi-fi que la ville, la nuit l'est plus que le jour, le passé plus que le présent. » (Schafer, 2010, p. 77) Dans un environnement sonore hi-fi, il y a de l'espace, car « les sons se chevauchent moins fréquemment ; la perspective existe, avec un premier et un arrière-plan » (Schafer, 2010, p. 77). Dans le lo-fi, les effets de masque aplatissent et confondent : « La perspective s'évanouit. À un carrefour, dans une cité moderne, la distance est abolie, seule reste la présence. » (Schafer, 2010, p. 77) Pour rester perceptible et signifiant, un signal devra alors être plus fort, amplifié jusqu'à couvrir le fond sonore – et ainsi, le volume général tend à augmenter démesurément.

Il y a donc une sorte d'ambivalence dans le système de Schafer. D'un côté, le signal s'oppose à la *tonalité* du paysage sonore. Celle-ci constitue une dimension positive : non seulement c'est un élément musical, mais elle est de plus caractéristique du lieu, elle le spécifie. Mais d'un autre côté, le signal vient s'opposer au *bruit*, dans un sens qui renvoie à la fois au vocabulaire du son et à celui de la théorie de l'information : le signal est ce qui porte la signification, le bruit étant ce qui non seulement reste absolument insignifiant, mais de plus parasite la communication. Où la tonalité est ►

qualitative, le bruit est quantitatif, et si la première est le fond où viennent se détacher les figures sonores, le second empêche leur perception et abolit l'espace même.

On voit en tout cas combien le projet de R. Murray Schafer, qui fut à la fois théorique et pratique, à la fois social – lutte contre la pollution sonore – et musical, à la fois environnemental et physiologique, s'établit sur un postulat qui restera majeur pour les *sound studies* et l'écologie acoustique : les sons sont profondément inscrits dans un lieu, au point que la description d'un paysage sonore donné permet de comprendre à la fois la géographie, l'histoire, la culture, l'écosystème, la géologie du paysage physique – et de les comprendre *ensemble*.

### **Prolongements : cartographies et *field recording***

Le World Soundscape Project peut être considéré comme la matrice de nombreuses recherches artistiques ultérieures, qui pour la plupart partageront cette tension fondatrice entre vocation descriptive et formalisation musicale, entre écoute d'un lieu précisément situé et mise en œuvre esthétique. Ces pratiques doivent aussi à certaines tendances de la *musique concrète* développée notamment par Pierre Schaeffer dès les années 1940 ; elles se sont développées parallèlement à la diffusion du matériel d'enregistrement portable, notamment des enregistreurs magnétiques puis numériques, à partir des années 1950 – Nagra, Uher déjà mentionné, Perfectone, Stellavox, etc.

Parmi les projets situés le plus explicitement dans la suite des travaux de Schafer apparut, en 1983, l'association

française Acirène, rassemblant des professionnels du son, mais aussi de la culture, de la médiation et du patrimoine. L'enjeu était la sensibilisation à l'environnement sonore. Dès sa fondation, l'association mit en place un « Conservatoire d'échantillons sonores du paysage », fondé sur l'enregistrement et l'archivage de paysages sonores répartis sur un territoire donné. De cette démarche émergea un « inventaire de sites auriculaires remarquables ». Ceux-ci purent ensuite être valorisés de différentes manières, avec le soutien actif des populations locales. Des cartes touristiques IGN furent imprimées qui mentionnaient des « points d'ouïe » en plus des « points de vue ». Ces sites pouvaient voir la création de concerts ou de sculptures sonores, ou être reliés par des performances-promenades le long de « parcours audio-sensibles » permettant l'écoute de « paysages sonores partagés » (Tixier *et al.*, 2020). La démarche d'Acirène prolongeait clairement l'« écologie acoustique » envisagée par Schafer, articulant un travail d'enregistrement et de préservation de paysages sonores en transformation permanente – sinon en voie de disparition – avec une mise en jeu esthétique qui se conçoit aussi comme un moyen de diffusion d'une pédagogie de l'écoute au niveau local.

Parallèlement, des artistes de plus en plus nombreux se tournèrent vers l'enregistrement de terrain, le *field recording*, comme méthode d'investigation sonore et musicale de lieux et comme méthode de création. Les enregistrements sont ensuite montés pour constituer un disque commercialisé, voire une installation. Le montage peut transformer la matière sonore de manière importante, soit en superposant différentes couches sonores, soit en modifiant électroniquement les enregistrements ; il peut aussi se restreindre à un simple bout à bout de témoignages sonores inaltérés, seulement sélectionnés pour leur beauté

propre ou pour leurs caractéristiques exemplaires. Certaines de ces œuvres, conçues et diffusées comme des pièces musicales ou en tout cas à valeur esthétique, se présentent d'ailleurs comme des cartographies sonores de lieux précis : par exemple, *A Sound Map of the Hudson River* (1989) d'Anne Lockwood, enregistré sur quinze sites au long du fleuve Hudson, de juin à décembre 1982 ou, plus récemment, *The Hebrides Suite* de Cathy Lane, portrait des îles Hébrides initié en 2008 et qui donna lieu à un disque (2013), des performances, des expositions, des installations et des publications [fig 5]<sup>9</sup>.

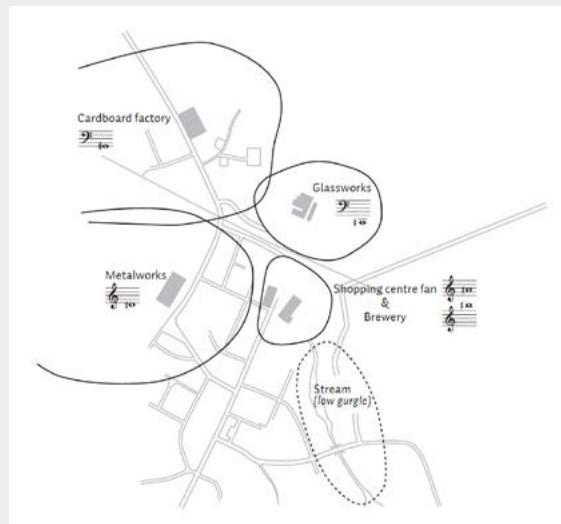
En 2012, le musicien anglais Peter Cusack publia un album intitulé *Sounds from Dangerous Places*, constitué d'enregistrements réalisés sur des zones marquées par des menaces ou des catastrophes écologiques, de Tchernobyl aux champs d'extraction de pétrole d'Azerbaïdjan, des abords de centrales nucléaires aux décharges à ciel ouvert. Il expliquait :

**La plupart des endroits où je suis allé récemment sont choisis pour les problèmes environnementaux qui les entourent. [...] Je ne les choisis pas pour le son, mais à cause des problèmes. Ces projets utilisent le son comme méthodologie pour explorer les problèmes, tout au moins c'est comme ça que je le vois aujourd'hui<sup>10</sup>.**

Ce « journalisme sonore » (*sonic journalism*) formulé par Cusack n'empêche pas un grand nombre des enregistrements d'être manifestement orientés par une sensibilité musicale – dès la première piste, l'agencement des sortes de gazouillements frénétiques d'un compteur Geiger avec la voix qui en égrène les chiffres et les sons naturels (souffles de vent et bruits de

pas dans les herbes) en arrière-plan dessine un portrait assez bouleversant de ce site exceptionnel qu'est devenu le village fantôme de Kopachi, évacué après la catastrophe nucléaire de Tchernobyl. Mais même musical, c'est l'exigence du lieu qui prime. Le son y est bien une *méthodologie de description des lieux*.

Avec l'émergence des dispositifs numériques, ces cartographies sonores ont pu se greffer sur de nouveaux



**Fig. 3 : Tonalités dominantes à Skruv, Suède (© R. Murray Schafer, 1977)**

modes d'échanges et de circulation. On a vu alors apparaître des entreprises parfois participatives, de plus ou moins grande ampleur, adossées à des cartes interactives de type Google maps.

En 1998, Peter Cusack commença le projet *Favourite Sounds* : ilregistra ses « sons préférés » de la ville de Londres, et demanda à des habitants de proposer

<sup>9</sup> Voir <https://hebrides-suite.co.uk> (consulté le 10 décembre 2021).

<sup>10</sup> [M]any of the places I've been to recently are chosen for the environmental issues surrounding them. [...] I'm not choosing them for the sound, but because of the issues surrounding them. These projects use sound as a methodology for exploring the issues, at least that's the way I look at them now. (Lane & Carlyle, 2013, p. 193, notre traduction)

les leurs – un disque fut ensuite publié (Cusack, 2001). Le programme est centré sur l'expérience urbaine, mais il est explicitement positif : Cusack renverse la condamnation d'une ville dominée par des volumes sonores agressifs, en un désir d'*habiter*, au sens le plus fort, la bruyante cité. Le projet se développa sur d'autres villes : Chicago (2004), Pékin (2005, voir Cusack, 2007), Manchester et Prague (2008-2009), Berlin (2013, voir Cusack, 2013), Hull (2017), etc. Ces projets variés se trouvent reliés par un site internet (favourite-sounds.org) où chaque son peut s'entendre à partir d'une carte Google interactive. C'est depuis le *lieu* donc, ou du moins depuis ses coordonnées géographiques, qu'on peut avoir accès au son. Ces divers projets concertés sont complétés par des enregistrements envoyés par des participants bénévoles, partout dans le monde.

D'autres projets de ce genre existent, par exemple Radio Aporee développé par Udo Noll à partir de 2006 (aporee.org), ou le London Sound Survey rassemblant depuis 2009 les enregistrements de Ian Rawes (soundsurvey.org.uk). Chaque fois, l'idée est de laisser entendre des sons *situés*, et de constituer un portrait de l'état du monde par le son, agaçant l'*espace commun* figuré par l'imagerie satellite avec des *situations locales* toujours complexes, historiquement déterminées, perçues à travers les paysages sonores.

## Musiques situées

Ces questions ont en retour profondément marqué les musiciennes et musiciens. Bien sûr, la production musicale n'a jamais cessé d'être travaillée par la problématique du lieu, de l'espace, de la réverbération. Les joueurs de luth savaient se placer de manière à

être entendus de leurs auditeurs, comme les facteurs d'orgue savaient qu'il aurait été absurde de concevoir l'instrument sans prendre en compte l'architecture de la cathédrale où il trouverait sa place définitive. Mais les bouleversements dans la conception même de ce qui était musique – et bruit – engendrés par les avant-gardes du vingtième siècle auront amené à radicalement réviser l'importance de l'inscription du son dans un lieu. La salle de concert close et silencieuse, le studio d'enregistrement méticuleusement insonorisé, purs espaces abstraits entièrement « hors-sol », se seront peu à peu ouverts, laissant entrer les sons du dehors, laissant s'entremêler bruit et musique. La quatrième des histoires d'une minute racontées par John Cage dans *Indeterminacy* en 1959 en est assez exemplaire :

**Un jour alors que les fenêtres étaient ouvertes, Christian Wolff jouait l'un de ses morceaux au piano. Les sons de la circulation automobile, les sirènes de bateaux, s'entendaient non seulement pendant les silences dans la musique, mais, étant plus forts, se trouvaient plus faciles à entendre que les sons du piano eux-mêmes. À la fin, quelqu'un demanda à Christian Wolff de rejouer le morceau avec les fenêtres fermées. Christian Wolff répondit qu'il en serait ravi, mais que ce n'était pas vraiment nécessaire, car les sons de l'environnement ne constituaient d'aucune manière une interruption de ceux de la musique<sup>11</sup>.**

Si Schafer évoque 4'33" de Cage dans *Le paysage sonore*, cette anecdote semble presque plus proche de ses préoccupations, mais aussi des tensions à l'œuvre dans son travail. Ce que voulait l'auditeur, c'était une interprétation dans un paysage sonore hi-fi, où chaque note serait clairement audible pour elle-même,

<sup>11</sup>

« One day when the windows were open, Christian Wolff played one of his pieces at the piano. Sounds of traffic, boat horns, were heard not only during the silences in the music, but, being louder, were more easily heard than the piano sounds themselves. Afterward, someone asked Christian Wolff to play the piece again with the windows closed. Christian Wolff said he'd be glad to, but that it wasn't really necessary, since the sounds of the environment were in no sense an interruption of those of the music. » (Cage & Tudor, 1959, notre traduction)

sans fond perturbant. Ce que Wolff offre avec les fenêtres ouvertes, ce n'est pas une performance lo-fi, contrairement à ce que pourrait avancer Schafer, mais une intégration du fond dans la figure, une musicalité du brouillage. Wolff réintègre le fond sonore comme une réelle *tonalité*, comme un espace situé au sein de laquelle la musique va venir se déployer de manière singulière, spécifique de ce lieu et de ce moment, impossible à répéter. L'auditeur cherchait un piano intemporel sous les contingences du lieu, malgré lui ; le pianiste Wolff entendait les moteurs et les sirènes comme il entendait les notes de son instrument. Il n'y avait, pour lui en tout cas, pas de différence. La question ne porte plus sur le silence comme dans *4'33"*, mais bien sur le bruit, la tonalité, le lieu.

Les recherches liées à ces possibilités nouvelles et à ces interrogations persistantes se formulèrent de différentes manières. L'une des directions importantes fut l'exploitation musicale de l'interaction concrète entre les corps des interprètes et le lieu physique de la performance. Pendant trois jours, du 9 au 11 mai 1977, le saxophoniste Peter Brötzmann et le percussionniste Han Bennink parcoururent en voiture la Forêt noire, entre Donaueschingen et le barrage de Schwarzenbach, munis de divers instruments et d'un appareil photo. Ils enregistrèrent là, en extérieur, un album publié la même année sur le label FMP, *Schwarzwaldfahrt*. Sur les notes de pochette, on pouvait lire :

**Sur ce disque les musiciens jouent : clarinette en mi bémol, clarinette en si bémol, clarinette basse, saxophone soprano, saxophone alto, apeaux, alto, banjo, cymbales, bois, arbres, sable, terre, eau, air. Les enregistrements ont été réalisés par Brötzmann/Bennink sur Stellavox<sup>12</sup>.**

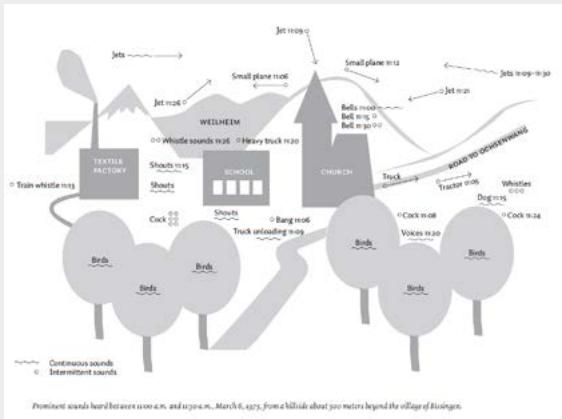
Les deux hommes font partie des personnalités majeures du *free jazz* et de l'improvisation libre en Europe ; ils ont joué ensemble à de maintes reprises à partir de la fin des années 1960, au sein de formations d'ampleurs diverses ainsi qu'en duo. Mais ce disque-ci est un peu différent : le son du saxophone est transformé par l'espace ouvert, par le froid qui crispe le corps de Brötzmann, par le vent qui traverse les instruments. Les chants des oiseaux, l'écoulement d'un ruisseau, occupent la bande sonore, et Bennink joue avec ses baguettes ou ses mains ou d'autres objets, tape sur des arbres, sur la surface d'un lac ou lance des cailloux dans l'eau sur un rythme parfois irrégulier. Déjà présente ailleurs dans leur musique, la dimension ludique, presque enfantine jusque dans sa gravité, prend ici un aspect d'exploration d'une nature à la fois familière et sauvage. L'album finit par dresser un portrait de la Forêt noire elle-même. La présence humaine y est sensible ; mais la quasi-absence de la voix, la préférence pour les bruits et les sons d'instruments à cordes ou à vent, intègre les humains dans l'ensemble du paysage sonore naturel plutôt que de les en détacher. Tout se passe comme s'ils voulaient enrichir la *tonalité* du paysage plutôt qu'y imposer des *signaux* qui lui seraient étrangers. ►

**12**

« On this record the musicians are playing : e-flat clarinet, b-flat clarinet, bass-clarinet, soprano saxophone, alto saxophone, birdcalls, viola, banjo, cymbals, wood, trees, sand, land, water, air. The recordings were made by Brötzmann/Bennink on Stellavox. » (Brötzmann & Bennink, 1977, notre traduction)

## Espèces d'espaces

Mais le rapport au lieu peut se jouer aussi sur un plan plus abstrait : celui du lien à la forme même de l'espace où se produit la musique, à ses caractéristiques physiques de résonance. Le 30 avril 1978, une autre figure



**Fig. 4 : Principaux sons entendus entre 11h et 11h30, le 6 mars 1975, depuis une colline située à environ 500 m du village de Bissingen (© R. Murray Schafer, 1977)**

nodale des musiques improvisées européennes, le saxophoniste anglais Evan Parker, enregistra son second album solo, *Monoceros*, à Wyastone Leys, propriété située à la frontière entre Angleterre et Pays de Galles où s'était installée depuis trois années la compagnie Nimbus Records, spécialisée dans l'enregistrement musical le plus exigeant. Grâce à ses ingénieurs du son, il avait pu adopter un procédé technique particulier, l'enregistrement direct sur disque. On lisait sur la pochette :

***Monoceros 1-4* est un ensemble d'improvisations au saxophone soprano enregistré par le procédé**

de la gravure directe. C'est une méthode d'enregistrement qui évacue la bande magnétique, par conséquent le ronflement, la nécessité de passer par des systèmes de réduction de bruit, et le montage sont tous éliminés. Le but dans la manière de placer l'unique microphone stéréo utilisé a été de trouver la position dans laquelle le son du saxophone dans cette pièce serait reproduit de la façon la plus convaincante. L'ambiance est l'acoustique naturelle de la pièce<sup>13</sup>.

La gravure directe correspond à une volonté de minimiser le souffle engendré par le dispositif électrique d'enregistrement, ce qui permet en retour d'ouvrir le rendu final à une représentation plus fine de la texture sonore et de la réverbération caractéristique de cet espace. Plaçant à l'arrière-plan la machinerie, c'est le lieu, la pièce, qui se trouve mis en avant. Par ailleurs, la gravure directe renvoie à un point éthique fondamental pour Parker, qui prolonge plusieurs débats importants à l'époque sur le statut à donner à l'enregistrement dans la pratique de l'improvisation libre<sup>14</sup>. Cette technique sans bande magnétique implique que le support n'est pas manipulable : on ne peut pas couper et coller dans un disque, il faut tout prendre ou tout rejeter en bloc.

C'est dans ce contexte d'affirmation d'une exigence technique radicale en lien avec une certaine idée de ce que doit être la musique improvisée qu'apparaît la question de l'espace. « Le son du saxophone dans cette pièce » : le déterminant est le seul mot souligné de ces notes de pochette. S'il s'agit d'éliminer le souffle magnétique ou ses systèmes de correction, l'enjeu n'est donc pas pour autant d'obtenir le son abstrait d'un saxophone pur, non médiatisé. Il s'agit que l'instrument soit entendu exactement dans le paysage sonore dans lequel il a été joué. Celui-ci peut être minimal,

<sup>13</sup>

« *Monoceros 1-4* is a set of improvisations on soprano saxophone recorded by the direct-cut process. This is a recording method which by-passes magnetic tape, consequently tape-hiss, the need for noise reduction systems and editing are all eliminated. The aim in placing the one stereo microphone used was to find the position in which the sound of the saxophone in that room was reproduced most convincingly. The ambience is the natural acoustic of the room. » (Parker, 1978, notre traduction)

<sup>14</sup>

Voir notamment sur cette question GRUBBS D., *Records Ruin the Landscape : John Cage, the Sixties, and Sound Recording*, Durham & London, Duke University Press, 2014, 248 p. (notamment, sur les musiques improvisées, le chap. 4, « The Antiques Trade : Free Improvisation and Record Culture », p. 104-134).

extrêmement « hi-fi » pour reprendre la notion de Schafer, presque complètement silencieux, il n'en reste pas moins marqué par une certaine « acoustique naturelle » qui doit former l'ambiance de la musique enregistrée car elle a été l'ambiance qui a accompagné la performance. Et il n'en reste pas moins qu'il aura donné au son du saxophone, joué là, une texture particulière, qui aura aussi alimenté la confrontation entre instrument et musicien qui forme la trame profonde de l'improvisation en solo. Si Parker a joué comme ça à ce moment-là, c'est parce que le saxophone sonnait comme ça et que l'espace résonnait comme ça. Ainsi, l'œuvre est finalement un échange à trois entre musicien, instrument et milieu.

La question du rapport à l'espace met en jeu une réflexion techno-esthétique de longue haleine chez le saxophoniste anglais. Il y revient par exemple lors d'une conférence donnée en 2017 :

**Comment la physique du comportement d'une anche est-elle affectée par l'acoustique de la pièce dans laquelle les vibrations de l'anche sont amplifiées à travers la colonne d'air conique d'un saxophone ? Récemment, j'en suis venu à penser qu'il y a un effet de couplage qui rend très difficile de décider de ce qu'est une « bonne anche » isolément de l'acoustique dans laquelle elle est jouée<sup>15</sup>.**

L'instrument est inséparable de l'acoustique de la pièce où il est joué, et chacune de ses parties – leur forme et leur matière – interagit avec tout le paysage sonore, jusqu'à l'impossibilité de les distinguer. Objets, corps et espace, lieux et intentions artistiques, textures et mélodies, rien n'est plus séparable. La cohérence profonde de cette conception chez Parker tient à sa

définition même du saxophone comme « colonne d'air conique », conception dont il rappelle régulièrement l'importance pour son travail et qu'il a qualifiée, lors d'une autre conférence, de « révélation dont j'ai encore du mal à communiquer la force » (Parker, 2018, p. 26). Car si le saxophone n'est plus d'abord un assemblage de matières dures manipulées par une volonté humaine, mais d'abord « un tube fermé qu'il s'agit d'ouvrir de différentes manières (de la même façon qu'on peut considérer qu'il s'agit d'un tube ouvert qu'il faut fermer) » (Parker, 2018, p. 26), alors ce qui fait musique est avant tout l'air même du lieu, modulé par les passages à travers le tube et ses ouvertures. L'instrument devient un médiateur par lequel l'espace entre en résonance avec lui-même. Il n'est plus le moyen d'expression d'une subjectivité, mais un dispositif de description des résonances spécifiques d'un certain point de l'univers, et de leurs perceptions par un corps humain et par la membrane d'un microphone placé en un autre point.

Cette capacité d'un instrument, peut-être plus particulièrement du saxophone, à s'inscrire dans un lieu, à en épouser avec précision la forme de l'espace, a continué d'être explorée et déployée. Les 3 et 4 septembre 1993, le saxophoniste suisse Werner Lüdi enregistra une série d'improvisations dans les immenses cavités de béton du barrage de Lucendro. Construit entre 1942 et 1947 au col du Gothard, dans la région du Tessin, Lucendro est le seul barrage suisse à être complètement vide : conçu à une époque où le béton était rare, il est constitué d'une série de dix-sept chambres. Dans les notes de pochette du disque qui résulta du processus, intitulé simplement *Lucendro*, le musicien précise le dispositif d'enregistrement :

**Dans la quatrième chambre du barrage, le ►**

<sup>15</sup>

« How is the physics of the reed behaviour affected by the acoustics of the room in which the reed's vibration are amplified by the conical air column of a saxophone ? Lately I have come to think that there is a coupling effect that makes it very difficult to decide what is a « good reed » in isolation from the acoustic in which it is played. » (Parker, 2017, p. 22, notre traduction)

**studio le plus simple qu'on puisse imaginer. Un microphone stéréo pend d'une échelle. Un enregistreur repose sur une caisse de bière<sup>16</sup>.**

Comme Parker pour *Monoceros*, Lüdi et son ingénieur du son Christoph Finé Renfer ont utilisé un seul micro immobile pour la prise de son. Le dispositif « le plus simple qu'on puisse imaginer » semble ainsi la meilleure option pour rendre compte du caractère spécifique de l'endroit choisi. Le résultat cependant est radicalement différent : on sent chez Lüdi que le conflit avec l'espace est réellement le point focal de l'enquête. On entend dans les sons émis que le musicien écoute leur retour, qu'il réagit à la profondeur de l'énorme écho qui les travaille, les prolonge, les épaissit. La différence tient à la taille de la cavité bétonnée mais aussi à celle de la « colonne d'air conique » : Parker est attaché au saxophone soprano, tandis que Lüdi joue ici du baryton et de l'alto. Le son monumental paraît arraché physiquement à l'air, aux poumons du musicien et à l'ampleur de cet espace de cathédrale vide qu'il s'agit de remplir – arraché aussi au fond d'humidité ruisselante coulant contre les murs, qui forme une tapisserie sonore, une *tonalité* dirait Schafer, sur laquelle les *signaux* épais du saxophone forment des figures abstraites. Dès le premier morceau, *Jenissei*, la prise de son rend compte de la réverbération, si longue que les aigus produisent en écho comme un chant ou un cri selon le volume, voire comme un accompagnement d'orgue. Mais le micro nous place aussi tout proche du corps de Lüdi : on entend les clés jouer, comme on entend parfois le souffle traversant la gorge et le tube de métal – surtout quand les bribes de mélodies descendent vers les graves et les bas volumes. La dernière pièce, *Solovki*, fait même sa matière de cette respiration montante ou finissante : c'est un drame du souffle, comme s'il fallait que l'on assiste à la fin de la musique sous la

forme de l'amplification jusqu'à l'absolu de l'expiration de l'humain. Entre les deux, une piste comme *Jürmala* va jusqu'à prendre sa matière dans les effets de bouche dans le saxophone, crachements ou vomissements de son, toujours proches de l'extinction même si surgissent par instant des éclats de fort volume, évoquant un monstre agonisant. L'espace gigantesque produit finalement une exploration intérieure : c'est le corps de Lüdi qui s'entend dans toute sa dimension organique et vivante, plutôt que le métal du saxophone ou le béton des murs humides. Le son de l'enregistrement nous entre à son tour dans le corps, renvoyant à la force physique que le saxophoniste aura dû déployer pour souffler ainsi, mais aussi en retour à la force physique exercée par le son sur le corps du musicien.

Ainsi, la prise de son engage ensemble le lieu et le corps. L'unique micro stéréo et l'enregistreur portable témoignent du formidable maelström qui emporta ensemble, en cette toute fin d'été 1993, la vaste cavité de béton humide d'un barrage suisse et le corps soufflant d'un musicien luttant contre l'espace. L'écho établit une distance par la longueur des sons, une persistance et des harmoniques spécifiques ; mais il nous place aussi dans un intérieur, dans un lieu à la fois grand et clos. C'est depuis cette intériorité matérielle que l'exploration de Lüdi – atteint de claustrophobie (Lüdi, 1996, p.17) – se tourne vers sa propre intériorité organique. L'espace alors est le contraire de la Forêt noire de Brötzmann et Bennink : là où tout s'ouvrait, dans un espace naturel sans échos, sans murs et pleins d'animaux, de choses, d'éléments, ici tout est clôture et nudité. Le paysage sonore constituait un lieu plein ; il est maintenant un espace presque vide, travaillé seulement par le ruissellement de l'eau et la hauteur de la voûte.

<sup>16</sup>

« In der vierten Kammer das denkbar kargste Studio. Ein Stereomikrophon hängt von einer Leiter. Auf einem Bierharass steht ein Aufnahmegerät. » (Lüdi, 1996, notre traduction)

## Espaces résonants

Un autre projet, développé quelques années plus tard, se trouve articuler ensemble ces problématiques de l'ouvert et de l'écho, de l'inscription dans des lieux, quoique sur de toutes autres bases techniques, même s'il s'agit encore une fois d'un saxophone. En juin 2006, le saxophoniste John Butcher et l'artiste sonore Akio Suzuki furent invités à improviser, ensemble et en solo, dans divers lieux exceptionnels lors d'une tournée en Écosse : les mégalithes de Stenness, l'immense cylindre métallique du réservoir de pétrole abandonné de Lyness, la grotte de Smoo où s'écoule une cascade, la glacière industrielle de Tugnet, le réservoir d'eau aujourd'hui vide de Wormit construit à même une colline en 1923, etc. Le titre du projet, *Resonant Spaces*, est devenu le titre du disque édité par Butcher en 2008 à partir de ses performances en solo<sup>17</sup> : le principe tenait essentiellement à la provocation d'une confrontation entre une pratique sonore fondée sur l'improvisation, la réactivité, l'échange, et des espaces « résonants », définis par des caractéristiques acoustiques radicales.

Dans un entretien de 2021, Butcher insiste encore sur cette importance du rapport avec le lieu envisagé comme une relation musicale, un mode d'interaction continu, de performance commune similaire à celui que l'on pourrait établir, au sein d'un groupe de musique improvisée, entre musiciens :

**Un espace inhabituel donne quelque chose avec lequel et contre lequel travailler. [...] Dans des environnements inhabituels, il faut être à l'aise avec le fait de faire de la musique en partenariat avec l'espace et non pas seulement importer un morceau de musique dans cet espace<sup>18</sup>.**

<sup>17</sup>

Akio Suzuki a également publié un album issu de ses performances (Suzuki, 2019). Enfin, les duos unissant les deux artistes sont parus sur un autre album encore (Butcher & Suzuki, 2017).

<sup>18</sup>

« An unusual space provides something to work both with and against. [...] In unusual environments you have to be comfortable with making music in partnership with the space and not just importing a piece of music into the space. » (Butcher, 2021, pp. 2 & 7, notre traduction)

<sup>19</sup>

« the music tonight was fully a product of the space [...] I wouldn't have played the way I did if I hadn't been in this space » (Kopf, 2006, p. 29, notre traduction)



Fig. 5 : *The Hebrides Suite* (© Cathy Lane, 2013)

Le lieu intervient en profondeur dans la production musicale, non pas uniquement en modifiant superficiellement la texture sonore par la réverbération, ou en perturbant l'écoute des auditeurs et des musiciens par l'imposition de « bruits de fond » ; la force acoustique du lieu est telle qu'il doit être compté comme une force active dans le processus de composition. Après son concert dans la grotte de Smoo, Butcher affirme :

**La musique ce soir était pleinement une production de l'espace. [...] Je n'aurais pas joué comme je l'ai fait si je n'avais pas été dans cet endroit<sup>19</sup>.**

La performance à Smoo joue de la réverbération dans cette vaste cavité naturelle, ainsi que du son de la cascade qui y coule, et des bruits produits par les enfants qui jouent durant le spectacle [fig 6]. Certains traits pourraient ressembler au jeu de Lüdi à Lucendro, bien que Butcher soit un saxophoniste très différent : travaillant surtout au saxophone soprano et un peu au ténor, parfois amplifiés, il s'oriente vers ►

des matières sonores moins épaisses, plus fluides et plus rêches à la fois. Mais selon les lieux, le mode d'interaction entre instrument et éléments, entre



**Fig. 6** : John Butcher jouant dans la grotte de Smoo (Écosse), enregistrement Ruari Cormack (© Garrard Martin)

organes et géographie, peut trouver de toutes autres formes. L'enregistrement de la performance parmi les pierres levées de Stenness, sur l'île principale des Orcades au nord de l'Écosse, a été, selon Butcher, majoritairement un échec. L'improvisation avait joué avec l'amplification du son du saxophone nécessaire afin qu'il parvienne à émerger des vents très forts qui traversaient le site. Mais les micros avaient échoué à la capturer. Un moment singulier pourtant a fonctionné et été conservé pour le disque. Ainsi que l'a raconté Biba Kopf :

**La performance de Butcher prend un tour encore plus impressionnant quand il se rend compte que le vent commence à jouer de son saxophone soprano amplifié sans qu'il y souffle lui-même.**

Il se met alors à sculpter en retour un solo désincarné à la beauté surnaturelle, en utilisant le saxophone pour prendre le vent à travers ses trous de notes, et produire ainsi des sons de flûte, tout en travaillant l'air piégé dans le tube en bougeant l'instrument et en appuyant sur les clefs<sup>20</sup>.

« Le vent et moi jouions tous deux l'instrument<sup>21</sup>», explique ailleurs le saxophoniste. La *Wind Piece* qui en résulte fait entendre un bouillonnement de souffles venteux, puissants et profonds, d'où émerge, comme en lutte, se détachant à peine, des sifflements doux qui finissent par se former en un flux modulé, aérien, riche d'une matière à la fois forte et fragile.

Sur le plan concret, les choix de prise de son étaient opposés au principe de l'unique micro stéréo adopté par Parker et Lüdi. Butcher expliqua :

**Techniquement l'enregistrement était plutôt basique, Ruari [Cormack] plaçant des micros monos autour de l'espace et devant moi. [...] Lorsque j'ai mixé l'enregistrement, j'ai choisi une petite sélection des micros pour capturer, aussi bien que je m'en souvenais, le son de chaque espace. La dispersion des micros était particulièrement utile pour toutes les pièces où je me déplaçais, comme dans la grotte de Smoo<sup>22</sup>.**

Pour *Monoceros*, on l'a vu, Evan Parker a mis en place une configuration de prise de son solide et immuable, vouée à rendre de la meilleure manière possible la sonorité de ce saxophone (joué par cet homme) dans cette pièce. Cela fait, le musicien délègue à un système technique rectiligne le soin d'inscrire la trace du moment et du lieu sur le disque, sans qu'aucune modification ne

**20**

« Butcher's performance takes an even more impressive turn when he notices the wind starting to play his amplified soprano without him blowing into it. He proceeds to reverse-sculpt a disembodied, eerily beautiful solo by using his horn to catch the wind across its tone holes to produce fluting sounds, while working the air trapped in the tube by moving the instrument and pressing the tabs. » (Kopf, 2006, pp. 27-28, notre traduction)

**21**

« The wind and I were both playing the instrument » (Loriot, 2021, p. 5, notre traduction)

**22**

« Technically the recording was quite basic, with Ruari [Cormack] placing mono mics around the space and in front of me. [...] When I mixed the recording I chose a small selection of the mics to try to capture, as well as I remembered it, the sound of each space. Having the mics spread out was useful for any pieces where I moved around, like in Smoo Cave. » (Butcher, 2021, p. 4, notre traduction)

puisse intervenir. La stéréo finale est celle produite par la machine microphone et les murs du studio, comme la temporalité finale est celle de la performance.

Dans *Resonant Spaces* au contraire, la matière captée n'est pas le son inscrit. Le disque est le résultat d'un processus de réinterprétation d'une matière excédentaire – une dizaine de pistes –, réinterprétation fondée sur un principe de fidélité dans la reconstruction du « son de chaque espace ». L'horizon est donc une reconstruction objective de la caractéristique acoustique du lieu, qui a elle-même joué un rôle si important dans la composition de la musique qu'elle en devient inséparable. Toutefois, cette reconstruction reste rattachée à une mémoire subjective, d'autant plus problématique que, ainsi que Butcher le remarque, l'écart n'est pas seulement entre l'événement et son souvenir, il est aussi d'emblée entre le son écouté qui produit la musique et le son capturé par les microphones :

**L'un des problèmes est que je réagis à ce qui arrive à mes propres oreilles, alors que les microphones répartis dans l'espace capturent une image diffuse avec tellement de réflexions. Les acoustiques sont très différentes dans des positions<sup>23</sup>.**

L'acoustique du lieu, si caractéristique et si importante, se décompose donc en une constellation d'acoustiques très différentes : chaque position au sein du site forme un lieu en elle-même. Les sons entendus par le musicien et ceux perçus par le public et par les appareils ne sont pas les mêmes. Face à cet agrégat de micro-lieux, le mixage intervient pour reconstruire un espace commun par modulation des points d'écoute. Là, le dispositif n'est pas fixé une fois pour toutes : on peut décider, au long de

la pièce, de donner plus de présence à l'un puis à l'autre des micros. L'espace de *Monoceros* est solide, matrice fixe sur laquelle vient se déployer la figure formée par la ligne du saxophone soprano ; celui de *Resonant Spaces* est fluide, sans contours précis, privilégiant la matière sonore. La spatialisation des sons se trouve intégrée aussi profondément dans leur texture que dans les processus de composition à l'œuvre ; mais ensuite, l'espace fait l'objet d'une remodelisation continue à partir de sources monophoniques, elles-mêmes dotées d'une spatialité restreinte mais réelle, sous la forme unique d'une perspective sonore rattachée au point d'écoute.

## En cinéma : le cas Huillet et Straub

Bien sûr, le cinéma s'est trouvé confronté à ces problématiques. Depuis l'émergence des premiers agencements de systèmes de prises de vues et de sons synchronisés, une des fonctions du cinéma a été précisément de localiser les sons, de les inscrire dans des sites, de les rattacher aux moments et aux lieux exacts de leur création. En son direct, l'image montre là où le son s'est produit. Elle offre une perspective visuelle encadrant le lieu, lui donnant sa cohérence plastique ; elle montre les murs qui créent la réverbération ou le vent dans les cheveux et les vêtements qui expliquent le souffle derrière les voix. Elle révèle le paysage visuel qui informe ce paysage sonore.

Les différences de dispositif avec la musique enregistrée engendrent tout de même des écarts qui ne tiennent pas seulement de l'articulation entre le sonore et le visuel. Les cas que nous avons évoqués jusqu'ici se tiennent tous au sein d'un système entièrement structuré par la stéréo : le support final, CD, vinyle ou cassette audio, doit ►

<sup>23</sup>

« One problem is that I'm responding to what's arriving at my own ears, and microphones out in the space capture a diffuse picture with so many reflections. The sonics are very different in different positions. » (Butcher, 2021, p. 2, notre traduction)

porter deux pistes et seulement deux. Lorsque Parker et Lüdi adoptent un unique microphone stéréo au départ, la chaîne de production dans son entier peut s'en tenir, du premier moment de l'enregistrement au dernier de l'écoute finale, à l'inscription et à la reproduction de ces deux pistes, inchangées. En cinéma, la salle constitue un dispositif d'écoute spécifique, à l'histoire complexe. Longtemps mono sur la base d'une seule enceinte placée derrière l'écran, elle est aujourd'hui en son multicanal, répartissant les sons dans l'espace de perception spectatorial d'une manière régulée et sophistiquée. L'adéquation entre prise et restitution des sons ne peut que difficilement se réaliser point à point – d'autant plus si l'on imagine que le film doit pouvoir s'adapter à des contextes d'exploitation différents, incluant des salles de dimensions et d'appareillages variés, mais aussi une édition DVD ou Blu-ray, elles-mêmes prévues aussi bien pour une télévision traditionnelle que pour un « home-cinema ». Or parallèlement, l'image a conservé un point de vue unique – à quelques exceptions près, que nous ne traiterons pas ici (split-screen ou stéréoscopie par exemple).

La problématique spécifique qui émerge ici est celle de l'articulation entre point de vue et point d'écoute<sup>24</sup>. Mais les disjonctions ou rapprochements entre ces deux points ne règlent pas la question de la forme de l'espace sonore – ni celle de l'espace visuel, pour laquelle il faudrait entrer notamment dans des considérations de découpage et de choix de focales. Danièle Huillet et Jean-Marie Straub figurent sans doute parmi les cinéastes qui auront le plus insisté sur le caractère crucial de cette question : l'espace doit être construit, le lieu doit être respecté.

À partir de *Klassenverhältnisse (Amerika/Rapports de classes, 1983)*, adaptation par le couple d'*Amerika*

de Kafka, Huillet et Straub adoptèrent le principe de découpage qu'ils nommèrent « point stratégique<sup>25</sup> » : il s'agit de découvrir pour chaque séquence – pour chaque espace – la seule et unique position de caméra qui vaudrait, inchangée, pour l'ensemble des plans. De petites variations sont possibles<sup>26</sup>, mais elles restent minimales. Il s'agit ainsi de respecter la forme même de l'espace, sa structure, en n'en faisant pas du « chewing gum » – comme le diront fréquemment Huillet et Straub – par de constants changements de points de vue. Dans certains films ultérieurs tournés dans un seul lieu – *Antigone (1991)* par exemple –, la caméra tiendra ainsi une unique position pendant toute la durée du long métrage.

Mais qu'en est-il du son au sein de cette politique straubienne de l'espace ? Deux principes fondamentaux structurent l'œuvre de Huillet et Straub. Le premier est le son direct, qui constitue pour eux dès *Nicht versöhnt (Non réconciliés, 1965)* un dogme absolu et essentiel. La bande sonore d'un film de Huillet et Straub n'est composée que de sons enregistrés sur le lieu et dans le moment même de la prise – à la seule exception de la musique, dans certains films. Rien n'est autorisé à interférer avec ce principe fondamental, qui fait de chaque plan un bloc d'images et de son rigoureusement interdépendants.

Le second principe est le choix systématique de la monophonie. Tous leurs films sont enregistrés avec des micros monos, et la bande sonore finale mixée en mono. Même lorsqu'ils adopteront le système Dolby afin de s'intégrer au mieux dans les dispositifs sonores dominants dans les salles occidentales, leurs films seront mixés en Dolby mono – c'est encore le cas par exemple de *Le Streghe (Femmes entre elles, 2008)*.

<sup>24</sup> Voir sur cette question Fiant, Hamery & Massuet, 2017.

<sup>25</sup> L'expression est de Straub : voir Bergala, 1984, p. 28 notamment.

<sup>26</sup> Voir Farocki, 1983, p. 243. Harún Farocki était l'un des acteurs principaux du film.

Sans doute le film le plus connu de Huillet et Straub aujourd'hui, *Chronik der Anna Magdalena Bach* (*Chronique d'Anna Magdalena Bach*, 1968) présente plusieurs pièces de Bach, jouées ou dirigées par Gustav Leonhardt – qui interprète le compositeur allemand – dans des lieux correspondant soit aux sites historiques de leur création, soit à des configurations similaires. Cette recherche historique quant aux contraintes spatiales des exécutions musicales fait bien sûr écho à la volonté de Leonhardt de jouer sur instruments anciens, revendication non évidente en 1968 et que le film contribuera à populariser. Il s'agissait de savoir comment sonne Bach avec les instruments de son temps, le film prolongeant cette question en se demandant comment sonnent ces instruments dans les espaces où, alors, ils étaient joués. La configuration musicien/instrument/lieu est essentielle pour le projet du film ; textures et résonances, ainsi montrées comme inséparables, y renvoient à la nécessité d'une histoire matérialiste de la musique.

Ce « parti pris des lieux » s'agence avec un ensemble de décisions fondamentales. Le film, donc, sera en son direct. Ensuite, les performances musicales seront filmées en un seul plan – les quelques exceptions envisagées au moment du découpage ont été abandonnées au tournage. En outre, précise Straub (Jousse, Vatrican, 1996, pp. 138-140), « dès notre première rencontre avec [August] Wenzinger, Leonhardt et [l'ingénieur du son Louis] Hochet, on a fixé un principe général : un micro par exécution. [...] On pouvait aller jusqu'à trois pour le chœur d'entrée de *Saint Matthieu*, mais pas quatre. » Enfin, tout est pris et enregistré d'emblée en mono. Ainsi, trois micros pour une seule piste sur le Nagra mono, cela impliquait que Louis Hochet mixât pendant la prise. « Pas de rattrapage possible. » Bien sûr, dans un tel contexte, les contraintes sur la prise de son – et sur les

musiciens – sont très importantes. « Le grand souci de Leonhardt, c'était la balance. La balance du chœur et de l'orchestre. La balance, en musique baroque, plus encore qu'ailleurs, doit être irréprochable. » (Jousse, Vatrican, 1996, pp. 139-140) Assurer la balance entre les éléments enregistrés lors d'une prise de son en direct avec un seul micro, implique une importante réflexion sur le choix du micro (ici des Neumann), sur son placement et son orientation dans l'espace physique où est exécutée la performance, par rapport aux musiciens et à la caméra, sur la disposition des musiciens dans le lieu, sur le lieu même où l'on jouera.



**Fig. 7 : *Chronik der Anna Magdalena Bach* (D. Huillet et J.-M. Straub, 1968, tous droits réservés)**

La matière sonore et musicale du film se trouve profondément structurée par ces choix, et *Chronik* fut parfois attaqué pour avoir rendu la musique concrètement inaudible. Prenons l'exemple « limite » donné par Straub, le chœur d'entrée de la *Passion selon Saint Matthieu* (BWV 244), exécuté dans la tribune d'orgue de l'église St-Wilhadi de Stade [fig. 7]. Le choral est ►

interprété par deux chœurs, placés à chaque extrémité de la tribune. Entre eux, répartis tout en longueur sur l'estrade allongée et étroite où tous doivent trouver place, se trouvent les deux orchestres et, au milieu de la tribune, l'organiste, les sept sopranos-ripienistes et Bach/Leonhardt dirigeant l'ensemble. La caméra a quant à elle été placée tout à l'extrémité gauche de la tribune, juste derrière le premier chœur. Même avec trois micros, c'est cette perspective sonore oblique qui est massivement adoptée. Nous entendons donc de beaucoup plus près le chœur n° 1, et de beaucoup plus loin le chœur n° 2. Sur les disques de l'époque, la stéréo est déjà utilisée : chaque chœur est isolé sur un canal, chacun ainsi « spatialisé » spécifiquement dans l'espace abstrait de l'auditeur, tout en respectant un équilibre neutre « idéal » entre les deux chœurs. Pour le spectateur de ce plan de 7'24", l'échange est dynamisé par la position physique de l'observateur/auditeur : les intensités, timbres, nettetés des voix du premier chœur sont différentes des voix du second. L'échange (premier chœur, proche : « Voyez ! » ; second chœur, lointain : « Qui ? » ; premier chœur : « Le fiancé. Voyez-le ! » ; second chœur : « Comment ? » ; premier chœur : « Comme un agneau. » Etc.) s'en trouve prendre une signification nouvelle, entendue autrement, par nous, ici, maintenant, dans une situation d'exécution et d'écoute singulière. Car, on peut le noter, la position de la caméra ne correspond pas à celle qu'aurait un spectateur assistant au concert dans l'église. Huillet et Straub placent leur caméra – et donc leur micro – là où l'on peut observer les musiciens au travail, pas là où l'auditeur est supposé être. C'est alors la forme de la tribune qui détermine les conditions matérielles de la prise de vue et de son, fondées sur l'exigence d'une inscription radicale de la musique dans le lieu de sa production, et d'une inscription de la forme de l'espace de fabrication sonore au cœur même de la texture musicale.

L'agencement d'une performance musicale avec un lieu, un point de vue et d'écoute, tel qu'il est réalisé ici, résulte du dispositif cinématographique, et des pratiques et conceptions de Danièle Huillet et Jean-Marie Straub. Il différencie les enregistrements et la perception de la musique dans *Chronik der Anna Magdalena Bach* de toute écoute musicale sur disque ou en concert : la musique y est produite différemment, enregistrée à partir de principes techniques et de présupposés idéologiques (politiques) différents, et entendue dans des conditions différentes. Pourtant, on y retrouve l'écho de préoccupations qui furent aussi, à leur manière, celles d'Evan Parker, de Werner Lüdi, de John Butcher.

Si l'idée de décrire des lieux par l'étude de leurs sons peut être datée de l'émergence des *sound studies* et singulièrement des travaux de R. Murray Schafer, les expériences artistiques comme les recherches historiques et théoriques n'ont cessé de reprendre ces questions sous des formats divers – disque, cartes interactives, performances, essais... À travers tout ce qui peut différencier ces approches, un point commun les unit : la reconnaissance du fait que les sons sont toujours *situés*, très profondément inscrits dans des configurations locales d'espace, de culture, de géologie, de climat. Ils en portent l'empreinte de part en part, au point qu'une description des sons, que leur écoute même, suffit pour que le site qui a contribué à les modeler s'en trouve *saisi* dans toute sa complexité. Pour l'écologie acoustique, les sons habitent les lieux, parmi les choses et au milieu des vivants : ils peuvent interagir ensemble, se nuire éventuellement, mais aussi construire ensemble des œuvres communes. La modernité musicale a fait entendre de maintes manières comment le paysage sonore se mêle à la musique, et devient musique lui-même – un paysage

qui à la fois inclut des sons et détermine comment ils sonnent. Si le son peut être spatialisé, c'est d'abord parce qu'il l'est toujours déjà.

### Bibliographie :

- > BERGALA, A. (1984), « Straub-Huillet, la plus petite planète du monde (Méthodes de tournage) ». *Cahiers du cinéma*, n° 364, octobre, pp. 27-31.
- > BUTCHER, J. (2021), « Resonant Spaces : Interview with Frantz Lorient ». *Recordedness*, July, [https://44995de9-b562-4a0f-b282-c8fdc42347a3.filesusr.com/ugd/32034b\\_b9f54a02b5a54742a9e5c486c7173695.pdf](https://44995de9-b562-4a0f-b282-c8fdc42347a3.filesusr.com/ugd/32034b_b9f54a02b5a54742a9e5c486c7173695.pdf).
- > FAROCKI, H. (1983), « Einfach mit der Seele, das gibt es nicht », entretien avec Jean-Marie Straub. *Filmkritik* n° 317, mai, pp. 242-247.
- > FIANI, A., HAMERY, R., MASSUET J.-B. (2017), *Point de vue et point d'écoute au cinéma. Approches techniques*. Rennes : Presses universitaires de Rennes.
- > JOUSSE, T., VATRICAN, V., avec BIETTE, J.-C. et LOSS, D. (1996), « Entretien avec Jean-Marie Straub et Danièle Huillet », in D. HUILLET et J.-M. STRAUB, *Chronique d'Anna Magdalena Bach*. Toulouse : Ombres, pp. 129-140
- > KOPF, B. (2006), « World of Echoes ». *The Wire*, n° 270, August, pp. 26-29.
- > LANE, C., CARLYLE, A. (2013), *In the Field : The Art of Field Recording*. Axminster : Uniformbooks.
- > OUZOUNIAN, G. (2020), *Stereophonica : Sound and Space in Science, Technology, and the Arts*. Cambridge (MA) & London : The MIT Press.
- > PARKER, E. (2017), « Practise/Practice/Praxis », in *Oscillate*, dir. Robert Stillman, Canterbury : Compost & Height.
- > PARKER, E. (2018), *De motu*. Nantes : Lenka Lente.
- > SCHAFER, R. M. (1969), *The New Soundscape*. Scarborough (Ontario)/New York : Berandol Music/Associated Music Publishers.
- > SCHAFER, R. M. (1970), *The Book of Noise*. Wellington : Price Milburn & Co.
- > SCHAFER, R. M. (dir) (1973), *The Vancouver Soundscape. World Soundscape Project Document No. 5*. Burnaby : Sonic Research Studio, Communication Studies Department, Simon Fraser University.
- > SCHAFER, R. M. (1977a), *The Tuning of the World*. New York : Knopf.
- > SCHAFER, R. M. (1977b), *European Sound Diary*. Burnaby : Sonic Research Studio, Communication Studies Department, Simon Fraser University.
- > SCHAFER, R. M. (1994), *The Soundscape : Our Sonic Environment and the Tuning of the World*. Rochester : Destiny Books.
- > SCHAFER, R. M. (2010), *Le Paysage sonore. Le monde comme musique*. Marseille : Wildproject, coll. « Domaine sauvage ».
- > TIXIER, N., AMPHOUX, P., CHELKOFF, G., LAROCHE, S., REGNAULT, C. & RÉMY, N. (2020), « Observatoires sonores. Origines, enjeux, formes, règles, usages ». <https://hal.archives-ouvertes.fr/hal-02919392>. ►

### Discographie :

- > BRÖTZMANN, P., BENNINK H. (1977), *Schwarzwaldfahrt*. FMP Records, FMP 0440.
- > BUTCHER, J. (2008), *Resonant Spaces*. Confront Records, confront 17.
- > BUTCHER, J., SUZUKI A. (2017), *Immediate Landscapes*. Ftarri, ftarri-987.
- > CAGE, J., TUDOR, D. (1959), *Indeterminacy : New Aspect of Form in Instrumental and Electronic Music*, reading by John Cage, music by David Tudor. Folkways Records, FT 3704.
- > CUSACK, P. (2001), *Your Favourite Sounds from London*. London Musicians' Collective, RESFLS1CD.
- > CUSACK, P. (2007), *Favourite Beijing Sounds*. Kwanyin Records/Sub Jam, kwanyin 022.
- > CUSACK, P. (2012), *Sounds from Dangerous Places*. ReR Megacorp, Berliner Künstlerprogramm des DAAD, ReR PC3&4.
- > CUSACK, P. (2013), *Favourite Berlin Sounds*. ReR Megacorp, ReR PC5.
- > LANE, C. (2013), *The Hebrides Suite*. Gruenrekorder, Soundscape Series, Gruen 127.
- > LOCKWOOD, A. (1989), *A Sound Map of the Hudson River*, Lovely Music, LCD 2081.
- > SCHAFER, R. M. *et al.* (1973), *The Vancouver Soundscape*. Ensemble Productions, EPN 186.
- > SUZUKI, A. (2019), *Resonant Spaces*. 901 Editions, 9ED007.

### BENOÎT TURQUETY

Diplômé de l'École nationale supérieure Louis-Lumière et docteur de l'Université Paris 8, Benoît Turquety est Professeur associé à la Section d'histoire et esthétique du cinéma de l'Université de Lausanne. Après un projet sur Bolex (2015-2019), il dirige maintenant un projet de recherche sur Nagra, financé pour quatre années par le Fonds national suisse de la recherche scientifique. Ses ouvrages *Inventer le cinéma. Épistémologie : problèmes, machines* (prix international Maurizio Grande 2015), et *Danièle Huillet et Jean-Marie Straub, « objectivistes » en cinéma*, tous deux publiés par L'Âge d'Homme, ont été traduits en anglais chez Amsterdam University Press. Ses autres ouvrages récents incluent *Medium, Format Configuration : The Displacements of Film* (Meson Press, 2019), *Des avant-dernières machines. Cinéma, techniques, histoire* (co-dirigé avec Selim Krichane, L'Âge d'Homme, 2020) et *Le « Direct » et le numérique. Techniques et politiques des médias décentralisés* (co-dirigé avec Caroline Zéau, Mimésis, 2022).

*Benoît Turquety is Associate Professor in the Film History and Aesthetics Department of the University of Lausanne. He is currently leading an SNF research project on Nagra recorders. He recently published *Inventing Cinema: Machines, Gestures and Media History and Danièle Huillet, Jean-Marie Straub: "Objectivists" in Cinema* (Amsterdam University Press, 2019 and 2020); and *Medium, Format, Configuration: The Displacements of Film* (Meson Press, 2019).*

# L'immersion sonore : réalités physique et virtuelle, réalités psychologiques et culturelles

| Corsin Vogel et Danièle Dubois

## Résumé

L'essor récent de technologies sophistiquées de diffusion multicanal du son, ainsi que de ce qui est communément appelé dans le domaine audiovisuel « la réalité virtuelle », pose des questions à la fois techniques, scientifiques, voire philosophiques, en particulier celle du rapport entre la réalité du son décrite par la physique et l'expérience subjective du son. Plus précisément, nous allons aborder ici la relation entre la description physique de la diffusion du son dans l'espace et le sentiment d'immersion dans le son. Du point de vue de l'ingénierie des systèmes audio, il s'agit de questionner le positionnement du curseur pour l'auditeur, entre une restitution acoustique la plus proche de la description physique, permise par les technologies contemporaines, et l'illusion référentielle suscitée par les stimulations sonores d'un objet audiovisuel fictionnel. Cette interrogation, en apparence technique, nous semble pouvoir bénéficier de connaissances et développements récents dans le domaine des sciences cognitives qui permettent l'analyse et la compréhension des différentes réalités en jeu : réalités physiques, « réelle » et « virtuelle », et réalités psychologiques, tant comme ressenti que comme connaissances culturellement et historiquement situées. À partir de plusieurs exemples et d'études menées autour de ces questionnements, nous allons discuter ces différentes réalités du son dans leurs approches à la fois physicaliste et cognitive. Nous serons ainsi amenés à convoquer les notions de réalité virtuelle et d'écoutes, et, dans le cadre contemporain d'une cognition située, d'introduire les concepts de validité écologique et d'illusions référentielles.

## Abstract

*The recent boom in advanced technologies for multichannel sound, as well as of what we usually call « virtual reality » in the audiovisual field, encompass technical, scientific and even philosophical questions, and particularly one pertaining to the relationship between sound reality, as described by physics, and subjective experience of sound. More precisely, we will deal here with the relationship between physical description of sound propagation in space and the feeling of sound immersion. From a sound system engineering standpoint, we have to question where to place the cursor for a listener, between an acoustical description of sound restitution as close as possible to physical concepts, allowed by the latest technologies, and the referential illusion aroused by sound stimulations of a fictional audiovisual object. What appears at first glance to be a technical question may take benefits from recent knowledge and developments in cognitive science, which allow analysis and comprehension of the different realities coming into play : physical realities, « real » and « virtual », and psychological realities, as a feeling as well as a knowledge that is culturally and historically situated. Leaning on examples and studies led around these questions, we will discuss these different realities of sound in both physical and cognitive approaches. We also will convoke notions of virtual reality and listenings, and, within contemporary knowledge of situated cognition, introduce concepts of ecological validity and referential illusions.*



Le concept d'immersion sonore au cinéma et en diffusion audio sonorisée est associé à une évolution technologique complexe et à une démultiplication impressionnante du nombre de sources sonores, allant de quelques dizaines de haut-parleurs avec le système Dolby Atmos ou l'ambisonie, à plusieurs centaines avec la synthèse du champ sonore WFS (pour *Wave Field Synthesis*) ! Ce développement des technologies audio questionne les conceptions mêmes des « réalités sonores » : « réalités virtuelles »<sup>1</sup> en tant que re-productions acoustiques du monde physique réel, ou productions originales historiquement situées en tant que réalités culturelles et historiques, dont l'étude peut dès lors faire appel à la contribution des connaissances acquises en sciences humaines. En particulier, ces différentes réalités peuvent être, les unes comme les autres, objets d'investigations psychologiques de leur perception et ressenti.

Il s'agit tout d'abord de noter et clarifier en quoi la réalité virtuelle, produite par les technologies contemporaines, diffère de la réalité, dans la mesure où elles sont toutes deux des réalités physiques, réelles et matérielles, décrites par les sciences physiques. En outre, la notion d'immersion sonore, dès lors qu'elle est considérée à la fois comme réalité acoustique et comme expérience subjective, impose de préciser les différences entre les réalités physiques et psychologiques. En inversant la question, le sentiment d'immersion est-il réductible à la description strictement physique de ces réalités ? Nous faisons l'hypothèse que la connaissance de la perception par l'humain de ces différentes réalités physiques puisse contribuer au développement de nouvelles propositions de réalités augmentées et esthétisées dans le domaine artistique.

## Le bruit de(s) porte(s) : une question d'écoute

Appréhender l'immersion sonore comme expérience subjective suppose de se positionner du point de vue de l'être humain qui perçoit et interprète les différentes stimulations acoustiques et donc, de prendre en compte la connaissance des processus de perception, de mémorisation de ces réalités physiques, qui constituent la réalité de l'expérience vécue, comme réalité psychologique, et d'identifier en quoi cette réalité diffère de la réalité acoustique des stimulations sonores.

Partons, par exemple, de deux observations relatives au « *bruit de portes* » : l'une décrite par Mzali (2002) dans son étude sur le confort acoustique dans les TGV, l'autre empruntée à la recherche en acoustique musicale relative aux musiques contemporaines et, plus particulièrement, à une composition de Pierre Henry (Henry, 1963).

La première étude (Mzali, 2002) propose une approche du confort acoustique du train qui prend en compte à la fois sa description *analytique* comme entité physique – du point de vue des sciences de l'ingénieur, au même titre que les confort thermique, visuel, olfactif ou vibratoire –, et l'évaluation *globale* du confort par les voyageurs à partir de leurs perceptions multisensorielles dans leur usage de l'espace du train - à travers la qualité de l'assise des fauteuils, du calme, de leur place pour les jambes, etc. Il apparaît dans cette recherche que le confort acoustique est lié pour les voyageurs au ressenti d'immersion dans une ambiance sonore à laquelle contribuent des éléments non directement liés aux seules stimulations sonores<sup>2</sup>. En effet, une analyse linguistique précise d'entretiens menés à bord des trains TGV a permis d'observer des

1

« Réalité virtuelle », emprunté à l'anglais « virtual reality », introduit en français une ambiguïté, liée à la différence sémantique des adjectifs « virtuel » et « virtual ». On peut même considérer qu'il s'agit d'un oxymore : *virtuel* qualifie une réalité comme potentialité non réalisée, et donc contradictoire à une réalité physique existante. Dans l'acception maintenant attestée de cette expression dans le champ de l'Intelligence Artificielle, *virtuel* est alors synonyme d'*artificiel*, en opposition à la réalité « naturelle », comme artefact produit par des calculs informatiques de traitement de l'information, et repose sur les présupposés ontologiques d'un « réalisme ingénu » (voir Mantovani et Riva, 1999) concernant la réalité.

2

Sur la notion d'ambiance sonore, voir Augoyard (2007), Augoyard et Torgue (1995) et les travaux du CRESSON, ainsi que Dubois (2012).

réponses différentes à deux questions qui semblaient, à première vue, assez semblables, mais qui se sont révélées mettre en évidence deux types de perception du « bruit des portes » dans l'évaluation globale du confort acoustique des voyageurs :

**Q1 : Qu'est-ce qui pour vous est désagréable ?**

**Q2 : Qu'est-ce qui pour vous est gênant ?**

La première question Q1, sur le caractère « *désagréable* » ou non du confort à bord du train, suscitait, en ce qui concerne la modalité acoustique, des réponses brèves et factuelles rapportant un *phénomène acoustique objectif*, identifié comme celui émanant d'une source sonore : par exemple « le bruit des portes ». La question Q2, demandant ce qui est *gênant*, mentionnait également « le bruit des portes », mais les réponses étaient plus complexes, incluant la mention d'éléments circonstanciels comme « *le bruit des portes au passage d'un voyageur* », témoignant alors d'un événement suscité par une stimulation acoustique et produisant un *effet subjectif* particulièrement négatif. Le « *bruit des portes* » en tant que tel, en tant que *son* (phénomène acoustique), est considéré simplement comme « *désagréable* », mais devient « *gênant* » dès lors qu'il est perçu et interprété comme l'indice sonore d'un événement pouvant affecter le voyageur (comme le passage d'un autre voyageur dans le couloir, avec ses bagages, ou ramenant une tasse de café du bar)<sup>3</sup>.

On peut contraster cette observation avec celle qui concerne également des bruits de portes, mais cette fois « entendus » (au sens schaefferien de *perçus* et *conçus*) comme éléments musicaux dans les *Variations pour une porte et un soupir* de Pierre Henry (1963). Dans ce contexte de création musicale, l'enregistrement du bruit de porte est travaillé par le compositeur comme matière sonore, pour produire là aussi des effets sonores, mais

qui ont cette fois pour visée de susciter une « écoute réduite », du son pour lui-même (Schaeffer, 1966). Cette « écoute réduite » s'oppose à l'« écoute ordinaire » (Chion, 1983), celle observée, par exemple, dans les TGV. Alors que la première est centrée sur la qualification des propriétés du son lui-même, comme phénomène acoustique « en soi » indépendamment de ses causes, l'écoute « ordinaire » s'attache à l'identification des événements, des sources ou des actions qui ont produit le son (Gaver, 1993 ; Castellengo, 2015 ; Guastavino, 2021 ; Paté & Gaillard, 2021). Gaver insiste sur le fait que tout phénomène acoustique peut susciter l'un ou l'autre de ces types d'écoutes et que :

**la distinction entre écoutes ordinaire et musicale se fait entre expériences, et non entre sons<sup>4</sup>.**

Dès lors que les sons sont perçus comme sons « en soi », abstractions génériques de « bruit de porte » ou formes acoustiques, ils sont susceptibles d'être générateurs d'émotions esthétiques, tel le peintre manipulant la matière colorée des pigments pour produire des effets colorés, ou le compositeur de parfums mélangeant des odorants pour créer des parfums comme formes olfactives originales sans référentiel dans le monde « réel », ordinaire, des couleurs ou des odeurs (Cance, 2021 ; Dubois, 2021). Ces sons, couleurs ou parfums diffèrent des sons, couleurs ou odeurs ordinaires « naturels », en ce qu'ils sont des *artefacts*, des réalités artificielles produites par des artistes, compositeurs dont les savoir-faire permettent de sélectionner certaines des propriétés des matériaux sonores, picturaux, olfactifs qui sont susceptibles de produire des effets à la fois semblables et différents, de la perception des objets « réels » ordinaires ainsi re-présentés. Cette re-présentation est cependant non exhaustive et vise à orienter les processus cognitifs, soit vers une écoute ordinaire « réaliste » évocatrice de ►

**3**

Le participe présent du verbe *gêner* traduit le caractère effectif de la gêne, alors que le suffixe *-able* de l'adjectif *désagréable* indique une propriété du son comme potentialité de l'effet produit par la stimulation sonore. Cette interprétation s'est trouvée corroborée dans d'autres corpus contrastant l'usage de participe présent ou passé versus une suffixation en *-able* (siège incliné/inclinable, plaisant/agréable, etc.)

**4**

« The distinction between everyday and musical listening is between experiences, not sounds » (Gaver, 1993, p. 1, notre traduction)

réalités précédemment éprouvées, soit au contraire vers une écoute réduite, en suscitant un décalage et ainsi, réflexion et émotion.

Ces deux exemples invitent à prendre en compte une diversité de réalités, tant celle des réalités physiques matérielles : « naturelle », mais aussi artéfactuelles, à savoir artistiques et/ou « virtuelles », que celle des réalités psychologiques, vécues, résultant des différents processus induits par ces différents types de stimulations sonores. En quoi ces réalités différent-elles ? Tout d'abord dans leur visée : si du point de vue perceptif, il s'agit soit d'identifier un événement pour ajuster un comportement adapté, soit d'éprouver une émotion artistique, que visent les réalités qualifiées de « virtuelles », produites par les technologies contemporaines ? Visent-elles à produire une description, une re-production exhaustive et « objective » susceptible de se substituer à une réalité du monde « réel » et, le cas échéant, laquelle ?<sup>5</sup> Ou bien ont-elles pour objectif de produire, à partir de ces mêmes outils technologiques, des effets sonores suscitant des expériences sensibles, semblables à celles antérieurement vécues par les auditeurs, qu'il s'agisse de leur écoute ordinaire du monde ou d'une écoute réduite suscitée par d'autres modes de productions sonores, comme la musique ? Laquelle de ces réalités est visée par les réalisations des ingénieurs du son, tant dans la réalisation de bruitages et de narrations fictionnelles cinématographiques ou radiophoniques, que dans la restitution de performances musicales ?

Une première ébauche de réponse peut être trouvée par un détour dans les recherches psychologiques et anthropologiques plus largement développées dans le domaine visuel, comme celles de Hutchins (2008). Hutchins a ainsi montré que les navigateurs des îles

Trobriand utilisent les étoiles comme « boussoles » pour orienter les pirogues dans leur navigation entre les îles. S'ils peuvent voir les étoiles individuelles comme telles, c'est cependant lorsque chacune d'elles est inscrite dans une constellation qu'elle fait sens comme indicateur de la direction à suivre. Et de développer l'idée, que nous avons évoquée précédemment, de deux processus psychologiques « voir » (ou écouter) et « voir comme » (ou écouter comme) : « voir une étoile comme » dans une collection d'étoiles, dans une constellation, est un acte d'imagination et non une simple perception comme « voir » une étoile<sup>6</sup>. Ces observations sous-tendent la notion de « système de cognition distribuée » ou « unité d'écologie cognitive » (Hutchins, 2010 ; Theureau, 2020) et permettent de relier différents types de connaissances, celles issues de l'expérience sensible individuelle et celles relevant des interprétations et représentations collectives, mythes et croyances tout comme connaissances scientifiques. Transposée dans la modalité auditive, cette conception d'une « cognition située » conduit à différencier « écouter » une source sonore localisée dans l'espace comme un acte perceptif, et « écouter comme » cette même stimulation, par exemple dans un ensemble orchestral, en s'immergeant dans une constellation d'objets sonores qui, par notre imagination, forment un tout dans une écoute musicale globale.

La question du réalisme se trouve dès lors déplacée d'une référence unique à la physique vers diverses représentations et interprétations de ce « réel », que différents individus et communautés d'individus y appliquent en fonction des situations d'écoute, de leurs attentes et de leurs connaissances préalables. La prise en compte de ces différentes réalités, qu'elles soient physiques, sensorielles ou cognitives, mais toutes aussi réelles, peut conduire à différents modes

##### 5

On pensera par exemple au domaine médical où la réalité virtuelle permet aujourd'hui des interventions chirurgicales à distance, mais aussi plus généralement à des problématiques de déplacements au sein de différents espaces où les notions d'immersion et de narration sont d'une grande importance (Gorini *et al.*, 2011).

##### 6

« Individual stars are **seen**, but groups of stars **are seen** as a related collection, a constellation. Seeing a constellation is an act of imagination, not a simple perception. A constellation may be **seen as** having a shape, a name and even a persona. This phenomenon of "seeing as" is both very old and absolutely fundamental to cognition » (Hutchins, 2008, p. 2013)

d'intervention sur le « réel ». En effet, un « bruit de porte », jugé « désagréable », suscitera une amélioration acoustique du son ; mais, évalué comme « gênant » à cause du passage d'un voyageur, ce même bruit entraînera une réponse spatiale relative à l'emplacement des sièges près des portes. Le « bruit de porte » comme son musical, comme artéfact médiatisé par des dispositifs instrumentaux, suppose, quant à lui, une éducation de l'auditeur et une certaine définition de la musique qui varie selon les époques et qui se trouve diversement acceptée ou récusée par différents publics.

Dès que l'on tient compte de la présence d'auditeurs, la question du réalisme peut se trouver reformulée : dans quelle mesure une seule description physique du signal acoustique peut-elle rendre compte de ces différentes réalités à la fois physiques et cognitives ? Ou encore, de quelles réalités, physiques et psychologiques, la *réalité virtuelle* que proposent les dispositifs techniques immersifs contemporains doit-elle, peut-elle et entend-elle rendre compte ? Comment identifier les caractéristiques de ces différentes réalités cognitives qui, certes toutes aussi réelles les unes que les autres, diffèrent néanmoins quant aux processus de leur objectivation : exploration expérimentale des réalités physiques, matérielles, ou investigation psychologique de la subjectivité. En outre, après avoir identifié le vécu des auditeurs, qu'ils soient immergés dans la réalité ordinaire ou dans une réalité virtuelle, comment implémenter cette réalité cognitive dans des dispositifs techniques tels qu'ils sont disponibles ou concevables actuellement ?

## L'immersion sonore : technologies et cognition située

La question du réalisme se pose dès la captation des phénomènes sonores à l'aide de microphones, où se dessine une influence déterminante sur la fidélité du signal acoustique « prélevé » dans un espace donné. Tout aussi déterminante est la manière dont ce signal physique sera restitué dans d'autres lieux, comme une salle de cinéma, un théâtre, un salon ou encore une cuisine, un parc ou une rue passante, à travers le choix du dispositif de diffusion (Figure 1 et Figure 2). Ce signal, diffusé sur un nombre plus ou moins conséquent de haut-parleurs, sera néanmoins privé des nombreuses informations non-acoustiques du contexte de l'espace où il a été prélevé. Il sera dès lors toujours susceptible d'être interprété différemment en fonction de son niveau de réalisme physique et, du fait de sa décontextualisation, il pourra également susciter des traitements cognitifs différents de celui de la situation « ordinaire » (Vogel, 2013). Les choix technologiques actuels permettent cependant aux ingénieurs et techniciens de proposer une ou plusieurs perspectives d'écoute prédéterminées par la prise et la restitution du son et, corrélativement, de redéfinir le concept même de réalisme. Ainsi, l'approche « audio orienté objet »<sup>7</sup> redéfinit-elle non seulement la spatialisation du son sur des systèmes immersifs en multicanal, mais également la façon de capter les sons, en abandonnant généralement la stéréophonie au profit de la multi-monophonie ou d'antennes microphoniques (cf. Messonnier *et al.*, ce numéro : chapitre 6).

Qu'il s'agisse de la captation du son ou de sa restitution, les choix techniques se doivent dès lors de prendre en compte la nature du son (musique, type de musique, voix, bruitage), liée au lieu d'émission et aux lieux ►

### 7

Le format audio orienté objet décrit un espace sonore en termes d'événements audios joués depuis des positions spécifiques d'une salle (Denizot, 2016). « [Il] diffère de la conception classique des salles de cinéma où le son est spatialisé par rapport au spectateur (système égocentrique), et propose un système allocentrique basé sur la spatialisation du son dans l'auditorium. » (Denizot, 2016, p. 31) Il contient des métadonnées qui accompagnent l'information audio brute et qui informent des données spatiales géométriques.



**Fig 1. Les échos de la Saline, parcours sonore immersif avec 40 haut-parleurs disséminés dans les arbres, Saline Royale d'Arc-et-Senans, 2016. Œuvre collective d'A. Bertini, P.-L. Cassière, B. Farey, G. Malatray et C. Vogel, informatique : G. Bertrand, commissariat : L. Viard, A. Catherine et L. Molliard. (© C. Vogel, 2016)**

d'écoute, comme c'est le cas notamment du point de vue de l'acoustique des salles. Ce sont alors les décisions des acteurs humains, tant les producteurs que les auditeurs, qui s'avèrent primordiales et qui doivent être explicitées. Le réalisme est alors lié aux performances du dispositif technique lui-même, à l'interpénétration entre son et lieu, aux intentions de l'artiste, du directeur artistique ou de l'ingénieur du son et ce, pour un public donné (cf. Turquety, ce numéro : chapitre 1). Le paradigme de la restitution sonore se trouve relativisé, non seulement en fonction des contraintes des différents espaces physiques, mais également en référence à des « humains », pour l'instant limités aux artistes ou à l'ingénieur du son, c'est-à-dire aux producteurs du son, avant que les « récepteurs », les auditeurs, ne soient eux aussi, à leur tour, pris en compte.

Un détour par l'évolution des recherches sur la réalité virtuelle, davantage explorée dans le domaine visuel,

permet là encore d'apporter des éléments de discussion. Les technologies de réalité virtuelle développées dans le domaine des Interfaces Homme/Machine reposent massivement sur des présupposés hérités de la tradition positiviste et des pratiques expérimentales de la psychophysique. Il s'agit de construire des dispositifs à même de proposer des simulations du réel qui soient les plus fidèles possibles au réel tel qu'il « existe » (Pausch *et al.*, 1997 ; Mantovani & Riva, 1998). Ce sont les caractéristiques techniques de ces dispositifs qui prévalent en faisant le pari que le degré de réalisme suit et suivra l'évolution des techniques (rapidité des processeurs, résolution des systèmes visuels de reproduction...) : plus le système est sophistiqué et complexe, calculé avec une précision toujours plus grande, moins le décalage avec le « réel » ne devrait se percevoir. Dans cette perspective objectiviste, l'utilisateur, mais aussi le concepteur, sont de fait exclus (Fuchs *et al.*, 2001 ; Grumbach, 2004 ; Milon, 2006). Cependant, selon Riva & Mantovani (1999) :

**la réalité n'est pas hors de ce monde, quelque part "extérieure" à l'esprit des gens, échappant à la négociation sociale et à la médiation culturelle ; la réalité est co-construite dans sa relation entre acteurs et leurs environnements à travers la médiation des artéfacts.<sup>8</sup>**

Il faut ainsi davantage concentrer les efforts sur l'interaction entre l'environnement tel qu'il a été créé, les personnes qui y évoluent, les objets qui en font partie et qu'il s'agit de « simuler ». Ce changement de positionnement épistémologique introduit le concept de *présence* explicitement défini comme construction cognitive - et donc culturelle -, au même titre que la description de la réalité décrite par les sciences physiques (Mantovani & Riva, 1999).

8

« Reality is not out there in the world, somewhere "outside" people's minds, escaping social negotiation and cultural mediation; reality is co-constructed in the relationship between actors and their environments through the mediation of the artifacts » (Riva & Mantovani, 1999, notre traduction).

On illustrera cette question à partir des recherches de Cance sur la pertinence des réalités virtuelles, développées à la demande de l'industrie automobile pour l'aide à la conception de simulateurs conçus pour évaluer la qualité visuelle de l'habitacle automobile (Cance, 2008 ; Cance *et al.*, 2009). Au cours de ces recherches, une attention particulière a été portée sur les couleurs des différents matériaux en proposant plusieurs dispositifs expérimentaux de présentation des habitacles : deux habitacles réels de véhicules à l'arrêt, un dispositif 2D restituant des images fixes sur un écran et un dispositif 3D d'immersion visuelle CAVE (acronyme de « *Cave Automatic Virtual Environment* »), avec suivi de mouvement des têtes des participants. Ces recherches ont mis en évidence que les dispositifs de simulation visuelle 2D et 3D induisent des rapports différents à l'espace, ainsi que des différences majeures selon la familiarité des dispositifs pour les personnes interrogées. Cette familiarité conditionne l'illusion de réel et, par là même, l'orientation des processus cognitifs : soit sur le « réel » représenté par le dispositif, soit sur la cohérence perceptive du dispositif lui-même, à savoir l'écart entre le dispositif et le « réel » perçu. Dans le dispositif 3D de réalité virtuelle, les individus peu familiers évaluent le dispositif non comme des représentations d'un monde préexistant, mais comme une expérience sensible nouvelle, originale, et ce, à la différence des ingénieurs qui ont conçu et construit le système. Pour ces derniers, le dispositif s'ajuste adéquatement, formellement au « réel ». Cependant, en réponse à la demande de l'industriel qui souhaitait valider le dispositif de réalité virtuelle comme mode de présentation des modèles mis à la disposition des acheteurs potentiels, le dispositif de réalité virtuelle rate sa cible et ne constitue pas un substitut « équivalent » au « réel » qui puisse être pertinent pour les clients. À l'inverse, la familiarité des clients avec le dispositif 2D et

le type de représentations visuelles qu'il offre permet, jusqu'à un certain point tout au moins, une illusion référentielle et une évaluation de l'ambiance elle-même et non du dispositif, comme c'est évidemment aussi le cas en concession dans des modèles vraiment « réels ». Cette familiarité confère ainsi au dispositif



**Fig 2 : Klangstall, quatre installations sonores immersives de C. Vogel dans une ancienne étable de montagne, Gadäwäg, Klosters, Suisse, 2017. Commissariat : Gasser, Derungs, Zurich. (© C. Vogel, 2017)**

2D un meilleur réalisme conçu comme une certaine « transparence référentielle »<sup>9</sup>. Cependant, il n'en demeure pas moins que si, même dans ce cas, ce n'est pas le dispositif qui est évalué, le mode d'évaluation reste contraint par le dispositif et renvoie davantage à l'évaluation de l'ambiance de l'habitacle comme image, représentation, et non comme un espace, réalité vécue par le sujet. Les mêmes problématiques se posent en réalité augmentée et sont abordées dans ce numéro (Bauer, ce numéro : chapitre 3).

Les recherches dans le domaine visuel que nous venons ►

## 9

La question de la familiarité des dispositifs renvoie ainsi à la question de l'interculturalité et au problème de l'adéquation des représentations aux différents modes de pensée, modes de structuration du monde, de l'expérience sensible et des connaissances (Misra et Gergen, 1993). Elle renvoie également à la problématique des expertises, qui peuvent être envisagées comme différents modes d'appréhension du monde, comme prises de position sur le monde qui le rendent familier (Morange *et al.*, 2005).

de rapporter montrent que le concept de réalisme se doit de prendre également en compte celui qui perçoit le monde « virtuel », comme monde construit, comme artéfact, et donc pose la question du *réalisme pour qui ?* Dans le domaine auditif, bien souvent, l'auditeur reste, soit absent, soit très abstrait, « universel », ou encore rapporté à des conditions « objectives » génériques d'écoute, prenant en compte des variations de sa position dans l'espace physique, comme le « *sweet spot* » ou la « zone d'écoute », dans différents types de lieux ou dans le choix de l'artiste ou du concepteur sonore. Les nouvelles technologies numériques aggravent en quelque sorte ce constat dans la mesure où elles permettent d'une part, la conception de dispositifs sophistiqués de plus en plus « réalistes », en créant des réalités virtuelles multimodales à partir du traitement unifié de l'information audiovisuelle ; et d'autre part, sur le plan conceptuel, ces technologies conduisent à développer, en psychologie, des modèles cognitifs reprenant la longue tradition de fascination de nos cultures pour leurs productions techniques. Ainsi, si Descartes, à l'ère du machinisme naissant, considérait les humains comme des automates, les technologies numériques de l'Intelligence Artificielle et des « réalités virtuelles » assimilent le fonctionnement cognitif humain à un système de traitement de l'information. La conceptualisation des processus perceptifs demeure influencée par les modèles et connaissances implantés dans les technologies contemporaines de chaque époque, même si, au sein des sciences cognitives, la connaissance du fonctionnement humain peut intervenir dans l'aide à la conception des artéfacts « cognitifs ». Les conceptions des ingénieurs concepteurs, comme d'ailleurs celle des psychologues cognitivistes<sup>10</sup> actuels, reposent de fait sur l'idée d'un auditeur abstrait, certes universel, mais non situé dans un espace sociologique et culturel, dont l'expérience

sensible s'inscrit dans un espace de pratiques qui puisse rendre compte de la diversité des écoutes qui ont façonné ses expériences antérieures. Ce sont ces expériences qui rendent réalistes les réalités virtuelles. Dans certaines situations cependant, les progrès technologiques permettent des avancées spectaculaires. Il en va ainsi de la réverbération active par exemple, qui confère aux musiciens une sensation de confort considérablement améliorée par rapport à l'acoustique naturelle d'une salle de répétition (Baranger, 2021). Les musiciens peuvent alors jouer en adoptant les conditions acoustiques de jeu d'une salle de concert ou en optimisant simplement la qualité d'écoute pour jouer. Ces réverbérations actives, installées dans des salles de spectacles et optimisées grâce à l'expertise pratique des ingénieurs du son, modifient la sensation d'immersion sans même que le public n'ait conscience de la présence d'un quelconque dispositif électroacoustique (Jullien, ce numéro : chapitre 5).

### L'espace sonore : réalités et illusions référentielles

Ces questions s'avèrent encore plus saillantes dès lors que l'on prend en compte les relations entre le son et l'espace, et le rôle décisif du son dans la perception de l'espace, qu'il s'agisse des recherches sur les ambiances sonores urbaines, les « *soundscapes* », ou des recherches sur le son musical, problématisées à travers le concept d'immersion. L'exemple du *deep listening*, qui consiste à s'approprier les caractéristiques acoustiques d'un lieu pour en proposer une expérience artistique, permet de constater que l'espace est bien un paramètre musical polysémique dont les implications perceptives sont riches et diversifiées (Féron, ce

<sup>10</sup>

*Cognitif vs cognitiviste* : l'adjectif *cognitif* qualifie ce qui relève de la connaissance, alors que *cognitiviste* renvoie à un des cadres théoriques en psychologie relatif à la prise en compte des phénomènes cognitifs influencé par le modèle computationnel, en termes de traitement de l'information.

numéro : chapitre 4). Par ailleurs, les recherches sur les environnements sonores ou sur les ambiances urbaines (Guastavino, 2021 ; Dubois et al., 2006) ont montré que les bruits urbains pouvaient également susciter, en situation « réelle », différents types de traitements cognitifs et ainsi, produire différentes réalités psychologiques : soit comme perception d'un indice sonore d'un événement du monde, comme une voiture qui passe, des enfants qui crient, des oiseaux qui chantent..., soit comme perception du son pour lui-même, comme le bruit de fond urbain, une intensité sonore trop forte sans relation imputable à une source particulière. Ainsi, lors du transfert des études de terrain sur les ambiances urbaines en laboratoire, afin d'identifier plus précisément les propriétés physiques du son responsables de la qualité des ambiances, il s'est avéré que ces deux approches perceptives étaient très sensibles au type de dispositif d'enregistrement et de restitution du son : les enregistrements et la restitution stéréophoniques et binauraux s'avéraient équivalents et suffisamment réalistes pour l'identification perceptive en laboratoire de sources sonores, alors que la sensation d'immersion dans l'espace sonore urbain requérait un enregistrement et une restitution spécifique, de type ambisonique (Guastavino, 2009).

C'est donc en fonction d'une conception *située* de la diversité des processus perceptifs humains que peut être établi le réalisme comme validité écologique des dispositifs techniques introduits précédemment. Leur finalité n'est alors pas de re-produire le plus exactement possible la réalité physique, mais bien de susciter chez celles et ceux qui les utilisent des processus psychologiques et des comportements similaires à ceux qui se produisent dans le monde « réel », ordinaire, comme c'est empiriquement le cas par exemple des appeaux qu'utilisent les chasseurs et qui constituent

des leurres adéquats pour attirer des congénères (Despret, 2012). Pour reformuler ces observations dans le domaine sonore, les « simples » dispositifs d'enregistrement et de restitution stéréophoniques satisfont cette illusion référentielle d'un espace sonore pour la reproduction en laboratoire des bruits de sources sonores (voitures, signaux d'avertissements, etc.) ou de perception frontale de l'espace. C'est ce qui a été observé lors de la comparaison des réactions des auditeurs à différentes situations d'écoute : in situ dans des situations ordinaires ou en laboratoire sur un système de diffusion, afin de vérifier les stratégies d'écoute adoptées. Les réponses identiques dans les deux situations ont permis de considérer que les stratégies d'écoute sont similaires et que l'écoute en laboratoire, dans laquelle le son a été abstrait du « monde réel », reste écologiquement valide, à condition toutefois d'être re-situé dans un cadre orienté par les consignes d'écoute (Vogel *et al.*, 1997 ; Vogel, 1999 ; 2000)<sup>11</sup>. En revanche, l'étude du sentiment d'immersion du public dans une ambiance donnée nécessite des dispositifs techniques immersifs de type ambisonique (Guastavino, 2007), *wave field synthesis* (Corteel, 2004) ou Dolby Atmos au cinéma pour des écoutes sur haut-parleurs, ou encore des systèmes binauraux pour des écoutes au casque, notamment les systèmes de réalités virtuelle et augmentée (cf. Bauer, ce numéro : chapitre 3). Avec ces récentes propositions immersives, ce sont les nouvelles approches orientées objet (cf. Jullien, ce numéro : chapitre 5), ainsi que la prise en compte des fonctions de transfert de la tête de l'auditeur<sup>12</sup> qui ouvrent de nouvelles voies de recherches.

En d'autres termes, il s'agit, pour l'étude en laboratoire des environnements sonores, non pas de viser de manière générique une reproduction réaliste la plus exacte et la plus exhaustive des signaux sonores captés ►

<sup>11</sup>

Sur les consignes et leur rôle dans la validité écologique des situations expérimentales voir Dubois et al. (2021)

<sup>12</sup>

Mesures des HRTF, pour *Head Related Transfer Functions*, qui diffèrent d'un individu à l'autre.

dans l'environnement par les dispositifs techniques, mais de produire une stimulation qui suscite une illusion, une *fiction référentielle* suffisante pour que l'auditeur réagisse « comme si » il était en situations réelles lorsqu'il déambule dans la ville. Le réalisme de cette illusion se doit alors de tenir compte des réalités psychologiques de l'auditeur et donc, non seulement de la réalité physique des stimulations qu'il éprouve, mais aussi de celles qu'il a éprouvées au préalable, qu'il a mémorisées et qui lui permettent d'interpréter le signal. C'est ce que nous désignons, dans le cadre de la recherche en perception, comme la *validité écologique* des dispositifs expérimentaux (Gibson, 1979 ; Dubois *et al.*, 2021). C'est d'ailleurs ce qui se trouve empiriquement réalisé par les savoirs faire des bruiteurs pour la sonorisation des films : ils réalisent des prototypes, quasi-caricatures de sons dont le succès est de donner l'illusion d'un réel ressenti, en tenant compte et en intégrant dans l'objet sonore produit des éléments déclencheurs d'émotions, comme peut l'illustrer le son d'une pastèque découpée pour susciter l'illusion et l'émotion d'un couteau transperçant un corps (*Psychose* d'Alfred Hitchcock, 1960).

Dans le domaine musical, le même type de remarques ont pu être également formulées. Turquety (ce numéro : chapitre 1) rapporte que Evan Parker a régulièrement souligné l'impossibilité de distinguer le son de l'instrument de l'espace dans lequel il est joué, comme, dans notre cas, le son du dispositif de sa restitution en laboratoire. Le réalisme se trouve ainsi relativisé à la position dans l'espace et au type de sources ; les questions de l'équilibre des timbres des différents instruments d'un orchestre par exemple, en fonction de leur rayonnement et leur directivité, restent posées, car les technologies actuelles de réalité virtuelle ne sont pas encore capables d'en rendre

compte. De même, si un grand nombre de paramètres physiques en acoustique des salles sont bien identifiés et donnent lieu à des modèles hautement précis et performants, qu'en est-il de la description acoustique des différents timbres qui conditionnent l'identification des sources sonores, et sur la base de laquelle s'effectue le jugement perceptif des qualités acoustiques à la fois de l'instrument et de la salle de concert ? Sans compter les critères de la qualité de la restitution d'un concert, en situation réelle, qui tiennent également compte des « réponses » de la salle selon son remplissage, et également des perceptions différentes des auditeurs selon leurs places, autant de paramètres liés au « contexte » de la production sonore en situation d'écoute en concert.

Cependant, à la différence de la recherche sur les environnements sonores urbains dans lesquels le bruit est subi, dans le domaine musical, la non-prise en compte de l'auditeur lui-même est encore plus problématique, car la musique est produite pour être écoutée. Ainsi, même dans le cas d'enregistrements musicaux, pour lesquels la position d'écoute est « figée », il s'agit de prendre en compte non seulement le type de son reproduit, le style musical et les interprètes, mais également les destinataires : l'amateur, le spécialiste d'un auteur/compositeur, d'un orchestre ou de son chef, sont, du fait de leurs différentes cultures musicales, diversement sensibles aux multiples interprétations et restitutions possibles d'une « même » œuvre.

## Réalités et productions culturelles

L'étude du « réalisme » des dispositifs audio ne peut ainsi se contenter de l'enregistrement et de la restitution d'un signal acoustique qui existerait « en soi », par

des dispositifs fussent-ils des plus sophistiqués, sans tenir compte ni de la diversité des objets sonores, non réductibles aux seules vibrations, ni des processus psychologiques qui permettent de faire sens pour les différents auditeurs potentiels et visés par la réalisation technique. C'est ce qu'illustre un dernier exemple relatif à la restauration d'un enregistrement ancien à travers l'évaluation de la qualité sonore de diverses versions d'un enregistrement monophonique de 1907 d'un opéra de Leoncavallo chanté par Caruso (Morange *et al.*, 2005 ; 2007).

Tout comme pour la restauration d'édifices architecturaux - on pensera par exemple aux vifs débats pour la reconstruction de la cathédrale Notre-Dame à Paris -, la restauration d'anciens documents sonores en vue de leur diffusion auprès du public soulève de nombreuses questions dont celles du réalisme de la restitution : doit-on laisser l'enregistrement sonore dans son état originel en tant que marque d'*authenticité*, avec tous les problèmes d'intelligibilité que cela peut présenter (craquements, bruit de fond...) ? Cette question de l'authenticité est d'ailleurs intimement liée au support physique : le disque vinyle ou le cylindre de cire. Que signifie l'authenticité d'un son seul dont le contenu présente des craquements supposant la présence du support ? Doit-on au contraire « nettoyer » certains bruits parasites ou essayer de retrouver, grâce aux technologies contemporaines plus sophistiquées, l'état originel ? Mais de quoi : du son fidèle tel qu'il a été émis, mais disparu à tout jamais, ou de l'enregistrement premier, à travers un pavillon et une aiguille traçant son sillon sonore dans la cire ? Ou plutôt de manière illusoire et mythique de la production du son telle qu'elle a été produite à l'époque dans une chambre d'hôtel ? Ou encore plus prosaïquement, s'agit-il de mettre le son au « goût du jour » en remasterisant

l'enregistrement ancien à des fins commerciales, au risque de dénaturer largement la version originale, par l'ajout de réverbération et d'effets stéréophoniques pour donner l'illusion d'un concert contemporain à la Scala de Milan ?

Dans un projet collectif réunissant des spécialistes en archives sonores, en techniques du son, en linguistique et en psychologie cognitive, une étude originale sur l'évaluation cognitive de différentes versions remasterisées d'un enregistrement musical ancien a été réalisée (Morange *et al.*, 2005). Il s'agissait de onze enregistrements numérisés, dont dix commercialisés sous forme de CD audio, édités entre 1989 et 2004, de l'opéra *Pagliacci*, de R. Leoncavallo (1890), tous développés à partir du même enregistrement sur cylindre de cire considéré comme original de E. Caruso en 1907. Une onzième version a été remastérisée par nous-mêmes, dans un souci de recherche d'un maximum d'authenticité du lieu et des conditions de captation. Le questionnement a sollicité des personnes d'âges différents (- de 30 ans et + de 60 ans, en 2004), celles-ci étant différemment familières aux supports analogiques et numériques et, par conséquent, à des types de qualité de rendu sonore spécifiques à différentes époques, allant de l'ère du disque vinyle et du phonographe à l'ère du CD audio, introduit en 1983. En outre, certaines personnes se trouvèrent différemment concernées par l'évaluation de ce matériau sonore : les « experts », tels les ingénieurs du son, acousticiens, amateurs de chant lyrique, musiciens ; les « non-experts », personnes pas ou peu intéressées par la musique et le son. De cette façon, les critères de qualité sonore - corrélée à l'expertise des sujets - et d'objet culturel historiquement situé - corrélé à l'âge des sujets - liés au document sonore conduisent à évaluer l'ensemble des critères ►

de catégorisation « subjective » de ces différentes populations et leurs couplages avec les descriptions de référence en sciences physiques. Ces recherches ont montré que l'écoute monophonique d'un opéra de Caruso suffisait à permettre l'abstraction nécessaire pour créer l'illusion référentielle de l'événement musical de l'époque, et les réponses de l'ensemble des auditeurs ont mis en évidence que les séquences les plus retravaillées, nettoyées, filtrées et réverbérées sont les plus appréciées, au détriment des séquences laissées proches du document sonore originel bruité, contenant les artéfacts liés au support analogique et pouvant susciter une certaine nostalgie liée au son d'époque. Les auditeurs experts de tous âges que nous avons interrogés ont préféré les sons nettoyés de leurs défauts techniques, équilibrés dans le spectre, ainsi que ceux ayant un apport de réverbération, ce qui modifie de fait le caractère authentique de l'archive sonore. Par ailleurs, il est intéressant de constater que la onzième version, remasterisée par nos soins, n'a suscité ni adhésion ni rejet de la part des participants. Il y a certainement eu dans cette approche « neutre » une attitude trop peu engagée pour aller dans le sens de choix forts, voire clivants, entre auditeurs cherchant le son d'époque d'une part, et participants sensibles à la transformation de l'archive sonore et à la création d'artéfacts anachroniques ou non situés d'autre part. Là aussi, nous retrouvons l'importance d'un choix, d'une direction artistique, qui oriente l'écoute des participants (Castellengo, 2010).

En tout état de cause, sur le plan du réalisme qui nous concerne ici, ces recherches ont mis en évidence le fait que les enregistrements sonores sont des objets complexes qui suscitent une diversité d'expériences sensibles, et dont l'évaluation dépend des expertises et motivations d'une grande diversité d'acteurs et

d'auditeurs. Ainsi, les enregistrements peuvent-ils être considérés comme :

- des *objets techniques*, analysés selon des critères de « réalisme » qui évoluent en fonction des technologies du son et de leur maîtrise ou intérêt par les auditeurs ;
- des *objets sonores* analysés selon des paramètres physiques, dont l'évaluation dépend des connaissances scientifiques en acoustique et dont la pertinence varie en fonction des types de sources sonores (voix, musique, parole, bruits de sources, environnements sonores...) ;
- et surtout, des *objets culturels* évalués par une diversité d'auditeurs en fonction de leur inscription dans une époque, contemporaine ou datée, leur expertise et intérêt vis-à-vis du sonore ou vis-à-vis du style, du genre de musique, de l'exécution de l'œuvre et de son contexte, dont le réalisme, comme fiction référentielle vécue, doit être pris en compte.

-

Les différents exemples que nous avons abordés relatifs aux jugements de scientifiques, ingénieurs, techniciens, ou d'artistes, musiciens amateurs, mélomanes, ou tout simplement d'usagers de notre monde, conduisent régulièrement à questionner, non plus le « réel » qu'il s'agirait de re-présenter, mais les différentes réalisations *des* réalités, virtuelles ou « réelles », tout aussi physiques et formalisables les unes que les autres. Celles-ci dépendent certes des connaissances physiques, tout comme des performances possibles des systèmes techniques. Cependant, dès lors que ces réalisations sont considérées comme des objets culturels à destination de divers publics, elles se trouvent - comme d'ailleurs *la* « réalité vraie » - également « situées » dans un

espace historique et sociologique, en fonction de points de vue qui peuvent être multiples et divers. Ainsi, d'un point de vue physicaliste, il ne s'agit plus de considérer par exemple l'immersion sonore ou un événement musical comme des objets spatio-temporels réductibles à une liste de propriétés acoustiques associées à des dispositifs technologiques, et indépendantes de la nature même de ces objets. Car, d'un point de vue cognitif, ces objets techniques font sens sous de multiples aspects, à commencer par la familiarité diverse des usagers à l'égard de ces objets. Aussi, en modifiant leurs propriétés acoustiques, il est possible de décider (ou non) de préserver le caractère historique de l'objet sonore, ou sa capacité d'immersion perceptive. Il en va de même dans le cas d'une production artistique, dont il faut alors rendre compte du style, de l'interprétation, de l'appartenance à un genre, une culture, un ensemble de valeurs symboliques, etc.

À l'aune des différents exemples décrits et analysés précédemment, nous pouvons ainsi établir cinq conceptions de l'immersion qui, malgré son caractère éminemment plurisensoriel, peut être étudiée en la restreignant au domaine sonore, mais en tenant compte des capacités cognitives humaines à traiter diversement les stimulations acoustiques selon les finalités de l'écoute : *l'immersion physique* « naturelle », dans le monde réel, espace mesurable et milieu de perception et d'action de l'être humain ; *l'immersion virtuelle*, dans une reconstruction physicaliste et cognitiviste du monde réel ; *l'immersion artéfactuelle*, réalité artificielle issue d'une proposition artistique de notre environnement ; *l'immersion écologique*, illusion référentielle et abstraction cognitive permettant de transposer le monde réel au laboratoire, à la salle de cinéma ou à toute représentation non diégétique et/ou anachronique, fictionnelle du monde ; et enfin,

*l'immersion culturelle*, nécessairement cognitive et qui passe par le prisme du vécu et de l'expertise. L'explicitation de ces cinq types d'immersion constitue une grille d'analyse qui bien évidemment n'exclut pas leurs interdépendances. Par ailleurs, le concept d'immersion, tel que nous l'envisageons dans cette approche pluridisciplinaire, tient compte tout autant de la réalité physique des stimulations acoustiques et de l'environnement de propagation, que de la diversité des traitements cognitifs associés à leur perception.

Pour reprendre Descartes dans ses *Méditations métaphysiques* (3.19) :

**Quant aux autres choses, comme la lumière, les couleurs, les sons, les odeurs, les saveurs, la chaleur, le froid, et les autres qualités qui tombent sous l'attouchement, elles se rencontrent dans ma pensée avec tant d'obscurité et de confusion, que j'ignore même si elles sont véritables, ou fausses et seulement apparentes, c'est-à-dire si les idées que je conçois de ces qualités, sont en effet les idées de quelques choses réelles, ou bien si elles ne me représentent que des êtres chimériques, qui ne peuvent exister.**

L'artificiel et le virtuel s'inscrivent donc, tout comme la « vraie » réalité, en tant que réalités matérielles, perçues et interprétées, dans un contexte historique et socio-économique où la création artistique est elle-même liée au développement technologique et donc à l'industrie de l'audiovisuel. Prendre en compte les propriétés des objets techniques dans leur production comme réalités artéfactuelles, situées dans leur contexte historique et culturel de création, c'est ainsi expliciter les finalités de la création pour des publics eux aussi pluriels. ►

## Bibliographie

- > Augoyard, J.-F. (2007), A comme Ambiance(s). In *Les Cahiers de la recherche architecturale et urbaine 20/21*, L'espace anthropologique, pp. 33-37.
- > Augoyard, J.-F., Torgue, H. (1995), *Répertoire des effets sonores, à l'écoute de l'environnement*. Marseille, France: Parenthèses.
- > Baranger, F. (2021), *La réverbération active dans le confort du musicien*, Mémoire de Master, Conservatoire National Supérieur de Musique et de Danse de Paris. France.
- > Cance, C. (2008), Expériences de la couleur, ressources linguistiques et processus discursifs dans la construction d'un espace visuel : l'habitacle automobile (Doctorat. Université Paris 3).
- > Cance, C., Giboreau, A., Dubois, D. (2009), Co-construction du sens et de l'espace en discours : l'exemple des couleurs. In Havu, E., Harma, J., Helkkula, M., Larjavaara, M., Tuomarla, U. (Eds.), *La langue en contexte*, Société Philologique, pp. 363-375.
- > Cance, C. (2021), Experiencing and talking about colors, in Dubois, D., Cance, C., Coler, M., Paté, A., Guastavino, C. (Eds) (2021), *Sensory Experiences : Exploring meaning and the senses*, Converging Evidence in Language and Communication Research, Vol. 24 ; John Benjamins Publishing Company, pp. 87-117. DOI: <https://doi.org/10.1075/celcr>
- > Casati, R., Varzi, A. (2008), Event concepts. In *Understanding events: From perception to action*, T. Shipley & J. Zacks (Eds.), New York: Oxford University Press, pp. 31-54.
- > Castellengo, M., (2015), *Écoute musicale et acoustique*, Eyrolles, Paris.
- > Castellengo, M. (2010) À l'écoute de nouveaux « sons anciens » : variations autour de la notion d'authenticité sonore, *Utopia instrumentalis : fac-similés au musée*, 71.
- > Chion, M., (1983), *Guide des Objets Sonores*, Buchet/Chastel, Paris, p. 23.
- > Corteel, E., (2004), Création et manipulation de scènes sonores pour la wave field synthesis. In: *Cahier Louis-Lumière n°2 : Espaces pluriels, images et sons*, pp. 62-83.
- > Denizot, F. (2016), L'approche orientée objet au cinéma : entre surenchère technologique et outil de création sonore, Mémoire de Master, Ecole Nationale Supérieure Louis-Lumière, Saint-Denis.
- > Despret, V. (2012), Que diraient les animaux si on leur posait les bonnes questions ? La Découverte, Paris, France.
- > Dubois, D. (2006), From psychophysics to semiophysics. Categories as Acts of Meaning. A case study from olfaction and audition, back to colors. In Plümacher, M., Holz, P. (Eds.), *Speaking of colors and odors. An interdisciplinary approach to cognitive and linguistic categorization of color vision and olfaction*, Amsterdam, the Netherlands: John Benjamins Publishing, pp. 167-184.
- > Dubois, D. (2009), *Le Sentir et le Dire – Concepts et méthodes en psychologie et linguistique cognitives*, L'Harmattan, Paris, France.
- > Dubois, D. (2012), Dénommer, définir, identifier, décrire une ambiance — A semantic analysis of the word "soundscape". In *Ambiances in action, Proceedings of the 2nd International Congress on Ambiances / Ambiances en acte(s)*, International Ambiances Network, Montréal, Canada, pp. 683-688.

- > Dubois, D., Bourguine, R., Resche-Rigon, P. (1992), Connaissances et expertises finalisées de divers acteurs économiques dans la catégorisation d'un objet perceptif, *Intellectica*, 15, 241-271.
- > Dubois, D., Cance, C., Paté, A. (2021), Procedures and outcomes, in Dubois, D. Cance, C., Coler, M., Paté, A., Guastavino, C. (Eds), *Sensory Experiences : Exploring meaning and the senses*, Converging Evidence in Language and Communication Research, Vol. 24;, John Benjamins Publishing Company, pp. 475-503. DOI: <https://doi.org/10.1075/celcr>
- > Dubois, D., Coler, M., Wörtche, H. (2014), Knowledge Sensory Experience, and Sensor Technology, In Rangacharyula, B.J.C., Haven, E. (Ed.), *The world in prismatic views* (pp. 97–133), Singapore: World Scientific Publishing.
- > Dubois, D., Guastavino, C., Raimbault, M. (2006), *A Cognitive Approach to Urban Soundscapes: Using Verbal Data to Access Everyday Life Auditory Categories*, *Acta Acustica United with Acustica*, 92(6), pp. 865–874.
- > Dubois, D., Cance, C., Coler, M., Paté, A., Guastavino, C. (Eds), *Sensory Experiences : Exploring meaning and the senses*, Converging Evidence in Language and Communication Research, Vol. 24, John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/celcr>
- > Gaver, W. (1993), What in the world do we hear? an ecological approach to auditory event perception, *Ecological Psychology*, 5(1), pp. 1–29.
- > Gibson, J.J. (1979), *The Ecological Approach to Visual Perception*, réédition de 1986, Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- > Gorini, A., Capideville, C. S., De Leo, G., Mantovani, F., Riva, G. (2011), The role of immersion and narrative in mediated presence: the virtual hospital experience. *Cyberpsychology, Behavior, and Social Networking*, 14(3), pp. 99-105. DOI: <https://doi.org/10.1089/cyber.2010.0100>
- > Guastavino, C. (2007), Categorization of environmental sounds. *Canadian journal of experimental psychology*, 61(1), pp. 54–63.
- > Guastavino, C. (2009), Validité écologique des dispositifs expérimentaux. Pour qui ? Pour quoi ? Pour quoi faire ? In Dubois, D. (Ed.), *Le Sentir et le Dire*, L'Harmattan, Paris, France, pp. 233–252.
- > Henry, P. (1963), *Variations pour une porte et un soupir*, pièce de musique concrète en 25 mouvements (48'17").
- > Hutchins, E. (2010), Cognitive ecology. *Topics in cognitive science*, 2(4), pp. 705-715.
- > Hutchins, E. (2008), The role of cultural practices in the emergence of modern human intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1499), pp. 2011-2019.
- > Mantovani, G., Riva, G. (1999), "Real" presence: how different ontologies generate different criteria for presence, telepresence, and virtual presence. *Presence*, 8(5), pp. 540-550.
- > Misra, G., Gergen, K. J. (1993), On the place of culture in psychological science. In *International Journal of Psychology* 28 (2), pp. 225-243.
- > Mzali, M. (2002), Perception de l'ambiance sonore et évaluation du confort acoustique dans les trains, Thèse de doctorat, Sorbonne Université, Paris.
- > Morange, S., Fontaine, J.-M., Vogel, C., Poitevineau, J., Dubois, D. (2005), Appartenances catégorielles d'un objet complexe, la voix : approches linguistiques et psycho-cognitives. In Journée Annuelle du Sensolier, *Catégorisation et catégories : des théories cognitives aux pratiques en évaluation sensorielle*, pp. 14–17. ►

> Morange, S., Dubois, D., Fontaine, J.-M. (2007), *Voices of Caruso : Cognitive evaluation and acoustic analysis of reedited editions Online*. In Maimets-Volk, K., Parncutt, R., Marin, M., Ross, J., (Eds.), *Proceedings of the third Conference on Interdisciplinary Musicology (CIMo7)*, Tallinn, Estonie.

> Newton, I. (1704), *Opticks, or, a treatise of the reflections, refractions, inflections & colours of light*. Courier Corporation, Optique (1979, 1999) Bourgois, C., Paris, France.

> Paté A., Gaillard, P. (2021), Exploring and talking about music, in Dubois, D., Cance, C., Coler, M., Paté, A., Guastavino, C. (Eds), *Sensory Experiences : Exploring meaning and the senses*, Converging Evidence in Language and Communication Research, John Benjamins Publishing Company, Vol. 24, pp. 213-247, DOI : <https://doi.org/10.1075/celcr>

> Riva, G., Mantovani, G., (1999), The ergonomics of virtual reality: Human factors in developing clinical-oriented virtual environments, *Studies in Health Technology and Informatics*, vol 62, pp. 278-284. DOI: 10.3233/978-1-60750-906-6-278

> Schaeffer, P. (1966), *Traité des objets musicaux*, Seuil, Paris, France.

> Theureau, J. (2020), *Cognition distribuée et Cours d'action / Distributed cognition and "Course of action"*, URL : <https://journals.openedition.org/activites/5308>, DOI : <https://doi.org/10.4000/activites.5308>.

> Vogel, C. (2013), Un art sonore situé : le sens et l'espace, *Etnográfica*, vol 17 (3), pp. 605-616. DOI : 10.4000/etnografica.3260

> Vogel, C. (2000), Les signaux d'avertissement : catégories et contextes d'émission. In Dubois D. (Ed.), *Espaces sensoriels et formes lexicales*, Vol. 4, LCPE, Paris, France, pp. 51-76.

> Vogel, C. (1999), Étude sémiotique et acoustique de l'identification des signaux sonores d'avertissement en contexte urbain, Thèse de doctorat, Sorbonne Université, Paris.

> Vogel, C., Maffiolo V., Polack, J.D., Castellengo, M. (1997), Sound characterization of urban environments: an approach based on ecological validity, *102ème Convention de l'Audio Engineering Society*, 22-25 mars 1997, Munich.

## CORSIN VOGEL

Corsin Vogel est artiste et professeur associé, titulaire d'un Master 2 en acoustique et vibrations (Aix-Marseille Université) et d'un doctorat en acoustique musicale (Sorbonne Université). Il s'est formé et a développé ses recherches dans un contexte pluridisciplinaire associant les sciences physiques et cognitives. Sa pratique artistique, nourrie du patrimoine culturel et des territoires, se déploie sous des formes variées : installations, compositions électroacoustiques et mixtes, mais également commissariat d'expositions et scénographie. Il enseigne actuellement, à l'École Nationale Supérieure Louis-Lumière et au Conservatoire National Supérieur de Musique et de Danse de Paris, la perception auditive et la méthodologie de mémoire, en y assurant le suivi des mémoires de Master Son. Il poursuit en parallèle ses recherches en acoustique environnementale et en perception auditive autour du son immersif

*Corsin Vogel is an artist and associate professor, holding a Master 2 in acoustics and vibrations (Aix-Marseille University) and a PhD in musical acoustics (Sorbonne University). He trained and developed his research in a multidisciplinary context associating the physical and cognitive sciences. His artistic practice, nourished by cultural heritage and territories, is deployed under various forms: installations, electroacoustic and mixed compositions, but also in exhibitions and scenography. He currently teaches at the École Nationale Supérieure Louis-Lumière and at the Conservatoire National Supérieur de Musique et de Danse de Paris, the auditory perception and the methodology for exploring memory, by ensuring the follow-up of Master's theses on acoustics. In parallel, he pursues his research in environmental acoustics and in auditory perception of the immersive sound.*

## DANIÈLE DUBOIS

Danièle Dubois est directrice de recherche émérite en psychologie cognitive et linguistique au CNRS, où elle était responsable du thème Langages, Cognitions, Pratiques et Ergonomie au sein de l'équipe LAM (Lutheries, Acoustique, Musique) rattachée aujourd'hui à l'Institut Jean Le Rond d'Alembert, Sorbonne Université, Paris. Elle a notamment publié l'ouvrage *Sensory Experiences: Exploring meaning and the senses*, en collaboration avec Caroline Cance, Matt Coler, Arthur Paté et Catherine Guastavino (John Benjamins Publishing, 2021). Cet ouvrage de synthèse de recherches en psychologie et linguistique cognitives explore les relations entre langage et cognition dans les différentes modalités sensorielles, dans une conception de cognition « située », fondée sur les significations que différents acteurs accordent à leurs expériences sensorielles (voir dans ce numéro : notes de lecture).

*Danièle Dubois is Director of Research Emeritus in Cognitive Psychology and Linguistics at the CNRS, where she was in charge of the theme Language, Cognitions, Practices and Ergonomics within the LAM (Lutheries, Acoustics, Music), now attached to the Institut Jean Le Rond d'Alembert at Sorbonne University, Paris. She notably published the book "Sensory Experiences: Exploring meaning and the senses", in collaboration with Caroline Cance, Matt Coler, Arthur Paté and Catherine Guastavino (John Benjamins Publishing, 2021). This book is a synthesis of research in cognitive psychology and linguistics exploring the relationship between language and cognition in the different sensory modalities, in a conception of "situated" cognition, based on the meanings that different actors give to their sensory experiences (see: reading notes at the end of this issue).*

## Influence du réalisme, de l'agentivité et des représentations sur l'immersion sonore en réalité augmentée

Valentin Bauer

### Résumé

Les récentes évolutions technologiques des casques de Réalité Virtuelle (RV) et Augmentée (RA) permettent à de plus en plus d'individus de s'immerger et d'interagir dans des environnements multisensoriels virtuels. Alors que la RA permet d'ajouter des éléments virtuels à l'environnement réel, la RV présente un environnement virtuel qui coupe le lien avec la réalité. Dans les expériences virtuelles, l'immersion est souvent qualifiée de façon technique, par la capacité d'un système à immerger l'utilisateur. Cependant, et en particulier en RA où l'utilisateur perçoit les éléments virtuels ainsi que l'espace réel qui l'entoure, d'autres dimensions subjectives entrent en jeu dans le phénomène d'immersion, en lien avec le contenu, le contexte de présentation, la perception de l'utilisateur de pouvoir agir sur l'environnement (agentivité), et ses représentations. Du fait de ces aspects multiples, saisir ce qui se joue dans l'immersion en RA demeure un défi, et d'autant plus pour l'immersion sonore, qui reste assez inexplorée. Pour questionner le phénomène d'immersion sonore en RA, cet article étudie trois aspects : le réalisme perçu par l'utilisateur, l'agentivité de l'utilisateur, et ses représentations. Dans ce but, une comparaison est menée entre les résultats issus de quatre études de RA qui sont représentatives de domaines variés : retours casques pour musiciens lors d'enregistrements studio, expérience participative inspirée de pratiques théâtrales, expérience multisensorielle pour enfants autistes nécessitant un fort soutien, et production musicale spatialisée. Les résultats suggèrent différents paramètres de l'immersion liés aux concepts de réalisme, agentivité, et représentations, ainsi que des futures pistes de recherche.

### Abstract

*The recent technological advances of Virtual (VR) and Augmented (AR) Reality headsets enable more and more individuals to immerse into virtual multisensory interactive environments. Although AR enables to add virtual elements onto the real environment, VR displays a virtual environment which is entirely separated from reality. Regarding virtual experiences, immersion is often described technically, by the capabilities of the system to immerse the user. Yet, and particularly regarding AR when users perceive the virtual content as well as their real surroundings, other subjective dimensions arise in the immersion phenomenon, related to the content, the presentation context, the user perception that they can act on the environment (agency), and their representations. Due to these various aspects, understanding what AR immersion encompasses remains challenging, and even more regarding audio immersion, which is still under-explored to our knowledge. To question the audio AR immersion phenomenon, this study focuses on three main aspects: the user perception of realism, the user agency, and the user representations. To that end, a comparison is drawn between the findings emerging from four AR studies being representative of various fields: musicians' headphone monitoring for studio recording, a participative performance drawing from theatre practices, a multisensory experience designed for autistic children requiring substantial support, and a spatialized music production. Findings suggest several immersion parameters related to the concepts of realism, agency, and representations, as well as future research directions.*

Les récentes évolutions technologiques, notamment des casques de Réalité Virtuelle (RV) et Augmentée (RA), permettent à de plus en plus d'individus de s'immerger et d'interagir dans des environnements multisensoriels virtuels. Alors que la RA permet d'ajouter des éléments virtuels à l'environnement réel, la RV présente un environnement entièrement virtuel qui coupe le lien avec la réalité. Cet article utilise le terme de Réalité Étendue (XR) pour inclure ces deux extrêmes (RV et RA), ainsi que les expériences intermédiaires de Réalité Mixte où la proportion d'éléments réels et virtuels est variée (Milgram et Kishino, 1994).

Dans ces expériences virtuelles, l'immersion de l'utilisateur est souvent qualifiée de façon technique, par la capacité d'un système à offrir une fidélité de rendu selon plusieurs canaux sensoriels (par exemple : audio, visuel, haptique) et une coupure du monde réel (Slater et Wilbur, 1997). L'hypothèse sous-jacente est que plus les capacités techniques sont élevées, offrant donc un rendu *réaliste* et des possibilités d'interaction convaincantes, plus l'utilisateur sera engagé dans l'environnement virtuel et s'y sentira *présent*, donc avec l'impression d'y être. Cependant, comme relevé par Cummings *et al.* (2015), cette hypothèse se justifie-t-elle en lien avec l'objectif de l'expérience ?

En particulier, en RA, l'utilisateur peut être impliqué dans un environnement réel augmenté, tout en percevant les éléments réels qui l'entourent. À l'immersion technique se rajoutent donc d'autres dimensions subjectives, liées au contenu présenté (par ex. : narration), au contexte (par ex. : lieu), aux interactions proposées (par ex. : interactions intuitives), et aux caractéristiques de l'utilisateur (par ex. : fatigue) (Shin, 2019). Ces aspects influencent également la perception que l'utilisateur a de pouvoir agir sur l'environnement virtuel, c'est-à-dire son *agentivité* (Bandura, 2018),

et donc son comportement et son implication. L'immersion subjective qui en découle correspond au concept d'immersion sonore proposé par Agrawal *et al.* (2019, p.5) : « un état d'engagement mental profond dans lequel les processus cognitifs [...] provoquent un changement de l'état attentionnel pouvant amener à expérimenter une coupure de la conscience du monde physique ».

Alors que saisir ce qui est en jeu dans l'immersion en RA reste un défi, la place du sonore dans cette expérience demeure assez inexplorée. En effet, le récent article de revue sur l'immersion en RA ne précise le rôle du sonore que dans le contexte multisensoriel (Shin, 2019) et les questionnaires sur l'immersion en XR se concentrent sur les aspects visuels.

Cet article questionne ce qui est en jeu dans l'immersion sonore en RA, au travers de trois aspects :

- Quel réalisme sonore pour quelle immersion sonore en RA ?
- Quel est l'impact de l'agentivité sur l'immersion sonore en RA ?
- Quel est le lien entre l'immersion sonore et les représentations de l'utilisateur ?

Pour aborder ces questions, nous tisserons des liens entre quatre travaux de RA précédemment menés. Deux concernent une augmentation sonore du réel, dans le cadre de retours casques pour musiciens en studio (Bauer *et al.*, 2018) et d'expériences collaboratives inspirées de pratiques théâtrales (Nagele *et al.*, 2021), et deux autres concernent une augmentation audiovisuelle, dans un cadre clinique avec la conception d'une expérience pour enfants autistes<sup>1</sup> (Bauer *et al.*, 2021), et pour une production musicale spatialisée (Bauer et Bouchara, 2021). En comparant ces articles issus ►

**1**

Cet article utilise le terme « personne autiste » et non « personne avec autisme », conformément aux préférences de personnes autistes (Bottema *et al.*, 2021)

de domaines variés et qui considèrent des participants neurotypiques<sup>2</sup> et autistes, cet article adopte une perspective large, pour suggérer des résultats et pistes de recherches qui puissent s'appliquer à de nombreux domaines.

Avant de présenter une perspective nouvelle des méthodes et résultats issus de ces quatre articles, nous allons présenter des études passées qui discutent des concepts de réalisme, d'agentivité, et de représentations des utilisateurs dans le cadre d'environnements augmentés. Nous présenterons ensuite les systèmes utilisés dans les quatre études, puis la méthodologie employée par cet article. La partie suivante présentera les résultats en trois sous-parties, liées aux concepts de réalisme, agentivité, et représentations. Enfin, une dernière section discutera ces résultats pour dériver des idées pour de futures recherches.

## 1. État de l'art

En 1965, Ivan Sutherland (1968) réalise le premier système de RA en *see-through*, c'est-à-dire avec une caméra qui filme l'environnement extérieur et le rediffuse dans le casque, image à laquelle sont ensuite rajoutés des objets virtuels. Ce système participe de sa vision de la « machine ultime », qui « permettrait de contrôler l'existence de la matière. Une chaise disposée dans un tel espace serait assez bien pour s'y assoir. Des menottes affichées enfermeraient et une balle affichée dans une telle pièce serait fatale » (Sutherland, 1965, p.2). Cette volonté d'une réalité alternative pour immerger l'utilisateur a alimenté les recherches pendant de nombreuses années (Billinghurst et al., 2015). Cependant, elle s'oppose dès les débuts de la RA à la vision de Jaron Lanier qui évoque un « médium à l'intérieur duquel nous pourrions "improviser la

réalité" collectivement » (Lanier et al., 1989). Lanier met l'accent sur le sens de l'environnement, qui émerge de la création, et/ou de la collaboration active d'utilisateurs au sein d'un environnement virtuel, pas forcément réaliste. De cette seconde vision émerge *DataGlove*, un dispositif avec un gant connecté et un casque de RV, qui permet la première performance musicale en RV « *The sound of one hand*<sup>3</sup> » en 1997. Cette opposition entre une vision *technique* et *subjective* de l'immersion, amène à questionner ce qui est en jeu dans le phénomène d'immersion. Cet article tente d'y apporter une première réponse concernant l'immersion sonore, au travers de trois dimensions, à savoir, réalisme, agentivité, et représentations.

### Quel réalisme sonore pour quelle immersion ?

Selon Olson (1972), le réalisme sonore dépend de quatre conditions : un large spectre, une large gamme dynamique, une fidélité spectrale de rendu sonore, et un espace acoustique fidèlement reproduit. Cette description *technique* sous-entend l'hypothèse souvent admise, et décrite par Cummings et al. (2015) que, plus les capacités techniques sont élevées, plus le rendu sera réaliste, et plus l'auditeur sera impliqué et immergé. Cependant, comme le note Pike (2021), cette hypothèse est parfois contredite par des expériences d'écoute spatialisée au casque. En effet, des comparaisons à l'aveugle entre binaural et stéréo donnent des résultats mitigés quant aux préférences des auditeurs. Ceux-ci préfèrent toutefois les versions en binaural quand ils savent lesquelles sont en binaural, et notent alors un meilleur réalisme perçu (Pike, 2021). Ainsi, de meilleures capacités techniques concernant le réalisme sonore n'amènent pas toujours une meilleure immersion des auditeurs.

<sup>2</sup>

Le terme neurotypique désigne toute personne ne présentant pas de trouble du spectre autistique, et ici plus largement toute personne ne présentant aucune condition neuro-développementale.

<sup>3</sup>

Lien vers le site de Jaron Lanier « The sound of one hand » (1997) : <http://www.jaronlanier.com/vr.html>

Le fait qu'un meilleur réalisme de rendu n'implique pas toujours une meilleure immersion amène à questionner quelles variables contextuelles influencent notre perception du réalisme sonore dans un mode d'écoute donné. En particulier, quel degré de réalisme est suffisant pour permettre d'immerger l'auditeur dans un contexte donné ?

Une première réponse est que le degré de réalisme sonore doit s'adapter au message à transmettre, et donc prendre en compte les besoins, avis, et volontés de toutes les parties prenantes de l'expérience (diffuseur, ingénieur du son, utilisateur...). Par exemple, alors que des ingénieurs du son peuvent vouloir utiliser le format Audio Orienté Objet (OBA) (Tsingos, 2017) pour réaliser des productions sonores spatialisées plus réalistes qu'en stéréo, les diffuseurs sont actuellement plus intéressés par le potentiel d'accessibilité de ce format (par exemple, pour les troubles auditifs) que par la production spatialisée (Pike, 2021, p.13). Ensuite, le degré de réalisme de l'information sonore et multisensorielle doit s'adapter aux utilisateurs. Cette adaptation est notamment critique dans le cas d'utilisateurs autistes, qui ont très souvent des difficultés d'intégration multisensorielle (Robertson et Baron-Cohen, 2017). Pour ces personnes, simplifier le réalisme de l'information pour donner uniquement les éléments nécessaires permet de favoriser l'engagement et l'immersion, alors que présenter une information multisensorielle plus réaliste nuira à la compréhension (Bauer *et al.*, 2022 ; Tarantino *et al.*, 2019). Cet article compare les résultats de deux études de RA qui adaptent respectivement le réalisme de rendu à un contexte de retours casques en studio et une expérience multisensorielle en hôpital de jour pour enfants autistes.

Une seconde réponse à cette question est que le degré de réalisme doit s'adapter à l'objectif du projet à réaliser.

Ceci est notamment suggéré par Barrett (2021, p.178) dans le cadre de composition musicale spatialisée, lorsqu'elle dit « Les sons peuvent être conçus pour se lancer autour de l'auditeur à des vitesses non naturelles, et la variété de l'immersion n'a pas d'homologue réel ». Dans cette démarche, recréer une situation réaliste concernant les trajectoires sonores n'a pas de sens. En revanche, puisque la création artistique générale vise à immerger l'auditeur, rendre de façon réaliste d'autres paramètres comme la qualité des sources sonores peut être intéressant pour le sens général de l'œuvre. La démarche de production musicale spatialisée en RA qui est discutée dans cet article partage cette vision.

### **Quelle agentivité en lien avec le sonore ?**

Alors que la littérature de XR considère souvent les utilisateurs comme des « agents statiques recevant passivement » l'immersion, les utilisateurs ont au contraire un rôle actif dans la formation de celle-ci (Shin, 2019). Cette considération amène à changer de point de vue, pour dépasser le concept d'immersion comme but à atteindre, couramment utilisé dans les mesures actuelles, et le penser au contraire comme un processus dynamique, réflexif, continu et interactif (Shin, 2019). L'étude de l'immersion subjective nécessite donc de considérer l'agentivité de l'utilisateur, en tant qu'acteur d'une performance dans un contexte spécifique.

Dans les expériences de XR, l'utilisateur interagit dans un cadre multisensorimoteur, contrôlé, et interactif. L'engagement du corps en XR est donc proche de celui permis par les installations audiovisuelles artistiques (Bauer et Bouchara, 2021). En effet, dans ces installations, l'utilisateur est considéré comme un acteur qui expérimente au travers d'interactions variées (par ex. : gestuelle, verbale). Dans ces installations, ►

comme dans les productions artistiques en XR, un déroulement linéaire prédéterminé de la performance peut être combiné à des interactions prévues par l'artiste, qui permettent à l'utilisateur de découvrir activement les différentes potentialités l'œuvre et d'en modifier le déroulement (Buckley et Carlson, 2019). Pour saisir le rôle de l'agentivité dans l'immersion ressentie, un rapprochement entre la performance en RA et dans le théâtre immersif peut également être fait, et notamment avec le *Theatre of the Oppressed* d'Augusto Boal (2002). Dans celui-ci, les spectateurs sont partie-prenantes de l'œuvre, afin de transmettre et co-crée des aspects narratifs, en lien avec le contexte de la performance. Redéfinir ainsi les liens entre spectateurs et acteurs, avec le contexte de création, vise à générer du sens en explorant des identités et des situations nouvelles (Boal, 2002). Au travers des quatre études présentées, nous allons examiner le lien entre immersion et agentivité.

Deux variables contextuelles générales influencent l'immersion sonore (Sungyoung, 2021) et donc aussi l'agentivité de l'utilisateur : le contexte culturel (par ex. : passé, attentes) et l'environnement d'écoute (par ex. : modes d'écoutes, lieu). En particulier, considérer l'environnement d'écoute peut aller jusqu'à l'inclure lors de la conception de l'environnement immersif. Par exemple, dans le cas de composition sonore spatialisée, Barrett (2021, p.176) note : « le haut-parleur comme objet sonore peut néanmoins être d'un intérêt potentiel pour des installations sonores artistiques sur site ». Concernant le théâtre au casque, Klich *et al.* (2017) mentionnent l'importance de considérer le casque comme objet de la performance. L'importance de la prise en compte du médium de RA lors de la conception est évoquée par trois études analysées

ici, respectivement sur la production musicale, la médiation multisensorielle pour enfants autistes, et les retours casques pour musiciens.

En plus de ces deux variables majeures, considérer les caractéristiques propres de l'utilisateur est essentiel lors de la conception. Ceci inclut par exemple son état au moment de l'expérience, qui influera sur son interaction (par ex. : fatigue), et ses besoins et capacités, afin d'ajuster l'interaction proposée. Cette prise en compte vise à éviter deux extrêmes : offrir trop d'interactions qui pourraient amener une charge cognitive, et ne pas en offrir assez ce qui pourrait rendre l'utilisateur passif et donc nuire à son implication. Cet aspect est notamment critique pour des utilisateurs autistes qui rencontrent souvent des difficultés à appréhender la nouveauté (Sinha *et al.*, 2014). Dans leur cas, un juste niveau de nouveauté doit permettre de favoriser l'exploration sans submerger la personne. Cet ajustement des interactions disponibles et de la nouveauté est également nécessaire pour des individus neurotypiques, comme dans le cas de musiciens qui enregistrent en studio. En effet, ceux-ci ressentent souvent une perte de contrôle sur leur jeu musical, car ils doivent se reposer entièrement sur l'écoute au casque, même si celle-ci peut être perçue comme gênante (Bauer *et al.*, 2018). Nous évoquerons ici cette prise en compte des caractéristiques de l'utilisateur notamment au travers de la conception d'une expérience de RA avec des enfants autistes et d'une expérience d'augmentation en binaural avec des musiciens enregistrant en studio.

## Représentations en jeu dans les situations d'augmentation

Selon les principes de la cognition incarnée (Wilson, 2002), l'interaction d'un individu dans un environnement implique deux processus : d'une part, ses représentations de l'environnement lui font adapter ses comportements et influencent ses réactions émotionnelles ; d'autre part, les processus multi-sensorimoteurs lors de l'interaction, et les émotions qui émergent, peuvent modifier ses représentations. Déterminer l'objectif principal visé en RA va donc influencer la conception de l'environnement. Par exemple, le but peut être de permettre aux utilisateurs d'interagir naturellement et sans effort avec un environnement directement compréhensible, car lié à leurs représentations. À l'inverse, l'objectif peut être de questionner leurs représentations en introduisant des éléments inattendus. Entre ces deux extrêmes, un continuum existe où la proportion d'éléments en lien ou non avec les représentations est variée. Dans notre cas, nous analyserons les résultats de quatre études qui varient différemment cette proportion, dans des contextes et avec des utilisateurs variés.

Concevoir un dispositif implique aussi de prendre conscience des représentations que les utilisateurs ont de celui-ci. En particulier, l'utilisation de casques audio ou de XR peut évoquer une expérience intime ou une coupure avec le monde extérieur. Comme le dit Barrett (2021, p.176) dans le cadre de concerts de musique spatialisée : les « casques empêchent l'expérience sociale ». Cette vision peut induire des réticences de la part des utilisateurs pour adopter le dispositif. Par exemple, des praticiens de la clinique de l'autisme peuvent être réticents à l'utilisation de casques de XR, car les enfants avec qui ils travaillent présentent déjà

des difficultés d'interaction sociale (Bauer *et al.*, 2022).

Enfin, prendre conscience des représentations existantes permet de les questionner au travers de la XR. Plusieurs études notent par exemple que la RA permet de questionner les concepts de réel et de virtuel. Comme dit Shin (2019, p. 11), la XR, « dans différentes plateformes, questionne le concept de réalité et nous permet de redécouvrir et d'explorer la réalité. » Cette exploration peut également être menée au travers de performances sonores où « les différences entre *réel* et *construit* [...] font partie de la recherche artistique » (Barrett, 2021, p.177). Questionner les représentations peut également concerner l'image de soi ou l'agentivité de l'utilisateur, comme étudié par Isnard *et al.* (2021) dans le cadre d'un scénario narratif de RV avec une modification en temps réel de la voix de l'utilisateur. Puisque la RA permet d'offrir un cadre multisensoriel contrôlé, elle présente donc un fort potentiel pour modifier nos représentations, en lien avec les principes de la cognition incarnée, et de ceux également développés par Tim Ingold (2002), selon qui les informations permettant de construire une représentation sont contenues dans la structure de la sensation. Ce travail de questionnement des représentations est amorcé ici au travers des quatre études présentées. ►

## 2. Méthode

Une méta-analyse de quatre études en RA audio ou audiovisuelle quant à l'immersion de l'utilisateur est proposée (Bauer *et al.*, 2018 ; Bauer et Bouchara, 2021 ; Nagele *et al.*, 2021 ; Bauer *et al.*, 2021). Les études varient en termes d'utilisateurs et de contextes : *BMR* concerne des retours casques pour musiciens d'improvisation lors d'enregistrement studio (Bauer *et al.*, 2018) ; *Please Confirm you are not a Robot* (Not A Robot) est une performance sonore participative inspirée du théâtre (Nagele *et al.*, 2021) ; *Magic Bubbles* est un environnement multisensoriel complémentaire aux pratiques de médiation dans la clinique de l'autisme (Bauer *et al.*, 2021) ; et *REM* est une production musicale spatialisée (Bauer et Bouchara, 2021). Le tableau 1 résume les caractéristiques des quatre systèmes d'augmentation du réel utilisés. Les études sont comparées selon trois dimensions, dérivées des études de Shin (2019) et Agrawal *et al.* (2019), à savoir le réalisme, l'agentivité et les représentations. Dans cette section, nous décrivons succinctement les systèmes et protocoles utilisés dans ces quatre études, les descriptions détaillées se trouvant dans les articles respectifs.

### Présentation des différents systèmes d'augmentation du réel

BMR vise à améliorer les expériences souvent négatives de musiciens lorsqu'ils utilisent des systèmes de retours casques en stéréo lors d'enregistrements studio, grâce au rendu binaural dans le cas de musiciens d'improvisation. Le système KLANG:fabrik (KF) a été choisi pour le rendu binaural, du fait de sa qualité sonore et de sa faible latence de 2,6 ms. Il permet un mixage de type Audio Orienté Objet (OBA). Cette étude



Figure 1 : Musicien utilisant le système BMR lors des tests.



Figure 2 : Quatre utilisateurs testant la performance *Please Confirm you are Not a Robot*.

n'utilise pas de système de suivi des mouvements de la tête (*headtracking*), car l'hypothèse initiale était que le démasquage sonore permis par l'écoute binaurale sans *headtracking* permettrait déjà d'améliorer l'expérience de retours casques en stéréo. La figure 1 présente un musicien utilisant le système BMR.

*Not a Robot* est une performance participative pour quatre utilisateurs en RA audio, inspirée de jeux décrits par Boal (2002). Le narrateur virtuel *Pi* conduit les utilisateurs au travers de quatre mini-jeux qui visent

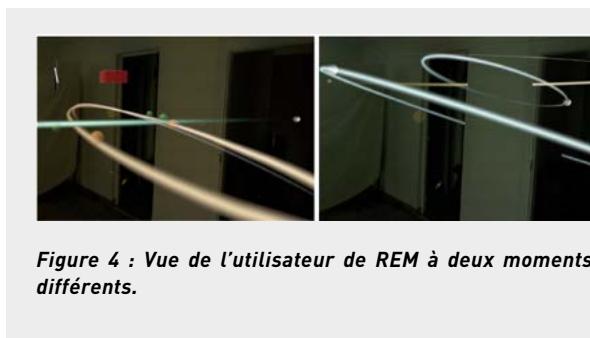


**Figure 3 : Visuels de Magic Bubbles ajouté à l'environnement réel.**

à interférer avec les dynamiques de groupe et à porter attention à l'impact négatif des outils numériques sur leur bien-être. Les lunettes *Bose Frames* sont utilisées pour la diffusion sonore, car elles offrent une transparence acoustique, trois interactions (hocher la tête, secouer la tête, toucher le côté des lunettes) et des capacités de *headtracking*, même si avec une latence perceptible (~200 ms). Développée avec Unity3D, cette production sonore en OBA est diffusée avec un rendu binaural à partir d'un codage ambisonique d'ordre trois grâce à *Google Resonance Audio* (Gorzel et al., 2019). Le rendu sonore est asymétrique, c'est-à-dire qu'il varie en fonction des utilisateurs. La figure 2 présente une photographie de la performance. Une vidéo est visible à ce lien<sup>4</sup>.

*Magic Bubbles* est un environnement de RA multisensoriel qui vise à sécuriser et rendre disponibles des enfants autistes nécessitant un fort soutien (avec déficience intellectuelle et peu ou pas verbaux). Sa conception est issue de 34 entretiens avec des personnes de la communauté de l'autisme, puis d'une collaboration avec deux psychologues cliniciens et une médecin psychiatre. La plateforme utilise de la RA en *see-through*, c'est-à-dire qu'une caméra stéréoscopique Zed Mini sur un casque HTC Vive-Pro capte l'environnement réel et le rediffuse dans le casque. Le mixage sonore est en

OBA et la diffusion sonore se fait avec les écouteurs du casque. Un écran permet au praticien de contrôler ce que perçoit l'enfant en temps réel. Le dispositif est



**Figure 4 : Vue de l'utilisateur de REM à deux moments différents.**

conçu pour être utilisé avec des praticiens dans la salle de psychomotricité d'un hôpital de jour. La figure 3 présente les visuels ajoutés à l'environnement réel que l'enfant perçoit. Une vidéo est disponible à ce lien<sup>5</sup>.

Enfin, *REM* explore les nouvelles possibilités expressives de la RA concernant la production musicale électronique spatialisée, c'est-à-dire par l'ajout de visuels et d'interactions. Les interactions consistent à se baisser pour actionner un filtre passe-bas, se déplacer dans l'espace pour déclencher des délais sonores, et regarder les sources pour leur faire changer de trajectoire. La plateforme de RA est identique à *Magic Bubbles*. ►

**4**  
Lien de Not a Robot :  
[https://valentinbauer.com/research/bose\\_frames/](https://valentinbauer.com/research/bose_frames/)

**5**  
Lien de la vidéo de Magic Bubbles :  
<https://valentinbauer.com/research/doctorat/>

**6**  
Un vidéo de présentation de REM est visible à ce lien :  
<https://valentinbauer.com/research/ar-music/>

Contexte général		Plateforme d'augmentation			Contexte d'augmentation		
	Objectif	Participants	Plateforme de travail	Rendu sonore	Diffusion	Lieu	Immersion
BMR	Retours casque pour musiciens en studio	1 solo et 2 trios	Station de travail audio-numérique ; Microphones de proximité	KLANG:fabrik - HRTFs anéchoïques propriétaires	Casques fermés	Salles de studio isolées et sèches	RA audio sans headtracking
Not a Robot	Performance participative inspirée de pratiques théâtrales	4 groupes de 4 personnes	Unity3D ; Sons préenregistrés	Google Resonance Audio - HRTFs propriétaires	Lunettes Bose Frames	Tout lieu avec peu de bruits ambiants	RA audio avec headtracking
Magic Bubbles	Mediation sensorielle destinée aux enfants autistes	Conception avec 3 praticiens puis tests avec 11 praticiens	Unity3D ; HTC Vive Pro et camera Zed Mini ; sons préenregistrés	Google Resonance Audio – HRTFs propriétaires	HTC Vive Pro (visuel et audio)	Salle de psychomotricité	RA audiovisuel en see-through
REM	Composition musicale spatialisée en RA	7 participants	Unity3D ; HTC Vive Pro ; Zed Mini ; Sons préenregistrés	Google Resonance Audio – HRTFs propriétaires	HTC Vive Pro (visuel et audio)	Tout lieu avec peu de bruits ambiants	RA audiovisuel en see-through

Tableau 1 : Caractéristiques des quatre systèmes d'augmentation du réel comparés.

## Protocoles d'évaluation

BMR a été évalué lors de trois tests, avec un musicien de « musiques du monde » faisant du re-recording, un trio de jazz (guitare, contrebasse, batterie), et un trio d'improvisation libre (clarinettes, percussions, accordéon). Tous les musiciens étaient pré-professionnels et avaient une connaissance des situations d'enregistrement en studio avec retours casques. Les tests étaient contrebalancés concernant les conditions d'écoute en binaural et stéréo, et débutaient par l'une pour finir par l'autre. Après chaque test, trois questions semi-dirigées étaient posées aux musiciens, concernant leur : confort d'écoute, perception de différences entre ces deux conditions (notamment liées à leur jeu musical) et de leur propre son. L'analyse qualitative utilisait la *grounded theory* (Charmaz, 2006). Cette méthode consiste à coder les données des entretiens en phrasés signifiants, et à les regrouper ensuite en concepts et catégories. Une comparaison constante entre les données, phrasés, concepts, et catégories permet d'ajuster ceux-ci progressivement via de multiples itérations.

*Not a Robot* a été testé par 16 participants avec différents niveaux d'expertise concernant l'audio spatialisé et le théâtre immersif. Après chaque mini-jeu, des questions concernaient l'engagement du participant et ses sensations. Après le test, 21 questions concernaient l'engagement, la perception de la voix du narrateur, la conscience de soi, le sentiment de présence, la conception du jeu et le lien avec le groupe. Les questionnaires mêlaient à chaque fois des questions quantitatives et ouvertes. Après l'expérience, une discussion guidée concernait l'utilisation du système pour d'autres contextes, la perception de divers aspects

de l'expérience (par exemple : son 3D), et l'aspect réflexif autour des outils numériques. L'analyse utilisait des méthodes mixtes, en mêlant la *grounded theory* et une analyse quantitative.

*Magic Bubbles* a été testé par 11 praticiens d'un hôpital de jour pour enfants spécialisé dans la prise en charge de l'autisme. Le but était de valider l'utilisation future de *Magic Bubbles* avec des enfants autistes de l'hôpital de jour. Les praticiens utilisaient l'application librement pendant environ 5 min, pendant lesquelles des notes étaient prises quant à leur comportement. Après l'expérience, sept questions semi-dirigées concernaient des aspects démographiques, l'expérience utilisateur et leurs préférences dans l'environnement. Trois questions ouvertes concernaient enfin leur sentiment général, des idées pour améliorer la conception ou d'autres éléments qu'ils voudraient évoquer. L'analyse des données était qualitative et descriptive.

*REM* a été composée pour explorer le processus de production musicale en RA. L'étude détaille ses étapes de pré-production et production et présente les retours de sept personnes qui ont donné leur avis pendant la phase de production. Ceux-ci avaient uniquement regardé une vidéo de la production car la situation sanitaire avec la Covid ne permettait pas de mener des tests « avec le système de RA » au moment de l'étude. Les retours sont analysés de façon qualitative et descriptive. ►

### 3. Mise en parallèle des résultats de quatre études de réalité augmentée quant à l'immersion perçue

Les résultats des quatre études en lien avec l'immersion sont successivement analysés ci-dessous selon trois dimensions : réalisme, agentivité et représentations.

#### Perception du réalisme

Dans le cas de BMR, le positionnement des sources sonores en binaural est non réaliste car il ne correspond pas au positionnement réel des musiciens en studio. En effet, il ne visait pas à reproduire fidèlement les positions réelles, mais à répondre aux besoins des musiciens quant à la qualité sonore, en bénéficiant d'un meilleur démasquage sonore permis par l'écoute binaurale en comparaison avec la stéréo. Cependant, tous les musiciens décrivent un meilleur réalisme perçu en binaural. Lors des entretiens, le réalisme évoqué témoigne notamment d'une fidélité de rendu sonore, au travers de trois dimensions. Tout d'abord les musiciens perçoivent les autres sources sonores avec plus de définition, de clarté, de subtilités, « de variété » (clarinettiste). En particulier, ceci est lié à l'écoute binaurale moins « compacte » contrairement à la stéréo « qui fait bloc » (musicien solo). Ensuite, les musiciens perçoivent mieux leur propre son. Par exemple, en binaural, le contrebassiste note un meilleur « contact physique/auditif », et le musicien solo que « ce n'était pas indispensable [...] d'avoir l'acoustique de ma voix ». Le musicien solo gardait en effet le casque en binaural alors qu'il l'enlevait en stéréo pour entendre résonner sa voix dans la salle. Enfin, deux des sept musiciens évoquent un plus grand naturel, plus proche d'un « instinct de l'oreille » (musicien solo). En plus de cette

qualité sonore accrue, six musiciens perçoivent une externalisation réaliste : « C'était comme si la musique était réelle autour de moi » (musicien solo).

Avec *Not a Robot*, la latence importante du « système de *headtracking*» (~200 ms) empêche de percevoir un positionnement sonore réaliste. Elle gêne la plupart des utilisateurs (11 sur 16) et entraîne des difficultés de localisation, notamment dans le mini-jeu où ils doivent identifier les sons dans l'espace puis les éteindre en interagissant avec les lunettes. Pourtant, dans le questionnaire post-expérience, 11 des 16 participants notent être captivés par la performance. De plus, l'environnement est jugé comme modérément (11/16) ou totalement (1/16) réaliste. Dans le cas du réalisme modéré, deux éléments nuisent au réalisme : voir l'environnement réel et en être conscient (6/16) et la qualité sonore limitée (2/16). Au contraire, deux éléments le favorisent : l'expérience de groupe (3/16), et le fait que la voix du narrateur « semble être la voix du contexte, du bâtiment, de l'institution, la voix de la BBC » (1/16). Le participant ayant perçu l'environnement comme totalement réaliste précise : « scénario de livre audio même si nous étions autour pour faire des choses bêtes. » Ainsi, malgré une localisation sonore non précise, trois variables permettent de percevoir un réalisme sonore : le lien entre les sons et le contexte, l'expérience de groupe et la narration.

Lors de la conception de *Magic Bubbles*, les 34 entretiens, puis la collaboration avec les praticiens, ont amené à simplifier l'environnement virtuel pour pouvoir l'utiliser avec les enfants autistes. Cette simplification correspond à utiliser peu de stimuli sonores et visuels ainsi que des visuels géométriques simples, et a été validée lors des tests. Pendant ces tests, les praticiens ne mentionnent jamais le concept de réalisme,

mais suggèrent au contraire une simplification des interactions proposées afin de présenter des actions-réactions claires. Dans ce contexte, la notion de réalisme a donc peu de sens. Cependant, cette simplification n'empêche par deux des onze praticiens de se sentir immergés, évoquant par exemple être « dans un autre monde » (éducateur). L'immersion peut être liée au réalisme de l'intégration des éléments virtuels dans le monde réel, bien que non mentionné par les praticiens, et à leur perception de leur capacité d'agir sur les éléments (agentivité), développée ci-après.

*REM* vise à immerger l'utilisateur en augmentant son écoute habituelle au casque par des visuels, pour lui donner l'impression que la musique existe réellement autour de lui. L'externalisation binaurale participe de cette réalité alternative. Dans ce sens, le fait de voir les sons autour de soi, en plus d'avoir un but esthétique, favorise leur externalisation. Les éléments visuels sont des formes géométriques simples avec des couleurs et formes en lien avec les sons, mais dissociées de l'environnement réel. Tous les participants ayant vu la performance l'ont appréciée, mais sans mentionner un aspect réaliste. Au contraire, leurs commentaires se concentraient sur le sens créé par le lien entre sons et visuels. Les notions évoquées étaient en lien avec l'agentivité et les représentations, développées ci-après.

## Agentivité

Trois dimensions de l'agentivité sont apparues lors de l'analyse des quatre études : le lien avec le cadre général d'utilisation, le lien avec d'autres acteurs, et l'impact de l'agentivité sur la création de nouveaux contenus. Ces trois dimensions sont développées ci-dessous.

### Agentivité dans un cadre structuré

Avec *BMR*, tous les musiciens décrivent un meilleur confort d'écoute dans ce contexte de retours casques. Le percussionniste note : « le binaural dans ce cadre-là [...] c'est très agréable ». La notion de *cadre* concerne le dispositif technique (lieu, équipement...), les personnes en présence (musicien et ingénieur du son) et l'objectif de la séance (enregistrer des morceaux). Dans ce cadre non naturel pour les musiciens, le binaural leur permet d'être plus *acteurs* de la situation, au travers d'un meilleur contrôle sur leur performance. Ce meilleur contrôle est notamment lié au meilleur confort d'écoute. Par exemple, le batteur note être « moins obligé d'être dans la tête » pour écouter, et donc plus immergé dans la performance. Ce meilleur contrôle favorise alors le jeu musical : « j'avais moins à me focaliser dessus, ça [*les idées musicales*] venait naturellement » (clarinettiste). Le dispositif peut ainsi être « oublié » (2/7), contrairement à la condition en stéréo où le guitariste adapte sa performance au dispositif et note « jouer par rapport à un son » entendu.

Dans *Not a Robot*, la narration définit la structure générale. Les participants acceptent les règles et tentent de les suivre (9/16), tout en gardant une liberté d'action. En effet, 14 des 16 participants disent suivre les instructions et environ la moitié les questionne (9/16). La majorité des participants fait donc l'effort de suivre les règles pour s'immerger dans la performance, car il faut suivre le « maître du jeu », ou pour « jouer et faire fonctionner le tout ». Cependant, les règles sont questionnées lorsqu'elles manquent de clarté (9/16) ou qu'une charge cognitive apparaît (7/16). Ainsi, de par leur contrôle dans ce cadre narratif structuré, les utilisateurs participent activement au phénomène d'immersion. Cependant, une non-compréhension du cadre nuit à ►

l'agentivité (3/16) et à l'immersion.

Le cadre de *Magic Bubbles* est celui d'un contexte clinique avec des enfants autistes nécessitant un fort soutien. La conception collaborative a révélé l'importance de proposer des actions-réactions simples pour rendre l'enfant acteur. Le test avec 11 praticiens a validé la plupart des interactions, mais révélé des améliorations possibles. En particulier, la possibilité de modifier l'environnement est apparue au travers de dessins (4/11), d'enregistrements (3/11) ou de dessins réels faits hors RA puis rajoutés dans l'environnement (4/11). De plus, certaines interactions ont été simplifiées. Par exemple, il a été conseillé de simplifier les deux bulles pour enregistrer et jouer le son, avec une seule qui enregistre et rejoue directement le son. Ces deux éléments (possibilités de création et simplification d'interactions) visent à favoriser l'implication et l'immersion de l'enfant par une meilleure compréhension des possibilités offertes par le cadre de RA.

Le déroulement général de *REM* est prédéterminé par le compositeur, mais l'utilisateur peut le modifier au travers de certaines interactions qui permettent de contrôler des effets sonores (comme les délais sonores, en explorant la limite de la zone d'interaction). Ces interactions visent à favoriser l'immersion de l'utilisateur en lui donnant un rôle actif dans la performance.

### **Agentivité avec d'autres utilisateurs**

*BMR*, *Not a Robot*, et *Magic Bubbles* visent à créer du lien avec d'autres utilisateurs au travers d'un espace virtuel intermédiaire. Dans chaque cas, cette collaboration participe à l'immersion.

*BMR* a permis aux musiciens de mieux se connecter avec les autres lors de la performance. En particulier, le batteur et le contrebassiste du trio jazz notent une meilleure entente basse/batterie en binaural, permettant un meilleur swing : « dès que tu mets de l'informatique [...] ça devient compliqué d'avoir cette cohésion [basse/batterie]. Et là pas du tout. Je trouve que le tout premier morceau [...] il swingue » (batteur). De plus, certains musiciens évoquent le fait de jouer ensemble dans le même espace virtuel. En effet, la clarinettiste dit : « ça me recréait un espace où on était tous présents dans ma tête ». Cette possibilité d'être acteurs en tant que groupe favorise donc une co-immersion dans la performance.

L'aspect le mieux reçu par les participants de *Not a Robot* est l'interaction via l'usage du corps en tant que groupe, qui inclue trois éléments principaux : interagir socialement, pouvoir briser la glace avec d'autres participants et s'amuser. En plus de la narration, les participants se sentent donc acteurs ensemble de la performance. Cette cohésion a pu influencer sur le fait que la majorité des utilisateurs (11/16) se sente captivée par la performance.

Dans *Magic Bubbles*, l'interaction enfant-praticien correspond à l'enfant qui porte le casque de RA et voit son praticien réel dans la salle de psychomotricité, qu'il connaît et qui peut le guider verbalement et/ou physiquement (tout en visualisant sur un écran ce que voit l'enfant). En effet, dans le cas de ces enfants, l'intérêt permettant ensuite d'explorer et de s'immerger dans une tâche est souvent suscité par un praticien connu de l'enfant. Les tests visaient donc particulièrement à vérifier que le dispositif favoriserait ce lien, d'autant plus que certains praticiens imaginaient que ce dispositif pourrait être enfermant pour ces enfants.

Puisque 10 des 11 praticiens se sont sentis connectés avec l'environnement réel lors des tests, le dispositif a été jugé positif pour mener des tests futurs avec les enfants.

### ***Impact de l'agentivité sur la création***

L'agentivité en tant qu'individu ou groupe a permis de générer de nouveaux éléments.

Dans le cas des retours casques, le plus grand contrôle perçu en binaural a impacté la création musicale, avec une influence positive sur la performance (5/7) et la créativité (6/7). En particulier, le trio d'improvisation a réalisé en binaural son enregistrement le plus long et avec le plus de dynamique.

Concernant *Not a Robot*, certains utilisateurs ont pris l'initiative de jouer un rôle. En particulier, pendant le mini-jeu qui les incitait à s'imiter à deux en lien avec les instructions reçues, 5 des 8 binômes ont discuté le scénario à improviser ensemble. Ces initiatives apparaissent aussi à la fin, lorsque la narration exclut un participant du groupe. Le fait d'être exclu et d'entendre des informations sonores que les autres n'entendent pas déclenche souvent le fait de jouer un rôle. Par exemple, un participant dit : « Et maintenant je dois tous vous ramener à la réalité, en prenant vos lunettes, et ensuite vous serez réellement libérés. ».

Concernant *Magic Bubbles*, les psychologues ayant collaboré à la conception ont insisté sur l'importance du jeu libre (sans scénario prédéterminé) pour permettre aux enfants de prendre des initiatives, « s'autoriser à » et favoriser l'émergence de comportements non prévus. Le jeu libre est, en effet, souvent utilisé dans les activités de l'hôpital de jour (dessin, etc.).

### **Représentations de l'utilisateur**

L'analyse des quatre études a révélé deux aspects du lien entre représentations et immersion :

1. Une conception de l'expérience proche des représentations des utilisateurs permet de favoriser leur implication. Ceci concerne *BMR* et *Magic Bubbles*, mais aussi certains aspects interactifs et narratifs de *Not a Robot*, et est détaillé dans la 1ère sous-section.
2. Un écart perceptif, entre les attentes de l'utilisateur et les réactions de l'environnement, pensé lors de la conception, ou émergeant lors des tests, peut favoriser l'immersion. L'impact de cet écart est détaillé dans la 2nde sous-section, en lien avec les quatre études.

### ***Concevoir l'expérience en partant des représentations***

*BMR* fait le lien avec les habitudes des musiciens d'enregistrements en studio, car l'équipement utilisé est le même en binaural qu'en stéréo. De plus, les musiciens des deux trios notent que le binaural permet de renouer avec une situation d'écoute familière, liée au contexte de répétition. Par exemple, le batteur avait demandé à positionner le bassiste sur sa droite en binaural, alors qu'il le voyait en face de lui. Pendant l'entretien, il explique : « j'avais l'impression d'avoir le contrebassiste à ma droite ce qui est le cas quand on joue en ...[répétition] ». Pour deux musiciens d'improvisation libre, cette perception familière se manifeste par l'impression d'être ensemble dans un espace virtuel partagé. Enfin, malgré un confort d'écoute accru en binaural, le musicien solo met longtemps à s'habituer à cette écoute, étant « déstabilisé par ce confort », du fait de sa grande habitude des retours en stéréo. ►

Dans *Not a Robot*, les interactions avec les lunettes *Bose Frames*, pour commencer et terminer chaque mini-jeu et coordonner la narration, sont jugées intuitives. De plus, la majorité des participants suit les instructions du narrateur virtuel (14/16) car elles sont nécessaires au déroulé de la narration, comme dans un « jeu de société virtuel ». Cependant, la conception n'a pas pris en compte certaines représentations. En effet, une charge cognitive apparaît (7/16), due à des problèmes de détection et un manque de feedback sonore (5/16), qui empêche de comprendre si l'interaction est correctement détectée. Un participant note : « Je n'étais pas sûr à certains moments si j'agissais de la bonne manière. » De plus, les observations et discussions de groupe montrent des difficultés à identifier les sons venant du sol. En effet, les participants ne s'attendaient pas à entendre des sons avec cette élévation négative, puisque non habituelle dans la vraie vie.

*Magic Bubbles* a été conçu à partir de 34 entretiens avec des personnes de la communauté de l'autisme, puis en collaboration avec trois praticiens cliniciens, pour être au plus près des représentations d'enfants autistes. Ce processus a fait émerger quatre principaux critères de conception : (1) l'environnement doit être inspiré de pratiques cliniques courantes, (2) le contenu doit être individualisé, (3) la RA est préférée à la RV pour favoriser le lien à l'autre, (4) et l'environnement doit être simplifié. Le premier critère a notamment amené à choisir du jeu libre, souvent utilisé en hôpital de jour, donc habituel pour ces enfants. Concernant le second critère, la bulle musicale a été individualisée pour intégrer un morceau musical aimé par l'enfant. De plus, tous les sons des bulles peuvent être ajustés en temps réel. Pour le troisième critère, la RA a été choisie pour que l'enfant voit sa salle réelle avec son thérapeute habituel, et donc être proche de ce qu'il connaît. Le dernier critère

correspond à la simplification effectuée pour favoriser l'agentivité de l'enfant.

*REM* part de la situation familière d'écoute au casque pour la questionner, en rajoutant des représentations visuelles correspondant aux représentations du compositeur. Cependant, la plupart des participants ne trouve pas que le lien entre audio et visuel soit signifiant, ce qui amène une confusion et charge cognitive.

### **Impact de l'écart perceptif pour modifier les représentations**

*BMR* visait à améliorer l'expérience de retours casques en stéréo grâce au binaural, en restant proche des habitudes de studio pour ne pas déstabiliser les musiciens. Cependant, le batteur et le guitariste du trio jazz notent que ce nouveau système a influencé positivement le jeu musical, par son ajout à leurs habitudes d'écoute. Le batteur dit : « dès que tu changes un truc dans ce que tu fais tous les jours habituellement, ça te donne un petit regain d'attention ». Cette attention nouvelle, permise par l'écart entre leurs habitudes et cette nouvelle écoute, a donc favorisé une exploration musicale nouvelle et l'émergence d'idées créatives :

« ça nous a fait un peu "Waah c'est trop cool !" Et si je fais ça, ça fait quoi ? Waah ça marche ! Et si je fais ça ? Enfin tu vois, c'est l'intérêt supplémentaire mais sans que ça nuise au son » (batteur).

Pour *Not a Robot*, trois écarts perceptifs apparaissent : (1) le fait que les participants ne s'attendent pas à entendre différentes informations sonores ; (2) le doute des

participants sur le fait que les sons entendus soient réels ou virtuels ; (3) le fait d'utiliser une nouvelle technologie, amenant un écart entre l'attente des participants et ce qu'ils découvrent. Le premier écart a poussé les participants à interagir ensemble, avec plus de communication non-verbale que dans la vie réelle, pour s'assurer d'entendre toutes les instructions du narrateur. Un participant écrit par exemple : « A poussé les barrières sociales. Je me suis basé sur mon langage corporel au lieu de signaux verbaux ». Ensuite, l'écart entre réel et virtuel est lié à certains sons qui étaient perçus comme réels, car reliés au lieu de l'expérience (le réfectoire de la *BBC Broadcasting House*). Un participant a, par exemple, vérifié que le téléphone virtuel entendu venait de sous un fauteuil, tandis qu'un autre participant note qu'il « pouvait entendre quelqu'un faire tomber une fourchette mais ne pouvait pas dire si ça venait des lunettes ». Lors de la discussion, un intérêt pour plus d'expériences qui questionnent le lien entre réel et virtuel apparaît et les participants suggèrent d'ajouter des objets physiques connectés pour pousser cette exploration plus loin. Enfin, l'écart lié à la nouveauté de l'expérience pousse les participants à chercher un sens et, dans ce processus, à co-créer la narration en tant que groupe. Par exemple, 5 des 16 participants décrivent positivement cette expérience « originale », « incitant à la réflexion » et « différente de toute autre expérience ».

Dans *Magic Bubbles*, un premier écart correspond à l'ajout d'imprévu par le praticien dans un environnement structuré pour favoriser l'exploration de l'enfant. Pour ajouter de la surprise sans le submerger, l'imprévu est adapté au rythme de l'enfant par le praticien. Par exemple, lorsque l'enfant portera le casque et explorera peu, le praticien pourra déclencher des sons au travers de l'interface dédiée sur l'ordinateur. Dans ce cadre, les

éléments sonores ont un intérêt particulier car notre capacité sonore est à 360°, contrairement à notre vision limitée à ce que nous voyons en face de nous. Ainsi, déclencher un élément sonore (sons des bulles, musique individualisée ou sons des manettes) sur le côté ou derrière l'enfant peut attirer son attention sur cette partie de l'espace et favoriser sa découverte. Ensuite, un second écart est lié au casque lui-même, qui permet de voir une partie du monde réel au travers d'une fenêtre d'augmentation avec un champ de vue limité (champ de vue horizontal de la caméra de 90°), entourée de noir. Ce décalage par rapport à la perception habituelle est positif pour un psychologue clinicien qui dit : « Ça me fait ressentir que l'on a quelque chose qui me sépare de vous ». Cet écart peut donc favoriser la compréhension de l'enfant qu'il interagira avec une machine. Un dernier écart perceptif correspond à la découverte d'un environnement nouveau. Dans le cas des praticiens, la surprise induite a favorisé le rire et l'exploration, comme en témoigne cette transcription verbale d'une vidéo d'une praticienne testant l'application : « Ouh ! [geste de recul et rire] ; Rien, ça m'a surprise [Interaction avec les deux mains, se tourne] ; Il y a plein de trucs ! ; Waah ! ».

La composition *REM* joue sur deux écarts perceptifs. Tout d'abord, l'écart créé en ajoutant des éléments virtuels à l'environnement réel connu vise à questionner nos représentations du réel, du virtuel, de l'imaginaire et des frontières entre ceux-ci. *REM* met ainsi l'accent sur notre perception dynamique du réel. C'est pour cela que la pièce s'appelle *Rapid Eye Movement (REM)* (mouvements oculaires rapides), qui correspond à une étape du sommeil où la majorité des rêves apparaissent. Le second écart perceptif correspond à l'opposition entre le déroulé linéaire musical prévu par le compositeur, et attendu par l'utilisateur, et la ►

découverte par l'utilisateur de sa possibilité d'agir sur l'environnement au travers d'interactions non mentionnées avant de débiter la performance. Cet écart vise à rendre l'utilisateur actif et à favoriser sa découverte lorsqu'il devient conscient de sa possibilité d'agir. La composition *REM* démarre une réflexion sur l'usage de ces écarts et de leur impact sur les représentations dans le cadre de production musicale spatialisée.

## 5. Discussion

Les résultats des quatre études présentent des dimensions distinctes du réalisme, de l'agentivité et des représentations de l'utilisateur, en lien avec l'immersion sonore perçue. Cette section discute ces trois aspects et suggère de futures perspectives de recherche.

### Perception du réalisme sonore en réalité augmentée

Dans les quatre études présentées, les participants ne mentionnent le concept de réalisme que de manière subjective et uniquement par rapport à certaines dimensions jugées pertinentes par rapport à leur perception de l'objectif de l'étude. Ces descriptions subjectives s'opposent donc à la vision technique du réalisme proposée par Olson (1972), liée à des caractéristiques telles que le spectre sonore ou la gamme dynamique. En particulier, dans *BMR* et *Not a Robot*, alors que le réalisme de localisation est limité, les utilisateurs s'immergent dans l'expérience au travers d'autres dimensions du réalisme. Dans *BMR*, ce réalisme concerne la fidélité de rendu sonore et l'externalisation sonore, en lien avec les besoins des musiciens. Dans *Not*

*a Robot*, le réalisme émerge de l'aspect théâtral, qui concerne le lien entre les sons et le lieu de l'expérience, l'expérience de groupe et la narration. De ces deux études émerge également l'idée qu'un défaut de réalisme sur un paramètre (ici la localisation) n'empêche pas la perception d'autres paramètres du réalisme. Dans *Magic Bubbles*, la majorité des participants ont été impliqués alors que les stimuli audiovisuels étaient simplifiés. Les praticiens ne mentionnent pas le concept de réalisme, car non pertinent par rapport au but de l'étude. De même, nous pouvons supposer que les praticiens mentionneraient le réalisme de l'intégration audiovisuelle uniquement si celui-ci était insuffisant et potentiellement gênant pour les enfants. Enfin, dans *REM*, une externalisation réaliste des sources sonores était visée, mais les commentaires des participants ne la mentionnent pas. Comme les participants n'étaient pas directement questionnés par rapport au réalisme perçu, ils se sont concentrés sur d'autres aspects en lien avec le sens de la pièce (liens entre sons et visuels, et agentivité).

Les paramètres du réalisme que les participants mentionnent sont donc en lien avec l'objectif des études et la possibilité pour l'utilisateur de s'impliquer dans celles-ci. Ils sont donc liés au sens perçu de l'expérience. Cette vision s'oppose à la vision moderniste du réalisme sonore proposée par Olson (1972) et à la vision de la « machine ultime » (Sutherland, 1965), selon lesquelles plus les capacités de rendu évolueront, plus l'utilisateur sera immergé. Au contraire, cette vision incite à prendre en compte le sens de l'expérience pour analyser les paramètres pertinents du réalisme à considérer pour concevoir un environnement immersif. De plus, une considération uniquement technique du réalisme peut nuire à l'immersion. Par exemple, dans le cas d'enfants autistes, un réalisme trop grand de résolution des

stimuli audiovisuels peut nuire à la compréhension générale et donc à l'immersion (Tarantino et al., 2019). Ce résultat incite à mener des études futures pour évaluer les différents paramètres du réalisme sonore. Ceci permettrait ensuite d'adapter la conception d'expériences de RA en agençant l'importance de ces différents paramètres en fonction du sens recherché.

### **Agentivité de l'utilisateur en réalité augmentée**

Dans les quatre études, l'agentivité dans un cadre structuré en tant qu'individu ou que groupe influence positivement l'immersion, en lien avec les deux facteurs contextuels influençant l'immersion selon Sungyoung (2021), à savoir, environnement d'écoute et contexte général. Cependant, en lien avec Shin (2019), nos quatre études mettent l'accent sur le fait que les utilisateurs sont des agents participant activement à la construction du phénomène d'immersion, contrairement à une immersion imposée par un dispositif. Ainsi, les utilisateurs (neurotypiques ou autistes) peuvent agir dans un système s'ils en comprennent le cadre, donc tous les éléments liés à l'expérience : dispositif technique, personnes présentes, objectif, narration, etc. Dans toutes les études, la compréhension du cadre et de la liberté d'action dans ce cadre donne une *impression de contrôle* nécessaire pour s'immerger activement dans la performance. Ceci se manifeste par exemple en prenant le contrôle sur leur propre son dans *BMR*, en respectant mais questionnant les instructions dans *Not a Robot*, en dessinant ou enregistrant des sons dans *Magic Bubbles*, ou en pouvant modifier le déroulement de la performance dans *REM*. Ce phénomène d'immersion comme construction active est aussi favorisé par la présence d'autrui. Par exemple, la conception de

*Magic Bubbles* prévoit que le praticien puisse favoriser l'exploration en suscitant de l'intérêt. Dans *BMR*, la connexion avec d'autres musiciens permet de renforcer l'immersion. Enfin, dans *Not a Robot*, l'aspect préféré des participants est l'expérience de groupe, qui favorise le fait qu'ils se sentent captivés.

La mise en parallèle des quatre études suggère l'importance de considérer l'influence des prises d'initiatives des utilisateurs sur l'immersion ressentie. Cette dimension s'ajoute à celles proposées par Sungyoung (2021), en mettant l'accent sur le rôle actif de l'utilisateur comme *co-créateur* de la performance. Dans nos études, ceci se manifeste de plusieurs manières : création musicale dans *BMR*, fait de jouer un rôle dans *Not a Robot*, jeu libre dans *Magic Bubbles* pour favoriser ces prises d'initiatives, et possibilités d'interagir dans *REM*. Cette considération de l'utilisateur comme acteur se rapproche du *Theatre of the Oppressed* de Boal (2002), où la narration est co-créée par spectateurs et acteurs afin de créer du sens et s'immerger ensemble.

### **Représentations des utilisateurs et réalité augmentée**

La conception des quatre environnements de RA s'est faite en lien avec les représentations des utilisateurs pour favoriser leur compréhension, et donc leur immersion. Dans les quatre études, la conception prévoit en effet des éléments en lien avec leurs attentes, à savoir : un dispositif technique proche du dispositif habituel utilisé (*BMR*), une narration et des interactions intuitives (*Not a Robot*), un dispositif en lien avec les pratiques cliniques courantes et une prise en compte de la perception autistique (*Magic Bubbles*), ou une situation d'écoute habituelle (*REM*). Lorsque la conception ne prend pas ►

assez en compte les attentes, ceci amène une confusion (*Magic Bubbles*, *REM*). Or, construire ce lien nécessite de considérer les préconceptions possibles des utilisateurs. Par exemple, Barrett (2021, p.176) évoque le fait qu'un casque audio peut nuire à une expérience sociale. De plus, les praticiens de la clinique de l'autisme ont souvent l'impression que le casque de XR peut nuire au lien à l'autre (Bauer et al., 2022). Dans nos quatre expériences, ces pré-conceptions ont été prises en compte, par exemple en utilisant des informations asymétriques dans *Not a Robot* pour mêler expérience de groupe et intime, ou en utilisant la RA et non la RV pour favoriser le lien à l'autre dans *Magic Bubbles*. Ainsi, une juste proportion d'éléments faisant lien avec les attentes du participant est nécessaire pour favoriser son implication. Ceci suggère de sonder de façon large la manière dont des individus de domaines différents perçoivent la RA et la production sonore spatialisée, pour saisir les pré-conceptions existantes, et les prendre en compte lors de la conception future d'expériences.

Plusieurs écarts perceptifs - considérés lors de la conception ou apparus lors des tests - ont favorisé l'exploration de l'environnement. Par conséquent, une juste proportion d'imprévus et d'éléments répondant aux attentes des utilisateurs permettrait de favoriser leur implication. En revanche, un écart trop grand par rapport à leurs attentes peut entraîner une charge cognitive, ce qui est notamment apparu dans *Not a Robot*. Ces écarts perceptifs ont par exemple amené à trouver de nouvelles idées musicales dans *BMR*, à plus interagir de façon non verbale dans *Not a Robot*, à questionner les concepts de réel et de virtuel dans *REM*, ou à favoriser la compréhension du fait d'interagir avec une machine dans *Magic Bubbles*. Le potentiel de la RA pour questionner notre rapport au réel était déjà évoqué par Shin (2019, p. 11) et, dans le cas de

performances sonore, par Barrett (2021, p.177). Son potentiel pour modifier notre rapport à soi, ici exploré dans *Not a Robot*, avait également été exploré de façon prometteuse par Isnard *et al.* (2021), qui modifiaient la voix de l'utilisateur en temps réel en RV.

Ces résultats suggèrent l'importance d'étudier les possibles écarts perceptifs permis par la RA (par ex : réel/virtuel, rapport au corps), les facteurs influençant la perception de ces écarts par l'utilisateur (par exemple, son état) et la possibilité d'ajuster ces écarts de façon dynamique en fonction de l'utilisateur. En particulier, l'écart perceptif en RA apparaît comme prometteur pour déconstruire les concepts de « réel » et de « virtuel ». En s'inspirant de l'expression de Deleuze (1980), l'écart perceptif en RA pourrait permettre un *bégalement* du réel, qui amènerait à considérer une multiplicité d'expériences entre réel et virtuel. Dans cette multiplicité, un même élément pourrait être considéré comme réel ou virtuel, ou avec une proportion de réel et de virtuel. Dans *Not a Robot*, les participants mentionnent d'ailleurs vouloir participer à plus d'expériences qui questionnent leur rapport au réel, avec l'ajout d'éléments tangibles connectés. En lien avec cet objectif, changer le terme de « média immersif » pour « média perceptif », comme proposé par Pike (2021), permettrait de mettre l'accent sur la perception de l'utilisateur, en lien avec ses caractéristiques, son activité, et son environnement.

-

Saisir ce qui se joue dans l'immersion en RA, et notamment l'immersion sonore, reste un défi. Deux définitions de l'immersion s'opposent souvent : une immersion qualifiée de façon technique, et permise par le système de RA utilisé, et l'immersion subjective

de l'utilisateur. Cet article questionne donc ce qui est en jeu dans l'immersion sonore en RA, au travers de trois aspects : le réalisme perçu, l'agentivité de l'utilisateur, et ses représentations. Pour analyser ces trois aspects, nous avons mis en parallèle les résultats issus de quatre études de RA qui représentent des domaines variés : retours casques pour musiciens en studio (Bauer *et al.*, 2018), expérience participative inspirée de pratiques théâtrales (Nagele *et al.*, 2021), expérience multisensorielle pour enfants autistes nécessitant un fort soutien (Bauer *et al.*, 2021), et production musicale spatialisée (Bauer et Bouchara, 2021). Tout d'abord, les résultats montrent que le réalisme perçu recouvre de nombreux paramètres, liés par exemple à la narration, à l'intégration audiovisuelle ou à l'expérience de groupe. Ce résultat appelle à mener de futures études pour définir les différents paramètres du réalisme sonore perçu par des participants en RA. Ensuite, trois dimensions de l'agentivité sont apparues en lien avec l'immersion : la possibilité d'interagir dans un cadre structuré, d'interagir avec d'autres, et de prendre des initiatives. Enfin, favoriser l'implication de l'utilisateur nécessite de concevoir un environnement de RA en lien avec ses besoins. Cependant, la présence d'écarts perceptifs, donc d'imprévus par rapport à ses attentes, peut permettre de questionner ses représentations et de favoriser son implication.

## Bibliographie

- > AGRAWAL, S., SIMON, A., BECH, S., BÆRENSEN, K., FORCHHAMMER, S. (2019), « Defining Immersion: Literature Review and Implications for Research on Immersive Audiovisual Experiences », 147th Audio Engineering Society Convention.
- > BANDURA, A. (2018), « Toward a Psychology of Human Agency: Pathways and Reflections », *Perspectives on Psychological Science*, n°13(2), pp. 130-136.
- > BAUER, V., BOUCHARA, T. (2021), « First Steps Towards Augmented Reality Interactive Electronic Music Production », 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 90-93.
- > BAUER, V., BOUCHARA, T., BOURDOT, P. (2021), « Designing an Extended Reality Application to Expand Clinic-Based Sensory Strategies for Autistic Children Requiring Substantial Support: Participation of Practitioners », IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 254-259.
- > BAUER, V., BOUCHARA, T., BOURDOT, P. (2022), « Extended Reality Guidelines for Supporting Autism Interventions Based on Stakeholders' Needs », *Journal of Autism and Developmental Disorders*.
- > BARRETT, N. (2021), « Spatial music composition ». J. PATERSON et H. LEE (dir.), *3D Audio*, New York, Routledge, pp. 175-191.
- > BAUER, V., DEJARDIN, H., PRAS, A. (2018), « Musicians' binaural headphone monitoring for studio », 144th Audio Engineering Society Convention.
- > BILLINGHURST, M., CLARK, A., LEE, G. (2015), « A Survey of Augmented Reality », *Foundations and Trends® in* ►

Human-Computer Interaction, n°8(2-3), pp. 73-272.

> BOAL, A (2002), *Games for actors and non-actors*, Hove, East Sussex, United Kingdom: Psychology Press, 301p.

> BOTTEMA-BEUTEL, K., KAPP, S. K., LESTER, J. N., SASSON, N. J., HAND, B. N. (2021), « Avoiding Ableist Language: Suggestions for Autism Researchers », *Autism in Adulthood*, n°3(1), pp. 18-29.

> BUCKLEY, Z., CARLSON, K. (2019), « Towards a Framework for Composition Design for Music-Led Virtual Reality Experiences », 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 1497-1499.

> CHARMAZ, K. (2006), *Constructing grounded theory: A practical guide through qualitative analysis*, London, SAGE Publication, 208p.

> CUMMINGS, J., BAIENSON, J. N. (2016), « How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence », *Media Psychology*, n°19(2), pp. 272-309.

> DELEUZE, G., PARNET, C. (1996), *Dialogues*, Paris, Flammarion (Champs).

> FRANCOMBE, J., WOODCOCK, J., HUGHES, R., MASON, R., FRANCK, A., PIKE, C., BROOKES, T., DAVIES, W., JACKSON, P., COX, T., FAZI, F., HILTON, A. (2018), « Qualitative Evaluation of Media Device Orchestration for Immersive Spatial Audio Reproduction », *Journal of the Audio Engineering Society*, n°66(6), pp. 414-429.

> GORZEL, M., ALLEN, A., KELLY, I., KAMMERL, J., GUNGORMUSLER, A., YEH, H. (2019), « Efficient Encoding and Decoding of Binaural Sound with Resonance Audio », *Audio Engineering Society Convention on Immersive and Interactive Audio*.

> INGOLD, T. (2002), *The Perception of the Environment*, London, Routledge, 480p.

> ISNARD, V., NGUYEN, T., VIAUD-DELMON, I. (2021), « Exploiting Voice Transformation to Promote Interaction in Virtual Environments », 2021 IEEE

Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 75-79.

> LANIER, J., HEILBRUN, A., SACKS, B. (1989), « An Interview with Jaron Lanier », *Whole Earth Review* (Fall Ed.), pp. 109-119.

> MILGRAM, P., & KISHINO, F. (1994), « A Taxonomy of Mixed Reality Visual Displays », *IEICE Transactions on Information and Systems*, n°77, pp. 1321-1329.

> NAGELE, A. N., BAUER, V., HEALEY, P. G. T., REISS, J. D., COOKE, H., COWLISHAW, T., BAUME, C., PIKE, C. (2021), « Interactive Audio Augmented Reality in Participatory Performance », *Frontiers in Virtual Reality*, n°1.

> OLSON, H. F. (1972), *Modern sound reproduction*, New York, Van Nostrand Reinhold.

> PIKE, C. (2021), « 3D audio in broadcasting », J. PATERSON et H. LEE (dir.), *3D Audio*, New York, Routledge, pp. 1-18.

> ROBERTSON, C. E., BARON-COHEN, S. (2017), « Sensory perception in autism », *Nature Reviews Neuroscience*, n°18(11), pp. 671-684.

> SHIN, D. (2019), « How does immersion work in augmented reality games? A user-centric view of immersion and engagement », *Information, Communication & Society*, n°22(9), pp. 1212-1229.

> SINHA, P., KJELGAARD, M. M., GANDHI, T. K., TSOURIDES, K., CARDINAUX, A. L., PANTAZIS, D., DIAMOND, S. P., HELD, R. M. (2014), « Autism as a disorder of prediction », *Proceedings of the National Academy of Sciences*, n°111(42), pp. 15220-15225.

> SLATER, M., WILBUR, S. (1997), « A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments », *Presence: Teleoperators and Virtual Environments*, n°6(6), pp. 603-616.

> SUNGYOUNG, K. (2021), « Contextual factors in judging auditory immersion », J. PATERSON et H. LEE (dir.), *3D*

Audio, New York, Routledge, pp. 160-174.

> SUTHERLAND, I. E. (1965), « The Ultimate Display », IFIP Congress, pp. 506-508.

> SUTHERLAND, I. E. (1968), « A head-mounted three dimensional display », Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I on - AFIPS '68 (Fall, Part I).

> TARANTINO, L., De GASPERIS, G., MASCIU, T. D., PINO, M. C. (2019), « Immersive applications: What if users

are in the autism spectrum? », The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry - VRCAI '19, pp. 1-7.

> TSINGOS, N. (2017), « Object-Based Audio », A. ROGINSKA et P. GELUSO (dir.), Immersive sound: the art and science of binaural and multi-channel audio, New York, Routledge, pp. 258-289.

> WILSON, M. (2002), « Six views of embodied cognition », Psychonomic Bulletin & Review, n°9(4), pp. 625-636.

## VALENTIN BAUER

Valentin Bauer est doctorant à l'Université Paris-Saclay, CNRS, LISN, où il prépare une thèse autour de l'audio spatialisé, la Réalité Augmentée (RA) et l'autisme, supervisée par Patrick Bourdot et Tifanie Bouchara. En 2017, il sort diplômé de la Formation Supérieure aux Métiers du Son au Conservatoire de Paris (CNSMDP). Il y mène un mémoire de recherche sur les retours casque en binaural pour musiciens, en lien avec la cellule « Qualité et Innovation » de Radio France, et supervisé par Hervé Déjardin et Amandine Pras. En 2018, il travaille comme conseiller musical avec l'Orchestre National du Capitole de Toulouse. La même année, il débute un second Master, au Media and Arts Technology à Queen Mary University of London, où il réalise son stage avec la BBC R&D, supervisé par Chris Baume et Joshua Reiss, sur une performance participative en Réalité Augmentée (RA) audio. Fin 2019, il débute son doctorat centré sur les troubles sonores dans l'autisme et engage une collaboration avec l'hôpital de jour André Bouulloche, par une étude de terrain sur l'utilisation de la RA comme outil de médiation, ainsi que trois autres collaborations, avec un psychomotricien, une musicothérapeute, et le Multisensory Experience Lab à Aalborg University Copenhagen dirigé par Stefania Serafin. En parallèle, il mène une recherche sur la création sonore spatialisée en RA et assure des missions d'enseignement, au CNSMDP et à l'IUT informatique d'Orsay. Enfin, il mène une activité de création musicale, en particulier pour France Culture (2017) et la Cité des Sciences (2022).

*Valentin Bauer is a PhD candidate at Paris-Saclay University, CNRS, LISN, working around spatialized audio, Augmented Reality (AR), and autism, supervised by Patrick Bourdot et Tifanie Bouchara. In 2017, he graduates from the Advanced Music Production Program (FSMS) at the Paris Conservatoire (CNSMDP). For his Master's thesis, he investigates binaural headphone monitoring for studio recording sessions at Radio France "Quality and Innovation" Department, supervised by Hervé Déjardin and Amandine Pras. In 2018, he works as a music consultant for Toulouse Capitole National Orchestra. He then starts the Media and Arts Technology Program at Queen Mary University of London. For his second Master's thesis, he focuses on a participative Audio AR performance at BBC R&D, supervised by Chris Baume and Joshua Reiss. In 2019, he starts his PhD, focusing on the auditory perception of autistic individuals. To that end, he begins a collaboration with the day hospital André Bouulloche, with a field study using an AR sensory-based and mediation approach. He also starts three other collaborations, with a psychomotor therapist, a music therapist, and the Multisensory Experience Lab at Aalborg University Copenhagen. In parallel, he conducts a research project about AR spatialized music production, and teaches classes about research at the CNSMDP and about computer science at the IUT Orsay. Valentin Bauer has also been working on several music creation projects as a composer, for France Culture (2017) and the Cité des Sciences (2022).*

## Composer (avec) l'espace : une pratique musicale polysémique

| François-Xavier Féron

### Résumé

Au fil des siècles, l'espace est devenu pour beaucoup d'artistes une dimension à part entière occupant une place centrale au sein du processus créatif. Si la notion d'espace en musique peut souvent renvoyer à des concepts abstraits, elle est abordée dans cet article en tant que dimension sonore travaillée par les compositeurs et perçue par les auditeurs. Dans une perspective historique, nous reviendrons brièvement sur l'émergence du paradigme spatial dans les musiques de tradition « savante » depuis la Renaissance et nous nous pencherons sur quelques écrits théoriques de la seconde moitié du XXe siècle tentant de mieux circonscrire la notion d'espace. Nous présenterons ensuite, à travers des exemples musicaux puisés dans les répertoires des XXe et XXIe siècles, cinq manières d'approcher et de composer (avec) l'espace : la répartition spatiale des sources sonores, l'exploration ainsi que la simulation des propriétés acoustiques d'un lieu, la création de perspectives sonores et enfin, la prise en compte du rayonnement des sources sonores.

### Abstract

*Over the centuries, space has become for many artists an integral dimension, becoming a core element in the creative process. If the notion of space in music can often refer to abstract concepts, it is approached in this article as a sound dimension worked on by composers and perceived by listeners. From a historical perspective, we will briefly revisit the emergence of the spatial paradigm in classical music tradition since the Renaissance and we will look upon some theoretical writings from the second half of the 20th century in an effort to better define the notion of space. We will then propose, through musical examples drawn from the repertoires of the 20th and 21st centuries, five ways of approaching and composing (with) space: the spatial distribution of sound sources, the exploration as well as the simulation of the acoustical properties of a place, the design of sound perspectives and finally, taking into account the radiation of sound sources.*

« L'espace est une dimension immanente de la musique et il faut tenter de distinguer ce qui relève d'une simple conséquence des conditions acoustiques, de la véritable intégration de l'espace en tant que paramètre structurant. On trouve facilement des descriptifs de dispositifs et d'installations techniques, mais beaucoup plus rarement des documents permettant le croisement analytique entre la partition instrumentale et le vocabulaire spatial. Ce déficit de représentations ne signale pas systématiquement une faible conceptualisation, mais laisse tout de même planer un doute quant à la véritable prise en compte des fonctions de l'espace, de l'ensemble des niveaux perceptifs qui le constitue et de l'efficacité de leurs actions dans l'écriture musicale. » (Bioteau, 2013, p. 1337)

Les syntagmes associant l'idée de son à celle d'espace sont nombreux et ils se réfèrent à différents champs de pratique et de recherche spécifiques : « ceux de l'écologie sonore, de l'acoustique architecturale, de la sociologie urbaine, de la création musicale ou de la linguistique » (Pinard, 2014). En musique, l'espace est devenu au cours du XXe siècle une préoccupation majeure pour de nombreux artistes, au point d'être considéré comme une cinquième dimension sonore – au côté de la hauteur, de la durée, de l'intensité et du timbre. La notion d'espace demeure néanmoins relativement absconse, en raison de sa nature polysémique. Elle renvoie aussi bien à des concepts abstraits qu'à des procédés d'écriture et des dispositifs technologiques contribuant à la formation d'impressions spatiales. Si les ouvrages collectifs consacrés à la notion d'espace en musique se sont essaimés depuis les années 1990

(e.g. *Lien – L'espace du son I, II et III*, 1988, 1991 et 2011 ; *Les cahiers de l'IRCAM*, 1994 ; Genevois et Orlarey, 1998 ; Chouvel et Solomos, 1998 ; Born, 2013 ; Brech et Paland, 2015), une certaine ambiguïté sémantique tend à persister. Composer l'espace... Composer avec l'espace : qu'est-ce que cela signifie pour un artiste et comment cela se traduit-il à l'écoute pour le public ?

Cet article vise à révéler la nature protéiforme de la notion d'espace en montrant comment cette dimension musicale peut être approchée de multiples manières pour donner lieu à différentes formes d'expériences sonores. Il n'est nullement question ici d'aborder l'espace sous un quelconque angle métaphorique, symbolique, philosophique ou sociologique, mais de révéler sa tangibilité à travers l'écoute. C'est pour cette raison que nous avons choisi d'inclure une discographie, comprenant par ailleurs plusieurs enregistrements en 5.1, format permettant de mieux restituer la dimension spatiale de certaines œuvres. Nous reviendrons tout d'abord sur l'émergence au fil des siècles du paradigme spatial dans les musiques dites de tradition « savante » et nous nous pencherons, dans ce contexte, sur quelques écrits théoriques de la seconde moitié du XXe siècle tentant de mieux définir la notion d'espace. Nous verrons alors comment ce paradigme devient un paramètre structurant dans de nombreuses œuvres produites au cours des XXe et XXIe siècles. Pour cela nous nous concentrerons sur cinq types d'approches spatiales : l'éclatement des sources sonores dans l'espace (section 2), l'exploration des propriétés acoustiques d'un espace (section 3), la simulation et la restitution des caractéristiques acoustiques d'un espace (section 4), la création de perspectives sonores (section 5) et enfin, l'exploration du rayonnement spatial des sources sonores (section 6). ►

## 1. L'émergence du paradigme spatial dans les musiques de tradition « savante »

### Quelques jalons historiques

L'espace – en tant que lieu doté de qualités acoustiques et topographiques – a, depuis toujours, eu un impact sur les pratiques artistiques. L'emplacement et le contenu des représentations rupestres seraient ainsi liés aux spécificités acoustiques des cavités à l'intérieur d'une grotte. L'évolution des langages et genres musicaux en occident serait étroitement connectée à l'architecture des lieux dans lesquels la musique était exécutée (Forsyth, 1987). Durant la période classique par exemple, Joseph Haydn (1732-1809) et Ludwig van Beethoven (1770-1827) tirent profit du potentiel des grandes salles de concert pour sublimer les techniques d'écriture orchestrale. Par exemple, Haydn ajuste dans ses symphonies l'instrumentation en fonction des salles où elles devaient être créées, mais ce n'est ici pas tant pour générer une expérience d'écoute spatiale, que pour parfaire l'équilibre instrumental. En revanche, dans le cas des *cori spezzati* (chœurs brisés), une technique que les compositeurs vénitiens ont développée durant la Renaissance au sein de la basilique San Marco, l'espace est non seulement structurant, mais il devient aussi parfaitement tangible à l'écoute des œuvres. S'inspirant des modèles byzantins, cet édifice majestueux en forme de croix est doté de cinq coupoles et de nombreuses tribunes dans lesquelles des groupes de musiciens peuvent être disséminés. La distanciation entre ces groupes permet de rompre avec une écoute frontale resserrée et contribue alors au développement des techniques antiphonales (alternance entre deux chœurs) et polychorales. Les *cori spezzati* marquent ainsi les débuts de la spatialisation sonore qui consiste à explorer

musicalement la perception du positionnement des sources dans l'espace.

« Cette technique n'a pas été inventée par Adrien Willaert, *maestro di capella* à San Marco à Venise de 1527 à 1567, mais elle est devenue davantage connue grâce à ses performances. Ce type de musique, qui à ses débuts était interprétée par deux chœurs de petite taille, s'est musicalement complexifié au XVII<sup>e</sup> siècle, avec la répartition aux différents balcons de plusieurs ensembles dont le son splendide et grandiose envahissait les églises. »<sup>1</sup>

Adrien Willaert (~1490-1562), suivi de Andrea Gabrieli (1533-1586) et de Giovanni Gabrieli (1555-1612), développent cet art des *cori spezzati* en composant des recueils entiers de pièces spatialisées (SACD Cori Spezzati, 2006). Cette pratique se répand progressivement hors de Venise. Alessandro Striggio (~ 1540-1592) compose le motet *Ecce beatam lucem* (1561) pour 40 voix indépendantes (SACD Alessandro Striggio, 2012), œuvre qui aurait incité Thomas Tallis (1505-1585) à écrire le célèbre *Spem in alium* (1573), un motet pour 40 voix indépendantes réparties au sein de huit chœurs positionnés autour du public (SACD Thomas Tallis, 2003). Cette œuvre iconique est d'ailleurs au cœur d'une installation sonore de Janett Cardiff (1957-) : dans *The Forty Part Motet* (2001), l'artiste canadienne recourt à 40 haut-parleurs pour reproduire individuellement chacune des voix ; les spectateurs sont invités à se déplacer à l'intérieur de l'installation pour vivre, au centre, une expérience immersive totale résultant de la fusion des voix, ou bien pour expérimenter, en périphérie, le point de vue des interprètes en se rapprochant de tel ou tel haut-parleur.

1

« The technique was not invented by Adrien Willaert, *maestro di capella* at San Marco in Venice from 1527–1567, but became well known due to his performances. This music, which at the beginning was sung by two small choirs, developed into a complex musical event by the seventeenth century, with several ensembles placed in different balconies, filling churches with impressive, splendid sound. » (Baumaan, 2015, p. 57, notre trad.)

« J'ai positionné les haut-parleurs en ellipse dans la salle afin que l'auditeur puisse vraiment ressentir la construction sculpturale de la pièce de Tallis. Vous pouvez entendre le son se déplacer d'un chœur à l'autre, passant de l'avant à l'arrière, se répondant en écho ici et là, puis vous sentir submergé par les ondes sonores lorsque tout le monde chante. »<sup>2</sup>

Au début du XVIIe siècle, l'émergence de l'opéra conduit les compositeurs à explorer la mise en espace des interprètes – en les positionnant notamment en coulisse – afin de renforcer les effets de mise en scène. Sinon, durant les périodes baroque et classique, la spatialisation sonore demeure, dans le répertoire instrumental, une pratique assez rare même s'il existe quelques exceptions notables, comme la *Serenada Notturmo KV 239* pour 2 ensembles et le *Notturmo KV 286* pour 4 ensembles, dans lesquelles Wolfgang Amadeus Mozart (1756-1791) reproduit des effets d'écho (SACD Wolfgang Amadeus Mozart, 2006). En revanche, au cours de la seconde moitié du XIX siècle – durant la période romantique –, l'espace suscite à nouveau un fort intérêt.

Hector Berlioz (1803-1869), Richard Wagner (1813-1883) et Gustav Mahler (1860-1911) développent une véritable pensée de l'espace qui se traduit de différentes façons. Les procédés de spatialisation sont à nouveau explorés comme en témoigne, par exemple, l'utilisation du cor postal en coulisse dans la troisième symphonie (1895-1896) de Mahler (SACD Gustav Mahler, 2003), ou la disposition des quatre groupes de cuivres autour du public dans la *Grande messe des morts* (1837) de Berlioz (SACD Hector Berlioz, 2013). Mais il s'agit aussi d'explorer l'effet d'espace créé par le timbre des instruments et de susciter des impressions spatiales en opposant des

sonorités perçues comme lointaines et d'autres, plus proches. Les indications textuelles sur les partitions font d'ailleurs de plus en plus appel à un champ lexical se rapportant à l'espace. Le rayonnement instrumental, c'est-à-dire la manière dont les instruments projettent le son dans l'espace, est aussi pris en considération, notamment par Wagner qui repense le positionnement des musiciens au sein de l'orchestre en fonction de la puissance des instruments, mais aussi de leur rayonnement. Dans sa quête d'œuvre d'art total, le compositeur allemand fait construire une salle de concert dédiée à sa musique – le *Festspielhaus* de Bayreuth –, une salle qui s'inspire des théâtres grecs antiques et qui présente de nombreuses innovations architecturales répondant à des préoccupations d'ordre visuel comme acoustique. À l'aube du XXe siècle, l'espace apparaît donc déjà comme une préoccupation centrale et cela, de maintes façons : répartition des sources sonores dans un lieu, prise en compte du rayonnement instrumental, structuration du contenu musical pour engendrer des impressions spatiales, construction de nouvelles salles de concert...

### Vers une formalisation de la notion d'espace au XXe siècle

Au XXe siècle, les procédés de spatialisation se complexifient et se diversifient de façon remarquable permettant, entre autres, la confection de trajectoires sonores dans l'espace. Mais une quête de tridimensionnalité se fait aussi de plus en plus ressentir dans la manière d'agencer les matériaux au sein d'une œuvre. Cette émancipation de la pensée spatiale est étroitement liée au développement des technologies sonores – lutherie électronique, matériel d'enregistrement et de diffusion, synthèse ►

2

« I placed the speakers around the room in an oval so that the listener would be able to really feel the sculptural construction of the piece by Tallis. You can hear the sound move from one choir to another, jumping back and forth, echoing each other and then experience the overwhelming feeling as the sound waves hit you when all of the singers are singing. » (<https://cardiffmiller.com/installations/the-forty-part-motet/>, notre trad.)

sonore... –, ce qui a ouvert la voie à de nouvelles pratiques compositionnelles et donné naissance, après la Seconde Guerre mondiale, à de nouveaux genres musicaux – musique concrète, musique électronique, dub, *computer music*, musiques mixtes... L'utilisation des haut-parleurs dans les arts électroacoustiques a bouleversé de manière drastique la manière de travailler (avec) l'espace. Celui-ci devient alors un vaste terrain d'expérimentations dans de nombreux domaines – prise de son, sonorisation, travail en studio, composition musicale, installation sonore... –, concourant au développement d'un champ lexical (Merlier, 2006) et insufflant de nouvelles pratiques dans le domaine de la musique instrumentale.

À défaut de revenir en détail sur toutes ces innovations, il nous semble important de nommer quelques personnalités musicales ayant, à nos yeux, fortement contribué à l'émancipation de la pensée spatiale au cours du XXe siècle : Charles Ives (1874-1954), Edgard Varèse (1883-1965), Pierre Schaeffer (1910-1995), John Cage (1912-1992), Henry Brant (1913-2008), Iannis Xenakis (1922-2001), György Ligeti (1923-2006), Luigi Nono (1924-1990), Luciano Berio (1925-2003), Pierre Boulez (1925-2016), Pierre Henry (1927-2017), Karlheinz Stockhausen (1928-2007), Alvin Lucier (1931-2021), François Bayle (1932-), Pauline Oliveros (1932-2016), John Chowning (1934-), Roger Reynolds (1934-), Léo Küpper (1935-), Francis Bayer (1938-2004), Jonathan Harvey (1939-2012), Emmanuel Nunes (1941-2012), Annette Vande Gorne (1946-), Michel Chion (1947-), Philippe Manoury (1952-), Denis Dufour (1953-), Bertrand Merlier (1958-), Marco Stroppa (1959-), Marc-André Dalbavie (1961-)... Les compositrices et compositeurs figurant dans cette liste – qui n'est bien évidemment pas exhaustive –, ont intégré l'espace de manière structurelle dans plusieurs de leurs œuvres et ont, pour la plupart, décrit leurs

approches à travers des écrits théoriques. Dans l'article dont est tirée l'épigraphe de notre texte, le compositeur et musicologue Alain Bioteau (1962-2017) réalise une synthèse de plusieurs de ces écrits – notamment ceux de Stockhausen, Nunes, Brant, Chowning, Vande Gorne et Küpper – et rend compte de la variété des approches, comme de la permanence de certains traits. Avant de révéler cette diversité, nous aimerions revenir brièvement sur d'autres écrits théoriques, ceux de Bayer, Chion et Stroppa. Tous trois ont cherché à formaliser au mieux la notion d'espace en identifiant et définissant certaines approches compositionnelles auxquelles nous nous référerons par la suite.

Dans son ouvrage dédié à la notion d'espace sonore dans les musiques avant-gardistes instrumentales de tradition écrite, Bayer opère une distinction entre l'« espace acoustique extrinsèque » qui fait référence « à la spatialisation physique de l'œuvre musicale dans le lieu de son exécution et de son audition » (Bayer, 1981, p. 13) et l'« espace sonore » qui est porteur d'une « spatialisation intrinsèque, entièrement définie par la situation respective des sons les uns par rapport aux autres, tels qu'ils sont notés dans une partition musicale écrite » (Bayer, 1981, p. 13). Si l'ouvrage se concentre sur ce dernier concept, l'auteur souligne dans son introduction que le déploiement des sources sonores dans un espace acoustique extrinsèque est propice à la formation de trois principaux types d'effets : effets de relief, d'ubiquité et de mobilité. L'effet de relief se traduit par une impression de profondeur « qui provient de l'évaluation acoustique de la distance séparant l'origine supposée du son du sujet récepteur et qui semble nous révéler concrètement d'une façon particulièrement sensible l'existence d'une nouvelle dimension spatiale. » (Bayer, 1981, p. 13). L'effet d'ubiquité sonore prend forme lorsque les signaux acoustiques semblent provenir

de partout. Il intervient lorsque « les différentes sources sonores que le compositeur a pris soin de disperser au préalable dans l'espace sont entendues simultanément. » (Bayer, 1981, p. 14). Enfin, l'effet de mobilité « est directement lié à l'idée d'une circulation des sons d'un groupe instrumental (ou d'une source sonore) à l'autre » (Bayer, 1981, p. 14).

Dans son court essai intitulé *Les deux espaces de la musique concrète*, Chion opère, dans le domaine des musiques sur support, une distinction assez semblable à celle de Bayer, en soulignant la coexistence d'un « espace interne » fixé sur le support d'enregistrement, et d'un « espace externe » lié aux conditions d'écoute dans un lieu particulier. L'espace interne fait indissolublement partie de l'œuvre : il « est aussi constitutif des sons que le sont leur hauteur, leur timbre ou leur grain... » (Chion, 1988, p. 31). L'interprétation dite acousmatique – consistant à projeter au moment du concert une œuvre sur support à travers un orchestre de haut-parleurs – tend à « complexifier, dispatcher, redoubler, amplifier, contrarier, démultiplier, recouvrir, etc... les effets spatiaux internes à l'œuvre par une mise en espace externe » (Chion, 1988, p. 32).

Pour Stroppa, l'espace est une dimension essentielle qu'il explore intensément depuis ses premières œuvres produites au début des années 1980. Il a consacré d'ailleurs plusieurs textes à la question de l'espace (Stroppa, 1992 ; 2007), dont un resté inédit (Stroppa, 1991) dans lequel il met de côté les paramètres caractérisant l'acoustique des salles, pour décrire trois composantes primaires : « la localisation, qui est la perception du placement d'une source par rapport à un auditeur », « la profondeur, qui est la perception d'une distance, c'est-à-dire de l'éloignement d'une même source sonore par rapport à l'auditeur » et enfin, « l'image, qui est liée

à la nature de la source, à la façon qu'elle a de diffuser le son autour d'elle. » (Stroppa, 1991, p. 1)

Croiser tous ces points de vue permet d'avoir une vision assez large sur ce que signifie travailler (avec) l'espace. Dans la suite de cet article, nous parlerons d'*espace physique* pour désigner l'environnement matériel dans lequel sont produits des sons. Il ne s'agit ni plus ni moins de l'espace acoustique extrinsèque pour Bayer, ou bien encore de l'espace hôte pour Reynolds (Reynolds, 1978). Un espace physique est doté de caractéristiques acoustiques propres : une salle de concert, un studio d'enregistrement, une église, un salon, un amphithéâtre, un lieu en plein air..., sonnent différemment en raison de l'importance plus ou moins accrue des phénomènes de réflexion, transmission, absorption et diffusion. L'espace physique interfère ainsi avec les sons produits en les filtrant (résonances), en les prolongeant (réverbération) ou même, en les démultipliant (échos). Il est immanent à toute expérience sonore, à moins que l'auditeur ne se retranche dans une chambre anéchoïque – lieu en théorie totalement dépourvu de réflexions –, ou bien qu'il ne recoure à un casque audio fermé, ce qui lui permettrait alors de s'affranchir des caractéristiques acoustiques de l'environnement dans lequel il se situe. Mais il ne faut pas oublier – si on écarte le cas des musiques de synthèse pure – que le support audio porte en lui l'empreinte d'un autre espace physique : celui dans lequel a été réalisé l'enregistrement. C'est ainsi que les choses se compliquent car cet espace physique n'existe alors plus qu'à travers le son : il s'est muté en espace interne. Quoi qu'il en soit, l'espace physique apparaît dans de nombreuses œuvres comme un élément structurel ou, du moins, structurant. ►

Composer avec l'espace physique consiste avant tout à explorer la topographie d'un lieu pour y déployer un peu partout des sources sonores afin de spatialiser le son, approche dont nous expliquons certains des enjeux dans la section suivante. Mais composer avec l'espace physique consiste aussi à explorer les spécificités acoustiques d'un lieu, à révéler ses singularités pour en faire un partenaire musical, comme cela sera souligné dans la section 3. Ces caractéristiques peuvent aussi être restituées plus ou moins fidèlement suivant les dispositifs d'enregistrement et de diffusion mis en jeu, mais elles peuvent aussi être simulées instrumentalement à travers l'écriture ; l'auditeur est alors projeté dans un autre lieu, un espace artificiel comme nous le verrons dans la section 4. Afin d'illustrer la notion d'espace sonore, chère à Bayer, nous nous tournerons, dans la section 5, vers Varèse et Ligeti, deux compositeurs qui, plus que tout autre, sont parvenus à produire des sensations spatiales robustes sans avoir à disposer les musiciens d'une manière spécifique dans l'espace physique. Enfin, dans la section 6, nous verrons comment le rayonnement des sources sonores – instruments comme haut-parleurs – peut être le siège de nouvelles expériences spatiales. Nous nous intéresserons plus particulièrement dans ce contexte aux recherches pionnières de Stroppa qui, depuis plusieurs années maintenant, compose des figures de rayonnement en recourant à des « totems acoustiques » constitués de plusieurs haut-parleurs empilés les uns sur les autres.

## 2. Répartir les sources sonores dans l'espace physique pour composer (avec) la spatialisation

Le terme spatialisation est employé ici pour décrire l'ensemble des pratiques ayant trait à la pluralité de localisations des sources sonores et à leur éventuelle

mobilité. Si les *cori spezzati* ont ouvert la voie de la spatialisation sonore, c'est avec l'avènement de la stéréophonie – dont les prémices remontent au théâtrophone conçu par Clément Ader (1841-1925) à la fin du XIXe siècle – que cette pratique s'est émancipée. Afin de révéler la diversité des techniques de spatialisation et de leurs enjeux, revenons sur les investigations spatiales de quelques pionniers incontournables : tout d'abord Charles Ives et Henry Brant qui recourent à l'éclatement des musiciens pour clarifier le contenu musical ; Karlheinz Stockhausen et Iannis Xenakis qui développent une véritable dialectique entre sources statiques et sources mobiles ; John Chowning dont l'algorithme de spatialisation permet de simuler toutes sortes de trajectoires sonores ; Emmanuel Nunes qui, dans *Quodlibet*, explore de manière originale la topographie d'une salle à Lisbonne.

### Charles Ives et Henry Brant

Chez Ives et Brant, l'espace physique sert d'agent clarificateur d'une complexité musicale. Dans ses deux contemplations pour orchestre – *Central Park in the Dark* (1906) et *The Unanswered Question* (1908) (SACD Charles Ives, 2016) –, Ives superpose des matériaux sonores hétéroclites qu'il assigne à différents groupes de musiciens éloignés les uns des autres, de sorte qu'ils se mélangent le moins possible. Marqué par la découverte en concert de *The Unanswered Question*, Brant s'engouffre dans cette voie : depuis *Antiphony I* (1953) pour 5 groupes instrumentaux, il ne composera plus que des œuvres spatiales souvent dédiées à des lieux spécifiques (CD The Henry Brant Collection, 2005-2008). Il rédige d'ailleurs l'année suivante un court essai dans lequel il tente de théoriser sa démarche, essai qui sera approfondi plus tard dans un texte où il explique comment l'espace est devenu un aspect essentiel de sa musique (Brant, 1998).

« Dans ses œuvres, Brant sépare les strates sonores produites simultanément en recourant à la distanciation spatiale tout en jouant conjointement sur la dissonance, les différences de tempo et de métrique, les motifs mélodiques et l'instrumentation. Cet effet de juxtaposition issu de la simultanéité de plusieurs types de musique était recherché avant Brant, notamment par Berlioz, Mahler et Ives. Les écrits et les œuvres de Brant regorgent d'exemples montrant comment il utilise conjointement des indices spatiaux avec d'autres indices pour favoriser la ségrégation des flux auditifs. »<sup>3</sup>

### Karlheinz Stockhausen et Iannis Xenakis

Avec la naissance de la musique concrète à Paris en 1948 et celle de la musique électronique à Cologne, deux ans plus tard, la localisation des sons dans l'espace suscite de nouvelles formes d'expérimentations sonores. Stockhausen explore les possibilités de la quadriphonie dans *Gesang der Jünglinge* (1955-56) pour support 4 pistes (CD Karlheinz Stockhausen, 1991) : cette œuvre marque un tournant historique dans l'utilisation de la spatialisation puisque, pour la première fois, la position et le mouvement des sons dans l'espace sont des paramètres aussi importants que la hauteur, le timbre, l'intensité et la durée. Au même moment, le compositeur allemand travaille à l'écriture de *Gruppen* (1955-1957) pour 109 musiciens répartis à travers trois orchestres – chacun dirigé par un chef – disposés en fer à cheval dans la salle de concert. Contrepoint spatial, effets d'ubiquité, rotations sonores..., cette œuvre pose les bases d'une nouvelle grammaire de l'espace (CD Karlheinz Stockhausen, 1992). En 1958, Stockhausen expose ses recherches lors d'une conférence donnée à Darmstadt : intitulée « *Musik im Raum* » [Musique dans l'espace], elle sera publiée et traduite par la suite

dans plusieurs langues (Stockhausen, 1988). Viendront ensuite *Carré* (1958-1960) pour 4 orchestres et chœurs mixtes dirigés par 4 chefs, puis *Kontakte* (1959-60) pour support 4 pistes, pièce qui existe aussi dans une version mixte avec l'ajout d'une partie de piano et de percussion. Dans ces deux œuvres, le public est placé à l'intérieur d'un dispositif instrumental et/ou technologique de type quadriphonique permettant de renforcer le sentiment d'immersion sonore. Pour *Kontakte* (CD Karlheinz Stockhausen, 1991), Stockhausen met au point un système ingénieux permettant de contrôler la vitesse de rotation des sons : un haut-parleur est placé au centre d'une table rotative tandis que quatre microphones, situés autour de celle-ci, captent le signal émis. L'enregistrement 4 pistes ainsi obtenu permet de restituer les rotations sonores en disposant les haut-parleurs aux quatre coins de la salle de concert. Ce ne sont là que les prémices d'une aventure spatiale que le compositeur poursuivra jusqu'à ces derniers jours.

Architecte de formation, Xenakis travaille un temps auprès de Le Corbusier (1887-1965) avec qui il conçoit notamment le pavillon Philips à l'occasion de l'exposition universelle de Bruxelles en 1958. À l'intérieur de cet édifice éphémère dont les parois épousent des formes paraboliques et hyperboliques, sont disposés plus de 400 haut-parleurs à travers lesquels sont spatialisés *Concret PH* (1958) de Xenakis et le célèbre *Poème électronique* (1957-58) de Varèse. Au cours des années 1960, Xenakis se penche sur la question de la spatialisation dynamique du son instrumental. Dans *Eonta* (1964) pour piano et 5 cuivres, les musiciens – hormis le pianiste – occupent au fil de l'œuvre différentes positions sur scène, devant même à certains moments déambuler à l'intérieur d'une zone de promenade (CD Iannis Xenakis, 1988). Puis, dans *Terretektorh* (1966) pour 88 musiciens disséminés au milieu du public, il confectionne toutes sortes de trajectoires sonores, notamment des rotations dont il modifie sans cesse la direction et la vitesse (CD Iannis

3

« In his works Brant uses spatial distance in conjunction with dissonance, differences in tempo and metre, melodic patterns and instrumentation, in order to separate the music into simultaneous layers. The effect of juxtaposing several different types of music played simultaneously was used before Brant, by, for instance, Berlioz, Mahler and Ives. Brant's writings and pieces contain many examples of his usage of spatial cues working in conjunction with other cues to create stream segregation. » (Harley, 1999, p. 150, notre trad.)

Xenakis, 2003) : dans le cas du mouvement circulaire, explique-t-il, « on peut établir une progression uniforme qui ressemble au mouvement de l'aiguille des secondes sur une horloge : même durée temporelle, même distance. On peut aussi créer une accélération ou une décélération de ce mouvement par diverses fonctions mathématiques » (Harley, 1994, p. 15). Les musiciens sont par ailleurs munis de sifflets à sirène et d'instruments percussifs qu'ils doivent jouer, à certains moments, tous ensemble, produisant des effets d'ubiquité remarquables : la matière sonore semble littéralement explosée dans l'espace en une multitude de fragments ou bien provenir de partout en même temps. Tout comme Stockhausen, Xenakis explorera tout au long de sa vie la question de l'espace à travers sa musique.

### John Chowning

Pionnier de la synthèse sonore par ordinateur, le compositeur et chercheur John Chowning est profondément marqué par l'expérience de *Kontakte* en concert, œuvre qu'il a interprétée d'ailleurs en tant que percussionniste en 1967. Au sein de l'université de Stanford, il met au point à cette époque, non seulement la technique révolutionnaire de synthèse par modulation de fréquence, mais aussi un algorithme de spatialisation lui permettant de façonner des trajectoires sonores pouvant sortir du cadre délimité par les quatre haut-parleurs de son dispositif quadraphonique.

« En 1967, j'avais terminé le système de traitement spatial qui me permettait de positionner le son en des points de l'espace situés à l'extérieur du périmètre du carré que dessinaient les

haut-parleurs. L'année suivante, j'ai trouvé une solution graphique permettant de contrôler les variations de fréquence liées à l'effet Doppler en fonction de la vitesse radiale. J'avais réalisé que, pour modéliser correctement l'effet Doppler, il fallait utiliser des lignes à retard variable, mais cette option n'était pas pratique en raison des limites de calcul. »<sup>4</sup>

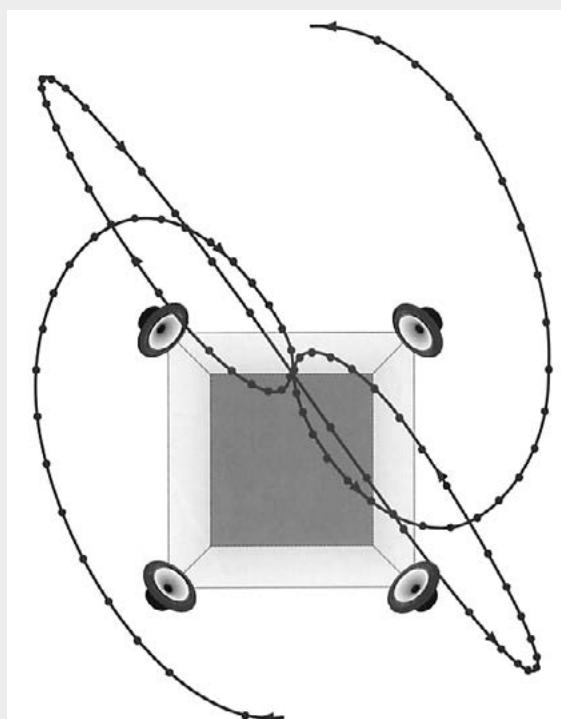


Figure 1. Une figure de spatialisation réalisée par John Chowning dans *Turenas* (1971-72) à partir de son algorithme de spatialisation. Les points représentent la position de la source apparente à intervalle de temps constant rendant ainsi des effets d'accélération et de ralentissement caractérisant cette trajectoire sonore (© John Chowning, avec son aimable autorisation).

L'algorithme de Chowning agit sur l'amplitude des signaux acoustiques émis par chaque haut-parleur pour définir la direction de la source, et sur le dosage entre le signal direct et le signal réverbéré pour modifier la profondeur apparente. Afin d'accroître la sensation de vitesse dans les mouvements rapides, il simule l'effet Doppler en élevant légèrement les fréquences lorsque les sources se rapprochent ou, au contraire, en les abaissant lorsqu'elles s'éloignent (Chowning, 1971). *Turenas* (1971-72) pour sons de synthèse fixés sur support 4 pistes révèle l'efficacité de cet algorithme en mettant en scène des trajectoires sonores d'une rare complexité reproduisant dans certains passages des figures de Lissajous (Figure 1) (CD John Chowning, 1988).

## Emmanuel Nunes

Nunes est peut-être l'un des compositeurs à avoir développé le plus une pensée théorique autour de la question de l'espace. « Le lecteur qui recherchera la plus grande acuité dans l'examen des rapports concrets entre écriture musicale et écriture de l'espace, des fonctions en jeu, de la construction d'une pensée et de la tentative d'une théorisation générale, trouvera dans les textes d'Emmanuel Nunes le plus haut degré d'intégration », souligne à juste titre Bioteau (2013, p. 1365). Nous nous contenterons ici d'aborder succinctement le processus de création de *Quodlibet* (1987-88) composé pour un lieu spécifique – le *Coliseu dos Recreios* à Lisbonne – dont il connaissait les défauts acoustiques pour s'y être souvent rendu lorsqu'il était jeune. Il décide ainsi de placer le public uniquement au parterre de cette immense salle octogonale. Les 79 musiciens occupent non seulement la scène, mais aussi les premières loges, les deuxième loges et la galerie supérieure

proche de la voûte. Ils sont regroupés en quatre catégories : un orchestre de 45 musiciens immobiles, 7 solistes immobiles, 6 percussionnistes semi-mobiles responsables de 10 sets d'instruments invisibles pour le public, et un ensemble de 21 instrumentistes mobiles (CD Emmanuel Nunes, 2001). Dans cette œuvre originellement conçue *in situ* – mais ayant fait l'objet d'une adaptation dans la salle Wagram à Paris en 1992 dans le cadre du festival d'Automne –, la stratégie de Nunes s'appuie « sur un découpage du *Coliseu* en des sous-espaces aux propriétés spécifiques, dont les deux paramètres qui ressortent sont la diffusion du son et la localisation des sources. » (Bioteau, 2013, p. 1353). Durant des tests préliminaires, le compositeur ausculte ces sous-espaces depuis le parterre en demandant à quelques instrumentistes de produire des sons depuis différents endroits. « Ceci m'a fourni, *explique-t-il*, une information sur certaines spécificités des endroits par rapport aux instruments, notamment en ce qui concerne, à l'audition, les degrés de reconnaissance de leurs localisations respectives, ainsi que l'influence de celles-ci dans une perception plus ou moins diffuse de chaque famille d'instruments » (Nunes, 2001, p. 8). Si ces tests acoustiques de nature empirique n'ont aucune prétention scientifique, ils l'ont aidé à développer une poétique de la distance – jouant sur les sensations de proximité et d'éloignement des sources – ainsi que des modulations instables, expression qu'il emploie pour se référer à « un *jeu de prédominance* entre une forte perception des trajectoires sonores et une localisation affirmée des sources mises en jeu » (Nunes, 2001, p. 8). Ce type de modulation devient instable lorsqu'on entend « un changement tellement rapide de situations qu'il devient par moments impossible de tracer nettement une frontière auditive entre localisations et trajectoires » (Nunes, 2001, p. 8). ►

4

« By 1967 I had completed the spatial processing system that would allow me to place a sound at a discrete location outside the perimeter of a square where there was a loudspeaker at each of its corners. In the following year I achieved a graphic solution to controlling the Doppler shift in frequency as a function of radial velocity. While I had understood that to properly model the Doppler effect required variable delay lines, it was an option that was not practical because of the computational limits. » (Chowning, 2011, notre trad.)

### 3. Explorer les propriétés acoustiques des espaces physiques

#### Révéler l'espace par le son et le son par l'espace

Sortir des salles de concert pour explorer des espaces urbains ou des espaces naturels ouverts permet d'explorer de nouvelles formes de spatialisation et d'intégrer les sons propres à ces environnements. Mais ce qui nous intéresse ici plus particulièrement, ce sont les manières de composer et de jouer avec les singularités acoustiques d'un espace physique. Dans les années 1980, le compositeur de musique électroacoustique Michel Redolfi organise ses premiers concerts subaquatiques, invitant les spectateurs à plonger les oreilles dans l'eau pour écouter ses œuvres diffusées par des hydrophones. Les propriétés acoustiques des milieux aqueux sont différentes de celles de l'air : dans l'eau, la vitesse du son est quatre fois plus élevée que dans l'air et la transmission des ondes acoustiques ne s'opère plus par voie aérienne mais par voie solidienne, sollicitant le corps dans son intégralité. La perception des sons et de l'espace en est totalement bouleversée. Il devient difficile pour l'auditeur de localiser les sources, ce qui l'amène à expérimenter une autre forme d'ubiquité sonore.

Dans le champ des arts sonores, il n'est pas rare que l'espace physique devienne une composante indissociable des œuvres. Avant de se tourner vers la vidéo, Bill Viola (1951-), très influencé par David Tudor (1926-1996), s'intéresse à la manière d'appréhender les sons comme quelque chose de matériel. Dans *Hallway Nodes* (1972), il diffuse un son pur de 50 Hz pour créer

une onde stationnaire à l'intérieur d'un couloir dont la taille (6 m 86) correspond à la longueur d'onde du signal sinusoïdal. L'espace physique façonné pour cette installation – un couloir étroit – se transforme en une étrange caisse de résonance que les auditeurs-spectateurs sont invités à traverser pour ressentir physiquement les ventres et nœuds de pression de l'onde stationnaire ainsi créée : aux extrémités du couloir, la pression acoustique est à son maximum, alors qu'au centre, elle est à son minimum, ce qui entraîne de fortes disparités d'intensité.

Alvin Lucier (1935-2021) n'a eu de cesse tout au long de sa vie d'explorer à travers ses installations et ses performances des phénomènes acoustiques et psychoacoustiques. S'il se sert de l'espace physique pour explorer le son, il lui arrive aussi de renverser la situation en se servant du son pour sonder l'espace (Davis, 2003) comme dans *I am sitting in a Room* (1970), une de ses œuvres les plus célèbres. Pour montrer que l'espace agit comme un filtre, Lucier enregistre à l'aide d'un magnétophone un texte qu'il récite à voix haute dans un lieu clos. La bande est alors diffusée par un haut-parleur positionné dans ce même lieu et un second magnétophone enregistre le résultat sonore. Le signal enregistré est à nouveau diffusé et ré-enregistré, la procédure étant répétée ainsi plusieurs fois. Au fil des occurrences, le discours devient progressivement inintelligible. On pourrait penser que cela est lié aux imperfections du dispositif technologique alors qu'en réalité, ce sont les propriétés physiques de l'espace qui sont au cœur de ce processus graduel de transformation sonore, les fréquences de résonance de la salle venant amplifier systématiquement les mêmes plages fréquentielles de la voix parlée alors que les autres fréquences sont atténuées. Le lieu n'est

plus seulement un simple cadre de production, il est à la fois un partenaire de jeu et un matériau structurel de l'œuvre. Le nombre d'occurrences et la vitesse de déformation varient considérablement suivant les espaces dans lesquels se déroule la performance (LP et CD Alvin Lucier, 2021). L'une d'elles, réalisée dans une chambre avec de la moquette au sol et des rideaux aux fenêtres, nécessita plus de 30 itérations et dure 45 minutes.

« Quoi qu'il en soit, la moquette et les rideaux ont affaibli les fréquences de résonance, si bien qu'il a fallu plus de temps pour les révéler, mais cela nous a donné un résultat encore plus beau [...]. Avez-vous remarqué que les mélodies semblent émerger ? Chaque salle est dotée d'une mélodie spécifique, cachée en elle jusqu'à ce qu'elle soit rendue perceptible. Vous savez, j'ai l'impression que nous sommes en quelque sorte dans la même situation que les compositeurs, lorsqu'ils ont commencé à percevoir les harmoniques. Les musiciens ont toujours été conscients de leurs effets, je pense, mais le timbre restait mystérieux jusqu'à ce que quelqu'un parvienne à démontrer l'existence de ces harmoniques. Aujourd'hui, nous commençons tout juste à composer en tenant compte de l'architecture, et je suis très heureux de prendre part à ces premières expérimentations. »<sup>5</sup>

## À la recherche d'espaces réverbérants

Chaque espace physique porte donc les germes d'une mélodie. Ce constat de Lucier résonne avec les investigations que Pauline Oliveros débute dans les

années 1980 en se rendant dans des lieux tellement réverbérants que le timbre des instruments se muait en raison des multiples réflexions. En 1984, dans le cadre d'un projet discographique collectif au titre évocateur (CD *Vor der Flut (Hommage an einen Wasserspeicher)* [*Avant l'inondation (Hommage à un réservoir d'eau)*, 1985], Oliveros est invitée à se produire à Cologne dans une citerne désaffectée où elle joue deux pièces pour accordéon. En 1988, elle réitère ce type d'expérience en investissant avec Stuart Dempster et Peter Ward – aka Panaiotis – la citerne souterraine de Fort Worden à Port Townsend dans l'état de Washington, un espace physique dans lequel le temps de réverbération avoisine les quarante-cinq secondes. Les trois improvisateurs y enregistrent un premier disque intitulé *Deep listening* (CD Pauline Oliveros, Stuart Dempster, Panaiotis, 1989). Ce terme renvoie à la profondeur même de l'acte d'écoute dont il faut faire preuve pour jouer de la musique dans de tels environnements. Il désignera par la suite un groupe musical – le *Deep listening Band* –, ainsi qu'une marque déposée renvoyant à une discipline méditative.

« En improvisant ensemble et en s'écoutant les uns les autres avec attention, nos styles se mélangent sur le moment et s'entrelacent, donnant naissance à une expérience musicale collective. C'est cela que j'appelle "l'écoute profonde". Dans cet enregistrement, l'espace acoustique avec son long temps de réverbération devient un élément unificateur décisif. S'écouter les uns et les autres, mais écouter aussi les transformations apportées par les modulations spatiales, sont des processus essentiels dans ce type de musique. La citerne, de façon effective, devient un instrument joué simultanément par ►

5

« Anyway, the carpet and drapes cut down on the production of the resonant frequencies so they took longer to achieve, but it gave us a more beautiful result [...] Did you notice that tunes seem to start? Every room has its own melody, hiding there until it is made audible. You know, I feel as though we're in the same situation as composers were when they first began perceiving overtones. Musicians were always aware of their effects, I think, but timbre was mysterious until someone could demonstrate their existence. Now we're just beginning to compose with architecture in mind, and I'm very pleased to be in on these first experiments. » (Lucier, 1995, p. 100, notre trad.)

les trois compositeurs. [...] Les caractéristiques tonales des sons produits par chaque interprète sont constamment modifiées en raison des interactions avec l'acoustique de la citerne, donnant l'impression que beaucoup plus d'instruments sont présents. »<sup>6</sup>

Depuis Oliveros, bien d'autres musiciens sont partis en quête d'espaces réverbérants. Le saxophoniste Werner Lüdi enregistre ainsi en 1993 un disque en solo à l'intérieur du barrage du lac Lucendro en Suisse (CD Werner Lüdi, 1996 ; voir Turquety, ce numéro, chapitre 1). Après avoir effectué une résidence exploratoire dans l'église du Gesù à Montréal en 2014, la saxophoniste Ida Toninato investit un immeuble en béton en voie de réhabilitation dont les étages libérés de toutes cloisons offrent une réverbération avoisinant les 12 secondes. C'est en partie dans ce lieu atypique qu'elle enregistre au saxophone baryton son premier album solo (Format numérique Ida Toninato, 2016) avant de réaliser une nouvelle résidence artistique qui se déroule cette fois-ci dans le Colorado, au sein de *The Tank – Sonic Art Center*, une citerne d'eau en métal qui, dans les années 2010, s'est muée en centre d'art (Figure 2, en haut). L'ombre d'Oliveros plane sur l'album que Toninato enregistre dans cette citerne dont la réverbération atteint les 40 secondes (Format numérique Ida Toninato, 2020). Viendront ensuite d'autres projets comme le film sonore réalisé par Frédéric Dallaire et intitulé *Le rêve d'Ida*, dans lequel on voit la saxophoniste – accompagnée de plusieurs preneurs de son – explorer avec son baryton différentes acoustiques comme par exemple celle d'une cale de cargo ou d'un bunker (Figure 2, au milieu et en bas). C'est d'ailleurs dans ce vestige de la guerre froide qu'elle réalisera par la suite en collaboration avec Claude Bastien (Studio Trebuchet, Montréal) une expérience de réalité virtuelle : *Listening to the Walls*.



Figure 2. Ida Toninato en quête de nouveaux espaces réverbérants avec, de haut en bas : 1) *The Tank – Center for the Sonic Arts (Rangely, Colorado, 19 septembre 2018)* (© Photo : Elysha Poirier) ; 2) *Le cargo Mystère (port de Montréal, Québec, juin 2019)* (© Photo : Antoine Amnotte-Dupuis) ; 3) *Le bunker du Norad (Saint-Sylvestre, Québec, 18 août 2020)* (© Photo : Antoine Amnotte-Dupuis).

6

« As we improvise together, and listen intensely to one another, our styles encounter in the moment, and intermingle to make a collective music. I call the result 'deep listening.' In this recording, the acoustic space with its long reverberation time is an influential unifying presence. Listening, not only to one another but to the transformative spatial modulations, is an essential process in the music. The cistern space, in effect, is an instrument played simultaneously by all three composers. [...] The tonal qualities produced by each performer are constantly changed by interaction with the cistern acoustics, making it seem as if many more instruments are present » (Oliveros dans le livret du CD Pauline Oliveros, Stuart Dempster, Panaiotis, 1989, notre trad.).

#### 4. Suggérer d'autres espaces en restituant ou en simulant une acoustique

La composition permet de simuler toutes sortes de phénomènes sonores dont ceux qui se rapportent à ce qu'on appelle les effets de salle, comme la réverbération ou l'écho. L'espace physique réel est alors dédoublé par un espace artificiel. Dans *Territoires de l'oubli* (1977) pour piano, Tristan Murail (1947-) parvient à projeter l'auditeur dans un autre lieu, extrêmement résonant (CD Tristan Murail, 1990). Durant la totalité de la pièce, la pédale tonale du piano demeure enfoncée, si bien que chaque corde frappée par un marteau du piano résonne jusqu'à son extinction naturelle ; mais les cordes ainsi excitées font aussi résonner par sympathie d'autres cordes de l'instrument. Il en découle un halo sonore continu et instable donnant l'illusion d'être dans un espace réverbérant. Des effets d'échos jalonnent aussi toute la partition sous la forme de petits groupes de notes répétés dans le temps, suivant différents paliers dynamiques.

Au milieu des années 1980, Luciano Berio et Pierre Boulez explorent à leur tour le potentiel des résonances par sympathie du piano. Dans la *Sequenza X* (1984) pour trompette et piano résonant, la partie de piano doit être interprétée silencieusement : à aucun moment les marteaux ne doivent venir frapper directement les cordes. Les actions du pianiste visent à libérer certaines cordes du piano qui vibrent par sympathie sur certaines notes jouées à la trompette. Ces dernières sont alors prolongées délicatement comme si elles étaient produites dans un espace plus réverbérant. Il s'opère ainsi à l'écoute de l'œuvre une sensation étrange : celle d'un espace physique artificiel qui ne cesse de se métamorphoser (CD Luciano Berio, 1998).

L'année suivante, Boulez compose *Dialogue de l'ombre double* (1985) pour clarinette et support spatialisé. Dans cette œuvre dédiée à Berio à l'occasion de son sixième anniversaire, l'espace est exploré de multiples façons. Idéalement, l'instrumentiste est placé au milieu de la salle avec le public autour de lui ; il doit alors pivoter sur lui-même au fil des sections de l'œuvre de sorte à ne pas toujours tourner le dos à la même partie du public. Celui-ci est par ailleurs entouré de six haut-parleurs servant à spatialiser de manière très précise la partie de clarinette enregistrée, Boulez souhaitant analyser le matériau sonore à travers sa répartition dans l'espace physique (Gerzso, 2015). Enfin, dans la version originale, un piano était placé en coulisse avec, sous sa table d'harmonie, un haut-parleur amplifiant à certains moments le son de la clarinette, de sorte à faire résonner par sympathie les cordes du piano. Le halo sonore était à son tour amplifié et diffusé depuis les coulisses afin de créer l'illusion, pour les auditeurs, d'être dans un espace plus grand et plus réverbérant. Cet effet est aujourd'hui entièrement contrôlé en temps réel par un logiciel informatique (CD Pierre Boulez, 1998).

Il existe dorénavant de nombreux outils technologiques permettant de restituer des acoustiques ou de les créer artificiellement et, cela, en temps réel. Dans *Lieux Perdus – Théâtre acoustique I* (2014) conçu par le compositeur Pedro Garcia-Velasquez (1984-) et le metteur en scène Benjamin Lazar (1977-), le son des musiciens jouant en direct est « transporté », grâce à des outils informatiques, dans d'autres espaces : chaque auditeur – muni d'un casque audio dans la salle de concert – voyage ainsi à travers différents lieux dont il perçoit les qualités acoustiques, mais aussi les bruits ambiants (Figure 3). Pour cela, Garcia-Velasquez a élaboré avec ►

le réalisateur en informatique musicale Augustin Muller une bibliothèque acoustique de réverbération à convolution 3D accessible en ligne sur le site dédié à ce projet de recherche et création (<https://www.lieuxperdus.com/>).

« Le voyage commence dans l'église. Elle est habitée au départ par la musique de Gabrieli, essence dont elle se vide progressivement. L'église venteuse laisse alors apparaître plus clairement ce qui subsiste à ses ruines, ce qui s'est transmuté et s'emplît des nouveaux sons. [...] Afin de permettre un plus fort ancrage de ce "monde virtuel dans les casques", les musiciens sont au départ dans une autre salle. Les musiciens et le public se rejoignent vers le milieu de l'œuvre. L'écoute au casque subsiste encore un temps avant que le chef d'orchestre n'invite les auditeurs à l'ôter, et à se transporter ainsi dans le dernier lieu à visiter, l'endroit même où ils se trouvent. » ([https://www.lieuxperdus.com](https://www.lieuxperdus.com/)).

## 5. Construire des effets de perspectives sonores

Si l'écriture musicale permet de suggérer des lieux comme nous l'avons vu dans la section précédente, elle permet aussi de créer des effets de perspective sonore sans qu'il n'y ait besoin de faire appel à des procédés de spatialisation. Varèse est sans aucun doute l'un des tout premiers compositeurs à structurer ainsi sa pensée musicale et cela, bien avant les débuts de la musique électroacoustique. Ses œuvres orchestrales mettent en jeu des complexes sonores plus ou moins disparates qu'il projette, tels des rayons sonores, en répartissant

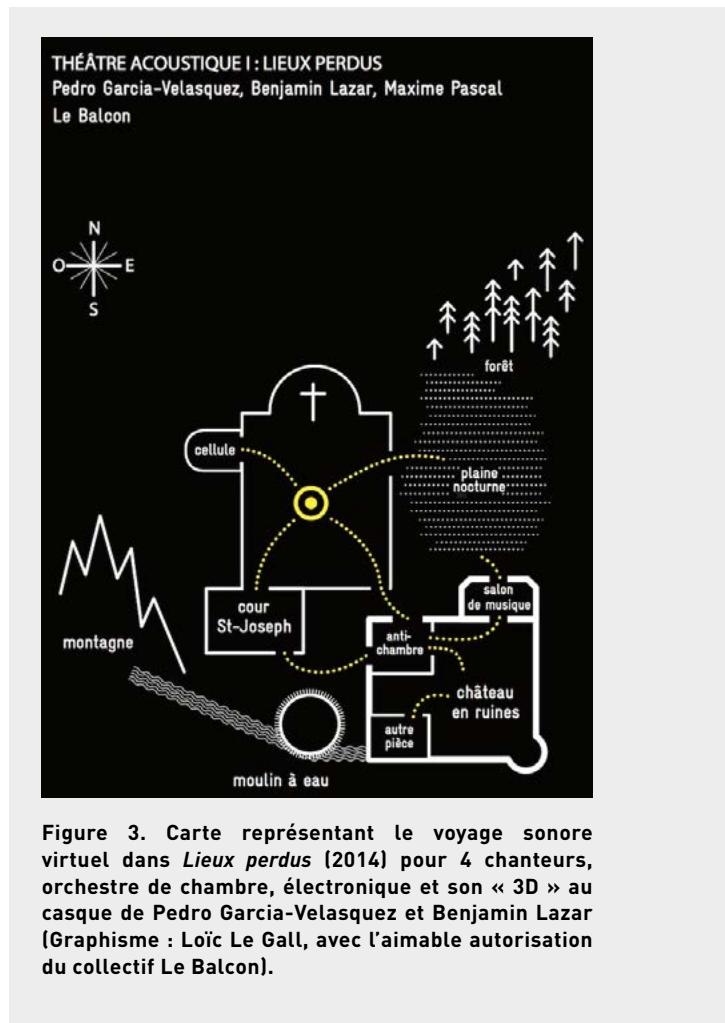


Figure 3. Carte représentant le voyage sonore virtuel dans *Lieux perdus* (2014) pour 4 chanteurs, orchestre de chambre, électronique et son « 3D » au casque de Pedro Garcia-Velasquez et Benjamin Lazar (Graphisme : Loïc Le Gall, avec l'aimable autorisation du collectif Le Balcon).

l'énergie suivant des profils dynamiques – croissants ou décroissants – dans différentes régions spectrales. Pour apprécier cette pratique, il suffit d'écouter le début de ses *Intégrales* (1923-25) pour 11 instruments à vent et 4 percussionnistes (CD Edgard Varèse, 1998), œuvre qu'il a conçue pour « la projection spatiale du son, susceptible d'être obtenue avec des médias acoustiques qui n'existaient pas alors » (Varèse [1954], 1983, p. 128-129).

Nul besoin donc de recourir à la spatialisation pour développer une pensée spatiale. György Ligeti a su lui aussi suggérer dans beaucoup de ses œuvres de fortes impressions spatiales où s'opposent la proximité et le lointain, comme dans *Lontano* (1967) pour orchestre. Dans cette œuvre au titre explicite, le compositeur hongrois, tel un orfèvre, parvient à creuser la matière sonore pour produire des effets de perspective particulièrement tangibles (CD György Ligeti, 2002). Lors d'un entretien radiophonique accordé au musicologue Josef Häusler, il revient longuement sur la construction de cet espace imaginaire.

« Les associations spatiales à l'intérieur de la musique jouent pour moi un très grand rôle, mais il s'agit d'un espace purement imaginaire. [...] À peu près dans le dernier tiers de l'ouvrage, il y a, après une plage sonore très statique et très douce, formée d'une seconde majeure et d'une tierce mineure, un mouvement très progressif dans des régions sourdes et graves. Ceci produit déjà une petite association spatiale. Et bien, cette progression s'éclaircit soudainement comme si la musique était transpercée par de la lumière par derrière. J'ai maintenant de continuelles associations lumineuses, qui jouent vraiment un rôle dans la pièce [...]. Il se produit un éclaircissement toujours croissant, la musique semble briller, rayonner. Ceci est également souligné par un *crescendo* et, en ce qui concerne les hauteurs, par une montée progressive vers des régions toujours plus hautes, jusqu'à ce que, tout à fait dans l'aigu, un son unique (*ré dièse*) se dégage et reste immobile, comme si cette lumière musicale était d'abord diffuse ; mais lentement, ce caractère disparaît, et la lumière devient un faisceau dirigé. Tout est associatif : à

ce moment, lorsque le *ré dièse* aigu se présente comme faisceau concentré de ce rayonnement musical, s'ouvre soudainement quelque gouffre, un énorme éloignement, un déchirement à travers la musique. Je forme à ce moment des associations spontanées avec l'admirable tableau d'Altdorfer, *la Bataille d'Alexandre*, qui est à l'Alte Pinakothek de Munich, et dans lequel les nuages – ces nuages bleus – se déchirent, derrière il y a le rayon lumineux doré du soleil couchant, qui transparait au travers.

Dans mon œuvre, l'image se casse soudainement. Pas totalement car le *ré dièse* résonne encore, mais à travers un *subito pianissimo* – on dirait presque qu'une lumière a été voilée – il demeure encore et continue à décroître (*verklingen*) ; une lumière peut aussi "shunter", si c'est une lumière musicale. Et cet éloignement s'ouvrant brusquement, cette étendue monstrueuse sont traversés par des sons de cors. Mais, en fait, les cors sont assis dans l'orchestre réel, il n'y a pas d'orchestre en coulisses : mais ils sonnent comme dans le lointain, pour diverses raisons associatives. D'abord, après le *tutti fortissimo*, il y a subitement *pianissimo*, et les cors jouent avec sourdine. » (Ligeti, 1974, p. 115)

## 6. Composer (avec) le rayonnement

Le rayonnement acoustique correspond à la manière dont l'énergie émise par une source sonore se diffuse dans l'espace autour d'elle. Il dépend bien évidemment de la nature de la source, mais aussi des sons émis, les fréquences aiguës tendant à être beaucoup plus directives que les fréquences graves. Il est très compliqué de modifier le rayonnement d'une source ►

acoustique sans opérer des modifications d'ordre structurel. Le rayonnement peut néanmoins être exploré musicalement en demandant aux musiciens d'orienter différemment leur instrument, à l'instar des cornistes qui peuvent aisément dresser le pavillon vers le ciel pour davantage emplir l'espace. Dans *Zeitmaße* (1955-56) pour quintette à vent, Stockhausen inscrit ainsi au-dessus de certaines notes un petit signe pour signaler aux musiciens qu'ils doivent pointer leur instrument vers le haut, ce qui modifie l'équilibre des timbres. Dans *Fonta* (1964) précédemment évoqué, Xenakis s'intéresse lui aussi à la question de la projection sonore en demandant aux musiciens (m. 335-375) de se déplacer librement d'un bout à l'autre de la scène dans des directions autant que possible opposées. À d'autres moments, alors qu'ils sont en position statique, il leur est demandé de pointer leurs instruments vers le plancher ou bien de pivoter lentement le pavillon à l'horizontale de gauche à droite suivant un angle obtus.

Ces différents exemples montrent qu'il est possible de composer avec le rayonnement mais peut-on envisager de composer le rayonnement lui-même ? Stroppa est un des rares compositeurs à s'être penché sur une telle problématique. Dans *Spirali* (1988) pour quatuor à cordes projeté dans l'espace (CD Marco Stroppa, 2009), il se sert ainsi de l'électronique pour travailler la mise au point d'un instrument ainsi que sa distance.

« Par "mise au point", j'entends qu'un son peut être non seulement centré sur un haut-parleur comme un point situé dans son milieu, mais également devenir une "surface" (dans laquelle la perception d'une directionnalité persiste, à ceci près que ce n'est plus un point précis, mais une région de l'espace), voire un son diffus partout, comme lorsqu'on écoute un instrument lointain

dans un espace très réverbérant. » (Stroppa, 2007, p. 75-76)

Dans la première version de *élet... fogytiglan* (1989-1997) écrite pour 11 instrumentistes – œuvre retirée du catalogue sous cette forme –, Stroppa recourt à un très grand nombre de microphones pour capter le rayonnement de certains instruments et l'agrandir grâce aux haut-parleurs soigneusement placés dans la salle, l'idée étant d'immerger le public dans le rayonnement instrumental et de créer ainsi une série d'images spatiales variant tout au long de la pièce. L'expérience n'ayant pas été fructueuse et compliquée à mettre en place, une seconde version sans amplification sera composée en 1997 pour 15 instruments. C'est à ce moment qu'il débute un cycle d'œuvres de musique de chambre sans électronique, mais dans lequel il explore l'espace frontal en positionnant les musiciens différemment sur scène, recherchant ce qu'il appelle « un "accord" (au sens musical du terme, comme on accorde des instruments) entre un espace réel, la scène, un espace inventé (le placement des instrumentistes à des endroits spécifiques de la scène) et une musique dont les matériaux ont été conçus pour "s'accorder" avec cet espace » (Stroppa, 2007, p. 81). Le positionnement et l'orientation des musiciens « donne une image non seulement visuelle, mais aussi acoustique qui fait sonner la scène d'une façon spécifique » (Stroppa, 2007, p. 81). Parmi les œuvres appartenant à ce cycle, figurent *Hommage à Gy. K.* (1997-2003) pour clarinette, alto et piano, *Ojia – Seven Strophes for a Literary Drone* (2005) pour violon, violoncelle et piano (Figure 4) (CD Marco Stroppa, 2018) et *Opus nainileven, 5 requies per la democrazia* (2003-2004) pour quintette à vent. Dans *Gla-dya – Études sur les rayonnements jumeaux* (2007) pour deux cors, et *Die Tanzende Hand. Versione triangolare* (2015) pour cor, trompette, trombone et

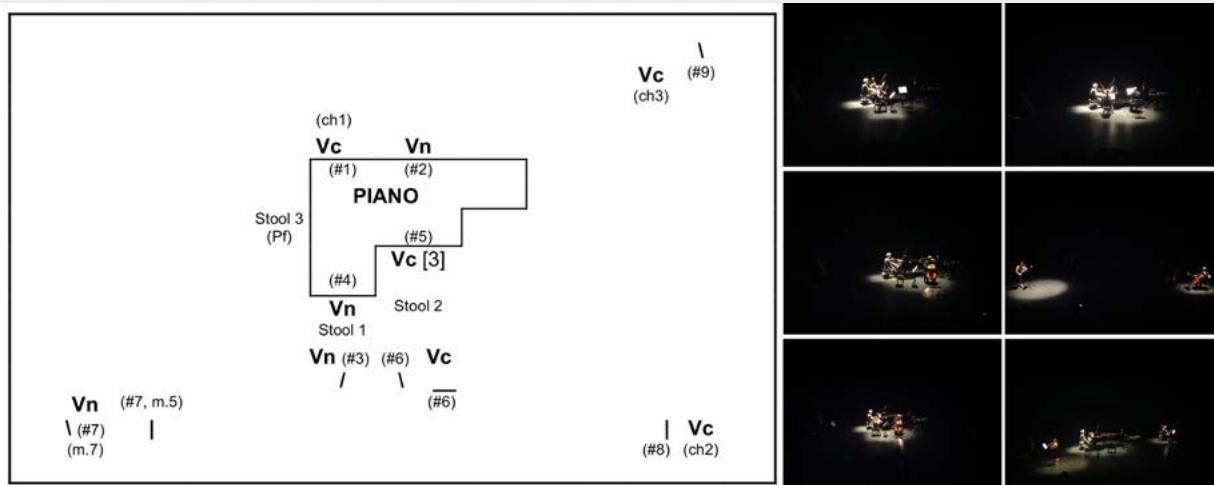


Figure 4. Marco Stroppa, *Ojia - Seven Strophes for a Literary Drone pour violon, violoncelle et piano*. Vue d'ensemble (à gauche) du dispositif scénique, incluant l'orientation des pupitres (#), des chaises (ch) et des tabourets hauts (stool) destinés au violoniste et au violoncelliste. Photos à droite d'une interprétation de l'œuvre par les solistes de l'Ensemble Intercontemporain : Jeanne-Marie Conquer (violin), Eric-Maria Couturier (violoncelle), Dimitri Vassilakis (piano) (© Marco Stroppa, avec son aimable autorisation).

tuba, le placement des musiciens change peu au cours de l'exécution, mais la question du rayonnement instrumental n'en demeure pas moins centrale.

Dans les années 1990, Stroppa débute aussi un cycle d'œuvres pour instrument soliste et « électronique de chambre » dans lequel il cherche à établir une relation plus intime entre le musicien et l'électronique, en utilisant systématiquement un dispositif de diffusion du son frontal. Dans *...of silence* (2007) pour saxophone et électronique de chambre, la quatrième pièce de ce cycle, le compositeur installe pour la première fois au centre de la scène un groupe de cinq haut-parleurs, autour duquel se place le soliste (Figure 5, en haut). Cet outil de diffusion est dérivé des recherches menées au début des années 2000 à l'Ircam autour de la *Timée*, un système de haut-parleurs dont il est possible de contrôler la directivité (Misdariis *et al.*, 2001) et

pour lequel ont été composées quelques œuvres, notamment *Duelle* (2001) pour mezzo-soprano, piano, violon et électronique de François Nicolas (1947-) et *Five Imaginary Spaces* (2002) pour piano et électronique de Oliver Schneller (1966-). Si la *Timée* est restée à l'état de prototype, il existe aujourd'hui une enceinte à rayonnement variable – Iko de IEM et Sonible – conçue avec la technologie audio 3D *Higher Order Ambisonics* (HOA). L'outil que développe Stroppa pour contrôler le rayonnement de la partie électronique de ses œuvres ou pour amplifier le rayonnement d'un instrument soliste – comme dans *Let me Let me sing into your ear* (2010) pour cor de basset et orchestre – consiste en une superposition de haut-parleurs traditionnels contrôlés individuellement par un programme informatique, pour produire ensemble différentes figures de rayonnement, mais aussi pour spatialiser le son autrement, en se servant des réflexions sur les parois de la salle. Après ►

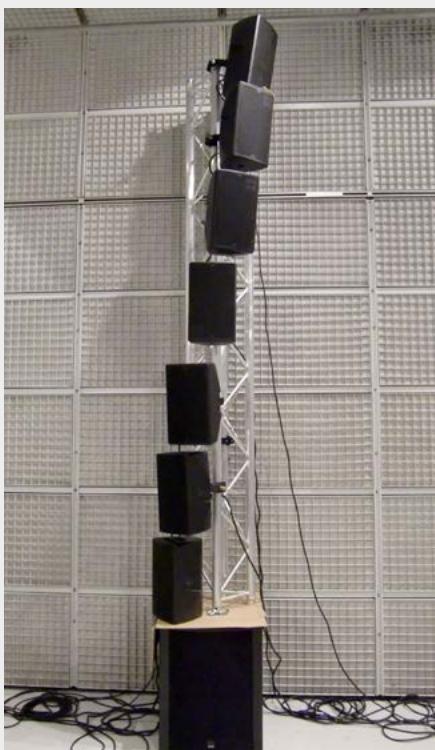


Figure 5. Totems acoustiques conçus par Marco Stroppa dans les studios de l'Ircam pour *...of silence* (en haut) et *Come play with me* (en bas) [© Marco Stroppa, avec son aimable autorisation].

*...of silence*, Stroppa poursuit cette idée en concevant dans *hist whist* (2009) pour violon et électronique de chambre diffusée par ce qu'il nommera dorénavant un « totem acoustique ». Ce type de dispositif connaîtra de nouveaux développements avec son opéra *Re orso* (2010-2012), puis dans son concerto pour électronique et orchestre *Come play with me* (2016) (Figure 5 en bas) ; CD *Donaueschinger Musiktage 2018*, 2019).

Si l'intégration de préoccupations spatiales dans la musique remonte à la Renaissance, ce n'est qu'au XXe siècle, principalement grâce au développement des outils de diffusion sonore et à l'émergence des musiques électroacoustiques, que l'espace apparaît comme un enjeu fondamental dans de nombreuses pratiques de création sonore. La localisation, le déplacement et le rayonnement des sources sonores, les propriétés acoustiques des lieux d'enregistrement et de diffusion, les qualités spatiales des timbres instrumentaux ou des sons électroacoustiques, etc. deviennent alors des paramètres compositionnels à part entière, source d'effets spatiaux tangibles. Pour illustrer ces différents aspects, nous n'avons pu sélectionner que quelques œuvres du répertoire « spatial » et avons donc dû laisser de côté un grand nombre de pièces tout aussi iconiques comme *Poésie pour pouvoir* (1958) de Boulez, *Prometeo* (1981-84) de Nono, *Cronaca del luogo* (1999) de Berio et bien d'autres.

Les recherches musicales et théoriques menées par quelques compositeurs soucieux de mieux formaliser la notion d'espace ont permis de clarifier certaines tendances, mais l'espace, en tant que dimension perceptive plurielle, devrait faire l'objet d'investigations musicologiques, acoustiques et psychoacoustiques conjointes et plus approfondies. Cela permettrait,

entre autres, d'établir un cadre formel regroupant l'ensemble des pratiques compositionnelles spatiales, de documenter et catégoriser les effets auditifs produits et, enfin, de vérifier le degré de corrélation entre les approches formalisées par les artistes et les sensations auditives vécues par les spectateurs. L'espace représente en ce sens un vaste terrain de recherches interdisciplinaires dont nous avons commencé à sonder certains aspects concernant notamment la perception des trajectoires sonores circulaires (Féron *et al.*, 2010) ou la question de l'interprétation acousmatique sur orchestre de haut-parleurs (Boutard et Féron, 2019). Mais beaucoup reste à faire.

## Remerciements

Je souhaite exprimer mon amitié et ma gratitude à John Chowning, Marco Stroppa, Ida Toninato, et les membres du collectif Le Balcon, pour l'ensemble de leurs investigations spatiales, pour leur disponibilité et pour avoir fourni à titre gracieux les différentes illustrations qui accompagnent ce texte. Merci aussi à Corsin Vogel et Delphine Roger pour leurs précieuses relectures.

## Bibliographie

> Baumaan, D. (2015), « Music and Space in the Renaissance », Brech M. et Paland R. (dir.), *Komposition für hörbaren Raum. Die frühe elektroakustische Musik und ihre Kontexte / Compositions for Audible Space. The Early Electroacoustic Music and its Contexts*, Bielefeld,

[transcript], coll. « Music and Sound Culture », pp. 45-65.

> Bayer, F. (1981), *De Schönberg à Cage. Essai sur la notion d'espace sonore dans la musique contemporaine*, Paris, Klincksieck.

> Bioteau, A. (2013), « Théorie et intégration technologique

de l'espace dans la composition. Des pionniers à Emmanuel Nunes », Donin N. et Feneyrou L. (dir.), *Théories de la composition musicale*, vol.2, pp. 1337-1366.

> Born, G. (dir.) (2013), *Music, Sound and Space. Transformations of Public and Private Experience*, Cambridge, Cambridge University Press.

> Boutard, G., Féron, F.-X. (2019), « Documenting acousmatic music interpretation: profiles of discourse across multiple dimensions », *Journal of Documentation*, vol.75 n°01, p. 99-119

> Brant, H. (1998), « Space as an Essential Aspect of Musical Composition » [1967], Schwartz E. et Childs B., *Contemporary Composers on Contemporary Music*, New-York, Da Capo, pp. 221-242.

> Brech, M., Paland, R. (dir.) (2015), *Komposition für hörbaren Raum. Die frühe elektroakustische Musik und ihre Kontexte / Compositions for Audible Space. The Early Electroacoustic Music and its Contexts*, Bielefeld, [transcript], coll. « Music and Sound Culture ».

> Chion, M. (1988), « Les deux espaces de la musique concrète », *Lien, L'espace du son I*, Ohain, Musiques et Recherches, pp. 31-33.

> Chouvel, J.-M., Solomos, M. (dir.) (1998), *L'espace, musique/ philosophie*, Paris, L'Harmattan, coll. « Musique et musicologie ».

> Chowning, J. (1971), « The simulation of moving sound sources », *Journal of Audio Engineering Society*, n°19, pp. 2-6.

- > Chowning, J. (2011), « *Turenas: the realization of a dream* », acte des *Journées d'Informatique Musicale*, Université de Saint-Etienne, 25-27 mai 2011.
- > Davis, R. (2003), « '...and what they do as they're going...': sounding space in the work of Alvin Lucier, *Organised Sound*, vol.8 n°2, pp. 205-212.
- > *Les cahiers de l'IRCAM - Recherche et Musique* (1994), n°5 « Espaces », Paris, IRCAM-Centre Pompidou.
- > *Lien – L'espace du son I, II et III* (1988/1991/2011), Ohain, Musiques et Recherches.
- > Féron, F.-X., Frissen, I., Boissinot, J., Guastavino, C. (2010), « Upper limits of auditory rotational motion perception », *Journal of the Acoustical Society of America*, vol.128 n°6, pp. 3703-3714.
- > Forsyth, M. (1987), *Architecture et musique. L'architecte, le musicien et l'auditeur, du 17e siècle à nos jours* [traduit de l'anglais par M. Haine et P. Haine], Liège, Pierre Mardaga.
- > Genevois, H., Orlarey, Y. (dir.) (1998), *Le son et l'espace*, Lyon, GRAME/ALÉAS, coll. « Musique et Sciences ».
- > Gerzso, A. (2015), « Pierre Boulez – *Dialogue de l'ombre double* », *ANALYSES – Œuvres commentées du répertoire de l'Ircam* [En ligne]. URL : <https://brahms.ircam.fr/analyses/dialogue/>
- > Harley, M. (1994), « Musique, espace et spatialisation : entretien de Iannis Xenakis avec Maria Harley », *Circuit – Musiques contemporaines*, vol.5 n°2, pp. 9-20.
- > Harley, M. (1999), « Spatiality of sound and stream segregation in twentieth century instrumental music », *Organised Sound*, vol.3 n°2, pp. 147-166.
- > Ligeti, G. (1974), « *D'Atmosphères à Lontano* : un entretien entre György Ligeti et Josef Häusler » (émission du Südwestfunk de Baden Baden diffusée les 19 juillet et 26 juillet 1968), *Musique en jeu*, n° 15, pp. 110-119 (Traduit de l'allemand par G. Tillier et D. Jameux. Publication originale parue dans Nordwall O. (1971), *György Ligeti: eine Monographie*, Mainz, Schott's Söhne).
- > Lucier, A. (1995), *Reflections / Reflexionen*, Cologne, MusicTexte.
- > Merlier, B. (2006), *Vocabulaire de l'espace en musiques électroacoustiques*, S. l., Delatour France.
- > Misdariis, N., Nicolas, F., Warusfel, O., Caussé, R. (2001), « Radiation control on multi-loudspeaker device: La Timée » - *XXth International Computer Music Conference (ICMC)* - septembre 2001.
- > Nunes, E. (2001), « La distance », texte inclus dans le livret du CD *Quodlibet*, Montaigne/ Naïve MO 782143, pp. 7-10.
- > Pinard, D. (2014), « L'espace sonore à la croisée des champs culturels : enjeux et perspectives », *Sciences humaines combinées* [En ligne], n°13, mis en ligne le 19 décembre 2017, consulté le 11 janvier 2022. URL : <http://preo.u-bourgogne.fr/shc/index.php?id=349>.
- > Reynolds, R. (1978), « Thoughts on Sound Movement and Meanings », *Perspective of New Music*, vol.XVI n°2, pp. 181-190.
- > Stockhausen, K. (1988), « Musique dans l'espace », *Contrechamps*, n°9, 1988, pp. 78-100. (Traduit de l'allemand par C. Meyer. Publication originale : « Musik im Raum », *Die Reihe*, n°5, 1959).
- > Stroppa, M. (1991), « Espace et Figure » [texte inédit].
- > Stroppa, M. (1992), « Die musikalische Beherrschung des Raumes », *Motiv - Musik in Gesellschaft anderer Künste*, n°4-5, Berlin, Verlag Constructiv, pp. 22-24.
- > Stroppa, M. (2007), « Accorder musicalement un espace réel et un espace inventé » [Entretien avec Jehanne Dautrey], *Rue Descartes*, n°56 « Musique, Architecture », pp. 70-81.
- > Varèse, E. (1983), *Écrits* (Textes réunis et présentés par Louise Hirbour. Traduction de l'anglais par Christiane Léaud), Paris, Christian Bourgois.

## Discographie (par ordre d'apparition dans le corps du texte)

- > *Cori Spezzati – Venetian Polychoral Music*, Chamber Choir of Europe / Nicol Matt (direction) (2006), Brilliant Classics [SACD Surround Sound] 93689.
- > *Alessandro Striggio – Mass For 40 And 60 Voices*, Le Concert Spirituel / Hervé Niquet (direction) (2012), Glossa [SACD Surround Sound] GCDAS 921623.
- > *Thomas Tallis – Spem In Alium: Music For Monarchs And Magnates*, The Sixteen / Harry Christophers (direction) (2003), CORO [SACD Surround Sound] 16016.
- > *Wolfgang Amadeus Mozart – Serenate, Notturmo, Eine Kleine Nachtmusik*, Le Concert Des Nations / Jordi Savall (direction) (2006), Alia Vox [SACD Surround Sound] AVSA 9846.
- > *Gustav Mahler – Symphonie no. 3*, Wiener Philharmoniker, Die Wiener Sängerknaben / Pierre Boulez (direction), Anne Sofie Von Otter (mezzo soprano) (2003), Deutsche Grammophon [SACD Surround Sound] 474 298-2.
- > *Hector Berlioz – Grande Messe des morts*, London Symphony Orchestra, London Symphony Chorus, London Philharmonic Choir / Sir Colin Davis (direction), Barry Banks (ténor) (2013), LSO Live [SACD Surround Sound] LSO0729.
- > *Charles Ives – Orchestral Works vol. 2*, Melbourne Symphony Orchestra / Andrew Davis (direction) (2016), Chandos [SACD Surround Sound] CHSA 5163.
- > *The Henry Brant Collection*, 9 volumes (2005-2008), Innova Recordings [CD].
- > *Karlheinz Stockhausen – Elektronische Musik 1952-1960* (1991), Stockhausen Verlag [CD] vol. 3.
- > *Karlheinz Stockhausen – Gruppen / Carré*, WDR Sinfonieorchester Köln, Sinfonieorchester Des Norddeutschen Rundfunks Hamburg, Chor Des Norddeutschen Rundfunks Hamburg / Karlheinz Stockhausen, Andrzej Markowski, Bruno Maderna, Mauricio Kagel, Michael Gielen, Pierre Boulez (direction) (1992), Stockhausen Verlag [CD] vol. 5.
- > *Iannis Xenakis – Metastasis, Pithoprakta, Eonta*, Ensemble Instrumental de Musique Contemporaine de Paris / Konstantin Simonovic (direction), Orchestre National de l'O.R.T.F / Maurice Le Roux (direction) (1988), Le Chant Du Monde [CD] LDC 278 368.
- > *Iannis Xenakis [s.t.] – WDR Sinfonieorchester Köln / Michel Tabachnik (direction), Ensemble Ars Nova / Marius Constant (direction), Orchestre Philharmonique de l'O.R.T.F / Charles Brück (direction) (2003), Edition RZ [CD] 1015-16.*
- > *John Chowning – Turenas, Stria, Phoné, Sabelithe* (1988), WERGO Digital Music Digital [CD] WER 2012-50.
- > *Emmanuel Nunes – Quodlibet*, Ensemble Modern, Orquestra Gulbenkian Lisboa / Kasper de Roo et Emilio Pomarico (direction) (2001), Montaigne/ Naïve [CD] MO 782143.
- > *Alvin Lucier – I Am Sitting In A Room* (2021), Sound On Paper Editions [coffret LP et CD].
- > *Pauline Oliveros, Vor Der Flut (Hommage An Einen Wasserspeicher)* (1985), Eigelstein Musikproduktion [CD] ES 2025-26.
- > *Pauline Oliveros, Stuart Dempster, Panaiotis – Deep Listening* (1989), New Albion [CD] NA 022.
- > *Werner Lüdi – Lucendro* (1996), Unit Records [CD] UTR 4064.
- > *Ida Toninato – Strangeness is gratitude* (2016), Kohlenstoff [Format numérique] KOHL 028.
- > *Ida Toninato – We Become Giants* (2020), Dragon's Eye Recordings [Format numérique] de 6031.
- > *Tristan Murail – Allégories, Vues Aériennes, Territoires de l'Oubli*, Ensemble Fa / Dominique My (direction et piano) (1990), Ac cord (una corda) [CD] 200842.
- > *Luciano Berio – Sequenzas*, Ensemble InterContemporain (1998), Deutsche Grammophon (20/21 - Music of Our Time) [CD] 457 038-2.
- > *Pierre Boulez – Répons, Dialogue De L'Ombre Double*, Ensemble InterContemporain / Pierre Boulez (direction), ►

Alain Damiens (clarinette) (1998), Deutsche Grammophon (20/21 - Music of Our Time) [CD] 457 605-2.

> *Edgard Varèse – The Complete Works*, Royal Concertgebouw Orchestra, Asko Ensemble / Riccardo Chailly (direction) (1998), Decade [CD] 460 208-2.

> *György Ligeti – The Ligeti Project II*, Berliner Philharmoniker / Jonathan Nott (direction) (2002), Teldec Classics [CD] 8573-88261-2.

> *Marco Stroppa – Traiettoria, Spirali*, Pierre-Laurent Aimard (piano), Quatuor Arditti (2009), Stradivarius [CD] STR 57008.

> *Marco Stroppa – Space*, Ensemble KNM Berlin (2018), WERGO [CD] WER 7372 2.

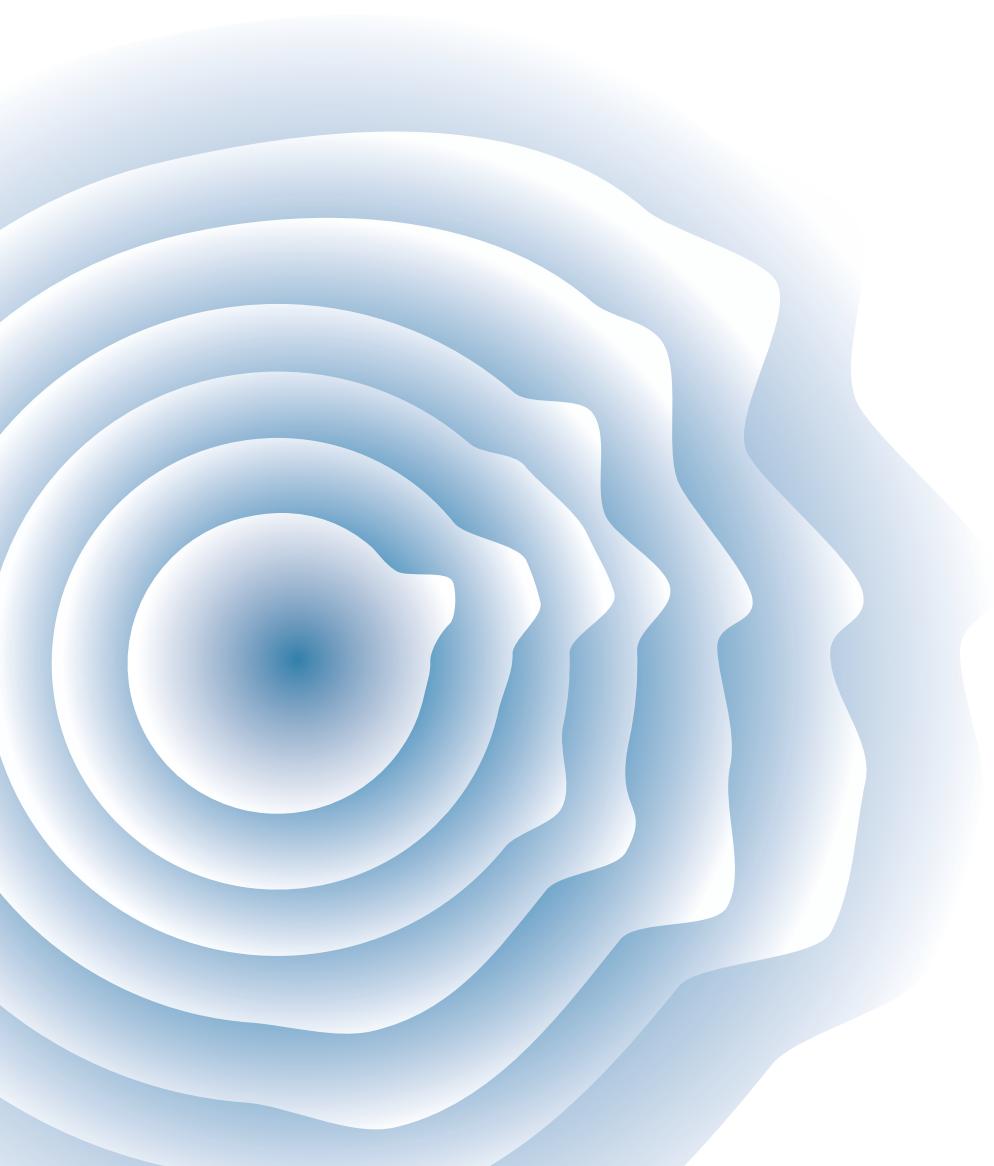
> *Donaueschinger Musiktage 2018* (Fedele - Bång - Stroppa - Zübel - Ivi evi - Filidei – Meier) (2019), Neos [SACD Surround Sound] 11914-15.

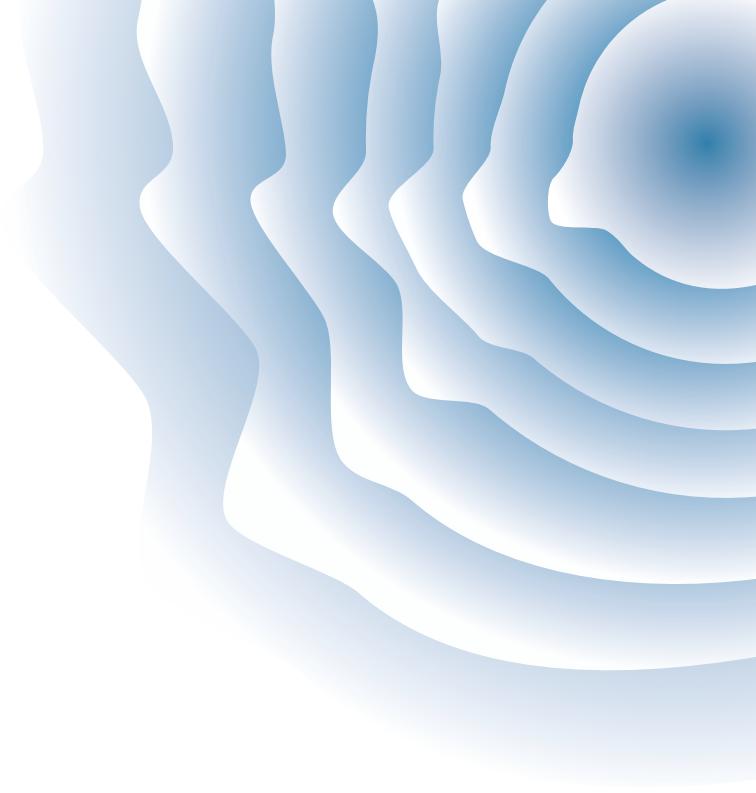
## FRANÇOIS-XAVIER FÉRON

Titulaire d'un master en acoustique musicale et d'un doctorat en musicologie, François-Xavier Féron est depuis 2013 chargé de recherche au sein du CNRS. Il a travaillé au Laboratoire Bordelais de Recherche en Informatique et au Studio de Création et de Recherche en Informatique et Musiques Expérimentales (LaBRI-SCRIME) avant de rejoindre en 2018, l'équipe Analyse des Pratiques Musicales au sein du laboratoire Sciences et Technologies de la Musique et du Son (STMS) à l'Ircam. Ses recherches de nature interdisciplinaire se concentrent sur les pratiques musicales contemporaines, du processus de création au travail d'interprétation en passant par l'analyse des œuvres et des phénomènes sonores. Il est membre collaborateur du Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT, Université McGill, Montréal), membre du comité scientifique de la revue canadienne *Circuit – Musiques contemporaines* et il co-dirige avec Alain Bonardi la base de données *ANALYSES – Œuvres commentées du répertoire de l'Ircam*.

*François-Xavier Féron holds a Master's degree in musical acoustics and a PhD in musicology (Sorbonne University). Since 2013, he has been a tenured researcher at the French National Centre for Scientific Research (CNRS). He worked at the Laboratoire Bordelais de Recherche en Informatique and at the Studio de Création et de Recherche en Informatique et Musiques Expérimentales (LaBRI-SCRIME) before joining, in 2018, the team Analysis of Musical Practices within the laboratory Sciences and Technologies of Music and Sound (STMS) at Ircam. His interdisciplinary research deals with various contemporary musical practices, focusing on creative processes, performances, and analysis of works and sound phenomena. He is also a collaborating member at the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT, McGill University, Montreal), co-editor in chief with Alain Bonardi of Ircam ANALYSES database and member of the scientific committee of the Canadian journal Circuit – Musiques contemporaines.*







**2.**

## **Pratiques et techniques**

## Enjeux de l'immersion dans une acoustique virtuelle pour les systèmes de spatialisation auditive

Jean-Pascal Jullien

### Résumé

Les compositeurs ont plutôt vu l'espace en trois dimensions comme une contrainte pour assurer une bonne audition au moment du concert. Parfois, ils ont pu en tenir compte dès le moment de l'écriture de la partition. Aujourd'hui, les outils de manipulation du son ont considérablement enrichi le monde des objets sonores accessibles au compositeur, et les systèmes de spatialisation auditive virtuelle permettent de considérer l'espace non plus comme une contrainte, mais au contraire comme l'opportunité de créer différemment de nouveaux objets sonores.

### Abstract

*For a composer, the three-dimensional space was rather seen as a constraint to ensure good hearing at the time of the concert, which could possibly be considered when writing the score. Today, sound manipulation tools have considerably enriched the world of sound objects accessible to the composer, and virtual auditory spatialization systems make it possible to understand space no longer as a constraint but on the contrary as the opportunity to create differently new sound objects.*

**1** Le niveau sonore et le contenu fréquentiel de ce qu'émet un instrument peut varier avec la direction d'émission. Un son direct et une réflexion peuvent donc avoir des contenus fréquentiels différents.

**2** Le signal source est un signal objet fourni en amont du système de spatialisation par un système de prise de son ou par la lecture d'un enregistrement, par exemple Dolby Atmos. À chaque signal objet sont associés des paramètres descriptifs variables dans le temps, comme sa localisation dans l'espace, qui sont interprétés par le système de spatialisation pour définir comment ce signal objet doit être diffusé sur les haut-parleurs.

L'expression de « spatialisation auditive » évoque avant tout l'illusion auditive, créée par un système de sonorisation, qu'une source sonore peut être localisée n'importe où dans l'espace. Si le même système possède un module d'acoustique virtuelle, il peut aussi donner l'illusion à l'auditeur qu'il est immergé dans un champ réverbéré similaire à celui d'une salle. Cela permet de corriger certains artéfacts que l'on passera en revue. On peut ainsi contrôler les trajectoires de déplacement dans l'espace des objets sonores et aussi des attributs spatiaux comme leurs étendues et leurs profondeurs. Cela se rajoute aux outils classiques du compositeur pour créer des formes musicales. L'approche artistique du compositeur Jean-Marc Duchenne, créateur de nombreuses œuvres spatiales, aboutit à une conclusion analogue en développant la notion de masse spatiale.

## 1. Terminologie

Les systèmes de réalité virtuelle mettent en œuvre des objets physiques aussi bien que des objets simulés. Ces deux types d'objets partagent certains attributs alors que d'autres sont spécifiques à chacun d'entre eux. Pour clarifier, on précise le sens donné à certains termes.

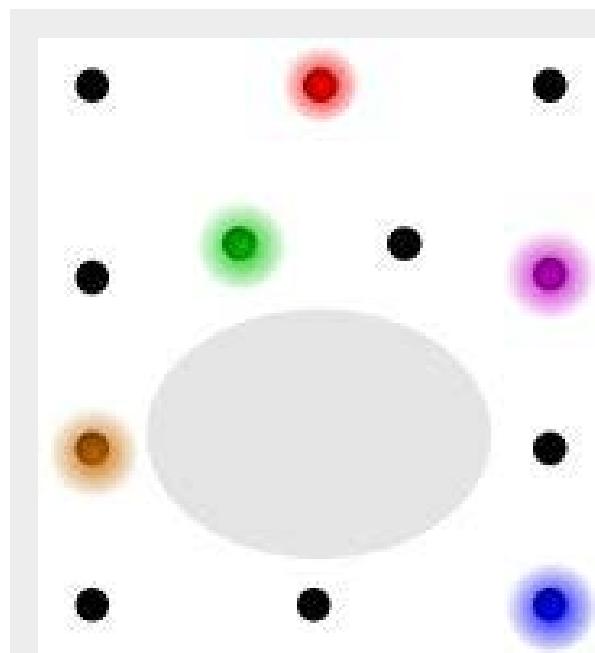
Une « source sonore » est un objet physique qui crée un champ acoustique, par exemple un instrument ou un haut-parleur. On lui associe un signal, une position dans l'espace et un diagramme de rayonnement orienté<sup>1</sup>. Un des objectifs d'un système de spatialisation auditive virtuelle est de donner l'illusion auditive qu'une telle source sonore existe quelque part.

Une « source virtuelle » est créée à partir d'un « signal source<sup>2</sup> » qui est diffusé sur plusieurs haut-parleurs, et auquel sont éventuellement appliqués des retards, des gains<sup>3</sup>, voire des filtrages différents. On l'appelle

<sup>3</sup> Un gain sur un haut-parleur est l'ajustement du niveau du signal à l'entrée de ce haut-parleur pour obtenir la puissance d'émission sonore souhaitée.

« source fantôme » lorsqu'on la localise entre plusieurs haut-parleurs. On distingue aussi deux types de sources virtuelles qui partagent le même signal source, la « source initiale » qui détermine la position perçue (comme le son direct dans une salle) et les « réflexions virtuelles » qui modifient la perception de la source sans modifier sa position (comme les réflexions acoustiques dans une salle).

Une « source perçue » est la forme (Gestalt) donnée par la perception auditive. Cette forme peut être produite par un ou plusieurs éléments sonores. Par exemple, en présence du seul son direct émis par un instrument, un auditeur perçoit une certaine forme musicale. Si l'on rajoute des réflexions apportées par une salle, et bien que les signaux du son direct et des réflexions soient à peu près identiques (au diagramme ►



Les haut-parleurs sont symbolisés par les points noirs et les couleurs représentent les signaux de source affectés aux haut-parleurs.

La zone d'écoute, en grisé, est située entre les haut-parleurs.

Figure 1 : représentation symbolique d'un espace cloisonné vu de haut (Duchenne, 2021)

de rayonnement de l'instrument près), l'auditeur perçoit une forme différente. La combinaison du son direct et des réflexions change le timbre et la présence de l'instrument. Elle lui donne aussi une « étendue », c'est-à-dire que la source n'est plus perçue comme occupant un point dans l'espace, mais comme ayant une certaine largeur ou étendue.

Un module « d'acoustique virtuelle » crée des réflexions virtuelles pour simuler l'acoustique d'une salle <sup>4</sup>. Comme dans une salle, les réflexions précoces modifient la perception des sources initiales. La réverbération tardive leur apporte une « profondeur » d'immersion dans le champ réverbéré<sup>5</sup>. Plus ce dernier est important, plus les sources semblent lointaines et profondes.

Un « système de spatialisation auditive virtuelle » est un système de sonorisation et de traitement du signal qui permet de créer, d'une part, des sources virtuelles et, d'autre part, une acoustique virtuelle. Comme il n'y a pas d'ambiguïté, dans la suite on se contentera de l'appeler « système de spatialisation ». On ne parle ici que de la partie « *rendering* » des systèmes de spatialisation dans une salle, c'est-à-dire des méthodes de diffusion des signaux sources sur un système de haut-parleurs installé dans une salle, autour d'une zone d'écoute étendue. Comme la zone d'écoute est étendue, il y a inévitablement des inhomogénéités de perception lorsqu'on se déplace sur cette zone.

## 2. Localisation des sources virtuelles initiales

Comment peut-on donner l'illusion qu'un son vient d'un certain point ? La solution naturelle est d'y placer un haut-parleur et d'y diffuser le signal source.

<sup>4</sup>

On distingue classiquement les réflexions précoces, qui arrivent après le son direct avec un retard inférieur à 80 ms, et la réverbération tardive qui est constituée de toutes les réflexions qui arrivent après ces 80 ms.

<sup>5</sup>

À noter que cette acoustique virtuelle se rajoute à l'acoustique de la salle où est installé le système de spatialisation. Pour un bon contrôle des effets perceptifs du système, il est donc souhaitable d'opérer dans une salle avec peu de réverbération.

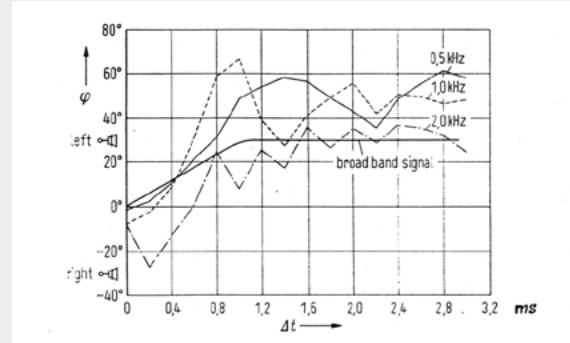


Figure 2 : Direction de l'événement auditif  $\varphi$  en fonction du retard  $\Delta t$  du signal du haut-parleur droit (d'après BLAUERT & COBBEN, 1978, in Theile, 1980)

Jean-Marc Duchenne (2021) appelle cette configuration « l'espace cloisonné ». C'est techniquement le plus simple à mettre en œuvre : un signal source est diffusé sur un seul haut-parleur. Les localisations sont excellentes, mais figées (sauf dans le cas, rare, d'un haut-parleur mobile). On verra plus loin comment Jean-Marc Duchenne a considérablement enrichi cette configuration pour réaliser ce qu'il appelle des profils de « masses spatiales » qui sont des attributs spatiaux des sources perçues.

### Création de sources fantômes

Pour acquérir une plus grande flexibilité dans le placement des sources, sans multiplier le nombre de haut-parleurs, on utilise l'illusion auditive qui fait entendre une source fantôme entre plusieurs haut-parleurs.

La stéréophonie sur deux haut-parleurs est la référence de base parce que les techniques associées de prise de son et de mixage ont acquis leurs lettres de noblesse.

Elles permettent d'obtenir un espace cohérent grâce à une prise de son d'ensemble et de créer des sources fantômes à partir de signaux monophoniques.

L'illusion auditive est causée par le phénomène de « sommation spatiale » : deux sons identiques, qui viennent de deux directions différentes et dont l'un est retardé de moins de 1 ms, sont perçus comme une seule source venant d'une direction intermédiaire.

La figure 2, illustrant le résultat des travaux de Blauert & Cobben (1978), montre comment la direction d'arrivée perçue (angle  $\varphi$ ) varie lorsque le deuxième son (émis à la même intensité par le haut-parleur de droite) est reçu avec un retard  $\Delta t$  variable. Les résultats obtenus avec un signal large bande sont bien lisibles : lorsque le retard est nul, la source fantôme est bien localisée au milieu des deux haut-parleurs ( $\varphi = 0$ ) ; lorsque le retard croît, l'angle  $\varphi$  croît linéairement jusqu'à la valeur  $30^\circ$  où se trouve le haut-parleur de gauche, celui qui émet en premier ; lorsque le retard est supérieur à 1ms, la source fantôme reste sur ce haut-parleur de gauche. C'est ce qu'on appelle l'effet de précedence ou effet Haas. Pour les fréquences pures, on retrouve des localisations plus incertaines, et assez variables d'un individu à l'autre. Pour une bonne écoute stéréophonique entre deux haut-parleurs, il faut donc éviter que le son d'un haut-parleur arrive après l'autre avec un retard supérieur à 1 ms, ce qui correspond à une différence de trajets de 34 cm. Il faut rester au milieu des deux haut-parleurs, ce qui est impossible lorsque la zone d'écoute est étendue.

La figure 4 donne une idée de la zone où les auditeurs bénéficient de cet effet de sommation spatiale et entendent bien une source fantôme entre les deux haut-parleurs. En dehors de cette zone, les auditeurs entendront toujours la source sur un des deux haut-parleurs. Si la source se déplace, on l'entendra « sauter »

d'un haut-parleur à l'autre, plus ou moins nettement suivant son étendue perçue. Pour minimiser ces sauts et discrétiser plus finement les trajectoires des sons, il faut augmenter la résolution spatiale en multipliant le nombre des haut-parleurs.

### **Perception auditive de la distance et de la profondeur**

L'indice auditif principal pour évaluer la distance à laquelle se trouve une source est son niveau sonore. Lorsqu'un auditeur écoute un instrument dont il est familier, il peut en estimer une distance. Ceci est approximatif, car l'instrument peut jouer plus ou moins fort et les premières réflexions de la salle peuvent créer un effet de loupe acoustique (distance perçue diminuée jusqu'à 40% au fond de la salle)<sup>6</sup>. L'auditeur peut aussi exploiter les variations de timbre de l'instrument quand il joue plus ou moins fort, ce qui affine son estimation. Dans une salle réelle, cette estimation auditive peut facilement être confortée ou corrigée par la vision.

Ce n'est pas le cas pour des performances entièrement sonorisées et sans images, ce qui nous intéresse plus particulièrement ici. Sans référence auditive ni visuelle sur une source virtuelle, l'auditeur ne peut pas se fier au niveau sonore qu'il perçoit pour en déduire une distance<sup>7</sup>.

Un nouvel indice peut être utilisé, c'est la comparaison des niveaux sonores de la source et du champ réverbéré d'une salle. En effet, dans une salle qui a une certaine réverbération, le rapport des deux niveaux est directement lié à la distance source-auditeur. Il faut connaître la salle pour utiliser pleinement cet indice car ce rapport diffère d'une salle à l'autre. En revanche, même si on ne connaît pas la salle, cet indice permet ►

**6**

Au fond d'une salle de concert ou d'opéra, l'énergie des premières réflexions peut dépasser celle du son direct. Comme ces deux énergies se cumulent pour donner l'impression auditive de distance, la source semble plus proche (effet de loupe) avec les réflexions que sans elles (écoute en plein air).

**7**

On ne développe pas ici le rapport entre la vision et l'audition mais on peut rappeler que la vision est capable de corriger de fortes distorsions auditives. Autrement dit, les indices visuels l'emportent presque toujours sur les indices auditifs.

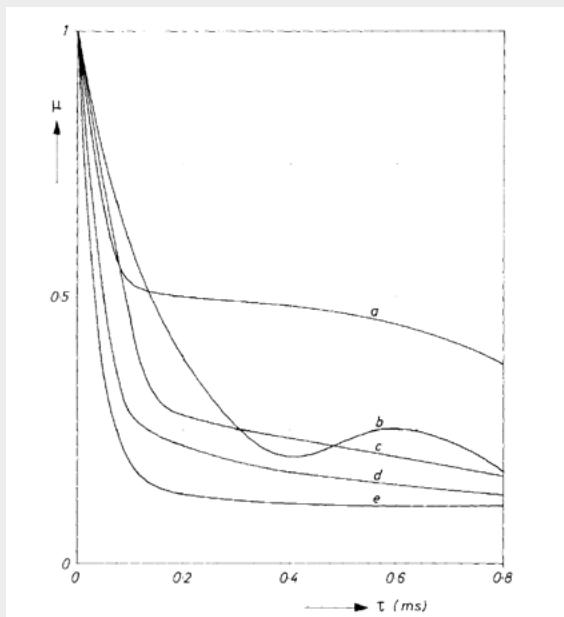


Figure 3 : Corrélation de différents signaux avec eux-mêmes en fonction du retard. (a) musique d'orgue (Bach), (b) orchestre symphonique (Glasunov), (c) clavecin (Frescobaldi), (d) chant et piano (Grieg), (e) voix d'homme parlée, d'après KUTTRUFF H.

dans tous les cas de distinguer entre un éloignement et une baisse du niveau d'émission de la source : dans le premier cas, le niveau du champ réverbéré reste quasiment constant, tandis que dans le second cas, il diminue avec celui de la source (Cremer & Müller, 1982, pp. 189-213).

Avec un système de spatialisation, le niveau du champ réverbéré, qu'engendre une source virtuelle dans la salle où est installé le système, décroît en même temps que celui de la source. L'auditeur interprète donc cela comme une simple baisse du niveau d'émission. Pour donner l'illusion d'un éloignement, il faut rajouter une acoustique virtuelle qui maintienne constant le niveau sonore d'un champ réverbéré virtuel. La plus ou moins grande immersion dans ce champ réverbéré permet

de contrôler l'éloignement/profondeur des sources virtuelles<sup>8</sup>.

### Maintien d'un niveau d'émission constant

Si, contrairement à ce qui a été vu dans la section précédente, on ne désire pas modifier l'effet de profondeur lors du déplacement d'une source virtuelle, il faut maintenir un niveau d'émission constant pour celle-ci. Comme le déplacement de la source se fait en variant des gains sur les haut-parleurs qui la diffusent, cela revient à respecter certaines règles sur ces gains. Voyons d'abord comment les niveaux sonores émanant des divers haut-parleurs s'ajoutent pour former un niveau global perçu par l'auditeur. Lorsque les signaux sont décorrélés, les niveaux s'ajoutent simplement (on dit en « intensité »), et lorsqu'ils sont parfaitement corrélés, ils s'ajoutent « en signal », ce qui se traduit par une amplification supplémentaire jusqu'à 3 dB. Dans notre cas, les divers haut-parleurs diffusent le même signal source et l'on pourrait s'attendre à ce que les signaux émanant d'eux soient corrélés. En fait, ce n'est pas vraiment le cas : comme les haut-parleurs ne sont pas tous à la même distance de l'auditeur, les signaux arrivent à ses oreilles avec un décalage temporel, et ce retard diminue la corrélation. La figure 3 permet de voir comment cette corrélation chute lorsque le retard augmente. On constate que, dès 0,2 ms, la corrélation tombe à une valeur faible pour la plupart des signaux. Les niveaux s'ajoutent alors en intensité<sup>9</sup>.

Le retard de 0,2 ms correspond à une distance de 7 cm. Il faut donc que les oreilles de l'auditeur soient quasiment à la même distance des deux haut-parleurs (à seulement 7 cm près !) pour que les niveaux s'ajoutent en signal<sup>10</sup>. Pour les autres auditeurs, les niveaux s'ajoutent en intensité. La figure 4 donne une idée de ces différentes zones.

8

A noter que l'auditeur a l'impression que la source s'éloigne sur l'axe où elle est perçue. Pour deux auditeurs différents, ces axes sont différents et les sources fantômes divergent. Cela reste vrai même si l'on multiplie le nombre des haut-parleurs.

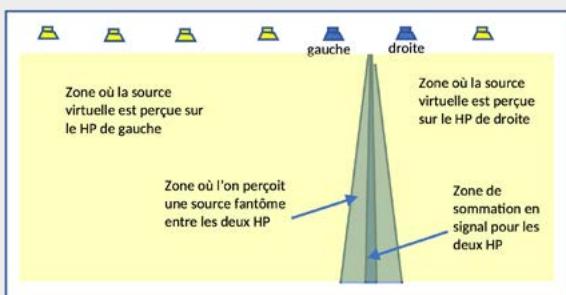
9

Les valeurs données dans la figure 3 correspondent aux signaux 'pleine bande'. Pour une fréquence basse, c'est sa demi-période qui donne la valeur pertinente du retard pour savoir si l'on somme en signal ou en intensité. Les zones de sommation en signal sont alors plus larges, ce qui explique la légère accentuation des graves dans ces zones.

10

Même si la tête est parfaitement entre deux haut-parleurs distants, les oreilles peuvent déjà être hors de la zone de sommation en signal !

La zone de sommation en signal étant très étroite, il n'y a pas vraiment lieu de s'en préoccuper : si une source fantôme se déplace entre ces deux haut-parleurs, les quelques rares auditeurs placés dans cette zone auront l'impression que la source s'approche légèrement d'eux avec un peu plus de chaleur dans le timbre, surtout au moment où elle passe au milieu. Pour tous les autres auditeurs, la puissance de la source virtuelle restera constante sur toute sa trajectoire de déplacement<sup>11</sup>.



**Figure 4 : Effets de sommation et de localisation lorsque deux haut-parleurs diffusent un même signal**

### Résolution spatiale cohérente avec notre discrimination perceptive

Pour minimiser le saut des sources d'un haut-parleur à l'autre, on peut rapprocher les haut-parleurs jusqu'à ce que ces sauts ne soient presque plus perceptibles parce que l'écart entre deux haut-parleurs est à peine discriminable. La figure 5 (Castellengo, 2015, p.130) donne une idée de cette discrimination perceptive en présentant les plus petits angles discriminables, M.A.A. ou *Minimum Audible Angle*, pour certaines incidences ou directions d'arrivée, et pour des fréquences pures.

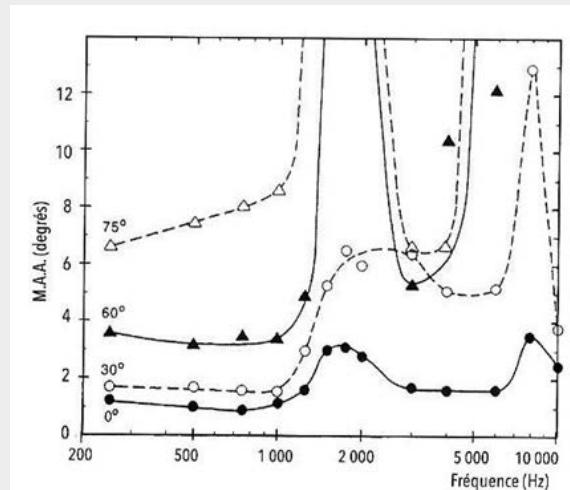
On voit que la discrimination est assez bonne en face, environ 2°, tandis qu'elle diminue sur les côtés où elle est supérieure à 5°. Par ailleurs, on retrouve ici la difficulté

<sup>11</sup>

Comme la contribution (en intensité) d'un haut-parleur est proportionnelle au carré du gain appliqué au signal diffusé, la règle revient à maintenir constante la somme des carrés de tous les gains. Cela revient à maintenir constante la puissance d'émission de ce signal sur tous les haut-parleurs utilisés.

<sup>12</sup>

La situation est encore plus favorable si l'on est à plus de 5m.



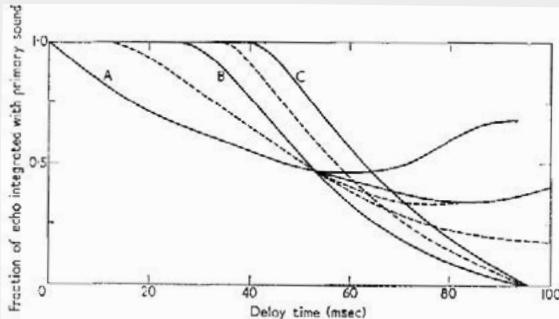
**Figure 5 : Angle minimum détectable dans le plan horizontal, en fonction de la fréquence, pour quatre incidences de la source : azimut 0°, 30°, 60° et 75°. D'après Moore (2013, figure 7.5)**

que nous avons à localiser correctement les fréquences pures, surtout autour de 2kHz.

Par exemple, pour un auditeur situé à 5 m du bord de la zone d'écoute, ces angles minimums, permettant d'éviter des sauts perceptifs, supposent de disposer un haut-parleur tous les 17 cm en face et tous les 45 cm sur les côtés ou derrière<sup>12</sup>. En d'autres termes, pour une salle de spectacle d'environ 300 places, de forme rectangulaire 24 m x 16 m, il faudrait recourir à environ 240 haut-parleurs uniquement dans le plan horizontal ! On comprend pourquoi on se contente généralement d'un nombre beaucoup plus faible au prix d'accepter pour certains auditeurs des artéfacts inévitables.

### Filtre en peigne

Un autre artéfact survient lorsque deux haut-parleurs émettent un même signal à peu près au même niveau, ►



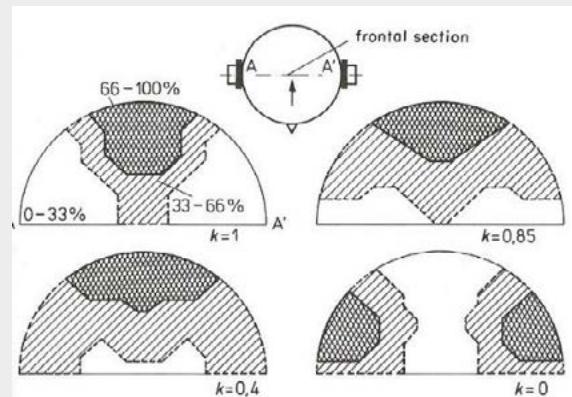
**Figure 6 : Fraction de l'énergie d'une réflexion intégrée au son direct en fonction de son retard et de son niveau par rapport au son direct +5dB(A), 0 dB (B), -5dB (C). Les pointillés sont des interpolations pour +2,5dB et -2,5dB. D'après Lochner & Burger (1964).**

ce qui arrive dès que l'on cherche à créer une source fantôme entre deux haut-parleurs. Un auditeur qui se trouve hors de la zone de perception d'une source fantôme (voir figure 4), reçoit ce signal de deux provenances et la version diffusée par le haut-parleur plus distant se trouve décalée par rapport à l'autre. Leur rencontre crée ce qu'on appelle un filtre en peigne : les fréquences qui sont en phase (le retard est un multiple de leurs périodes) sont amplifiées, tandis que d'autres, qui sont en opposition de phase, sont fortement atténuées. La perception du timbre s'en trouve perturbée.

Le même phénomène se produit également dans une salle où une première réflexion forte et à faible retard par rapport au son direct peut créer un filtre en peigne. Pour éviter cela, il suffit de rajouter des réflexions acoustiques<sup>13</sup> dont le retard n'excède pas 50 ms. Elles rajoutent de l'énergie aux fréquences atténuées et rétablissent un timbre spectralement équilibré. De même, si l'on veut s'affranchir d'un filtre en peigne pour une source fantôme, il faut aussi rajouter des réflexions

virtuelles qui arrivent avec des retards inférieurs à 50 ms sur toute la zone d'écoute.

Deux précautions supplémentaires s'avèrent nécessaires. D'une part il vaut mieux ne pas les diffuser sur les haut-parleurs de la source initiale (faute de quoi on créerait le même filtre) et d'autre part, afin de préserver la localisation, les réflexions virtuelles doivent arriver après les sons de la source initiale<sup>14</sup>. Dans une salle, les réflexions acoustiques arrivent toujours après le son direct, mais il n'en va pas de même pour les réflexions virtuelles. Pour assurer cela, il faut appliquer des retards au signal source, voire l'atténuer légèrement, avant de le diffuser sur des haut-parleurs. Ces précautions permettent que les réflexions virtuelles aient les mêmes effets bénéfiques que dans une salle. Elles pourront de surcroît atténuer les éventuels effets de sommation en signal évoqués plus haut.



**Figure 6 bis : Position des événements auditifs projetés sur une section frontale de la moitié supérieure de la tête pour un bruit à large bande et pour plusieurs valeurs de cohérence interaurale (écoute sur casque). Les zones différemment ombrées indiquent les régions plus ou moins importantes pour localiser l'événement auditif [d'après Chernyak & Dubrovsky, 1968, in Blauert, 1996]**

<sup>13</sup>

En acoustique, on rajoute des éléments diffusants qui multiplient les réflexions et diminuent leurs intensités.

<sup>14</sup>

On peut aussi diffuser le signal source sur un troisième haut-parleur (ou plus), sans se préoccuper de le retarder. Cela peut être un choix délibéré pour gérer la localisation (floue) des sources.

## Présence et étendue de la source

Ces premières réflexions ne changent pas la localisation de la source perçue, mais elles modifient sa présence et son étendue. La figure 6, qui provient d'une étude de Lochner & Burger (1964) sur l'intelligibilité de la parole, montre que, si le retard d'une réflexion reste inférieur à 50ms, cette dernière renforce complètement le signal de la source et augmente donc sa présence.

Les réflexions modifient aussi la largeur apparente de la source, son étendue. On ne dispose pas de données pour quantifier ce phénomène, mais le comportement qualitatif est connu : l'étendue croît avec le retard et l'énergie de la réflexion. Au-delà de certaines valeurs, qui dépendent du type de signal, la réflexion cesse d'être intégrée au son direct et devient un élément sonore distinct, un écho localisé dans la direction de la réflexion. Un autre moyen d'élargir la source est de multiplier les réflexions, ce qui diminue le degré de cohérence entre les oreilles. La figure 6 bis montre comment cette étendue croît lorsque la corrélation des signaux envoyés aux deux oreilles, appelée cohérence  $k$ , diminue (Blauert, 1996, p. 241)<sup>15</sup>.

## Reconstruction physique du champ acoustique

Depuis les années 1990, on a vu apparaître des systèmes de spatialisation beaucoup plus ambitieux. Avec ces « nouveaux systèmes », de conception radicalement différente par rapport aux « systèmes classiques » évoqués plus haut, l'objectif est de recréer le champ acoustique qui aurait été généré par les sources sonores que l'on cherche à simuler. C'est en quelque sorte un objectif de fidélité totale, car, si l'on plonge l'auditeur

dans le même champ acoustique, il a exactement la même perception auditive qu'avec les sources réelles.

Le projet est ambitieux, car il faut recréer un champ acoustique avec une précision spatiale de l'ordre du centimètre sur l'ensemble de la zone d'écoute ! En fait, la réalisation en est impraticable technologiquement dans la mesure où il faudrait mobiliser un nombre trop élevé de haut-parleurs et/ou de taille trop petite. Heureusement, comme nos deux oreilles sont distantes d'environ 17 cm, nous ne sommes capables d'analyser un rapport de phase entre elles que pour des longueurs d'onde environ deux fois supérieures à ces 17 cm, soit pour des fréquences inférieures à 1 kHz. On peut donc se satisfaire d'une première simplification qui consiste à limiter la reproduction exacte du champ acoustique à cette partie basse du spectre. La précision spatiale nécessaire est de l'ordre de la vingtaine de cm, ce qui permet d'utiliser des haut-parleurs du commerce.

On distingue aujourd'hui deux techniques assez différentes : la WFS (*Wave Field Synthesis*) et la HOA (*High Order Ambisonics*). Leurs principes sont décrits dans l'encadré sur la comparaison de la WFS avec la HOA (ce numéro, pp. 109-113).

La difficulté majeure qu'affrontent ces nouveaux systèmes réside dans le fait de couvrir une zone d'écoute étendue. Plus cette zone est grande, plus nombreux sont les haut-parleurs requis pour couvrir la zone.

L'IRCAM a installé des prototypes de ces deux systèmes dans l'Espace de Projection, salle parallélépipédique de 24 m x 15,5 m au sol et qui peut accueillir environ 300 auditeurs. Les lignes de WFS, installées sur les murs intègrent 264 haut-parleurs, soit environ un ►

<sup>15</sup>

Les trois valeurs supérieures de  $k$  peuvent être effectivement obtenues avec des réflexions diffuses. La valeur  $k=0$  n'est pas utile pour les premières réflexions mais cela montre l'intérêt de diffuser un champ réverbéré virtuel avec des signaux décorrélés pour ne pas former de source virtuelle localement située.

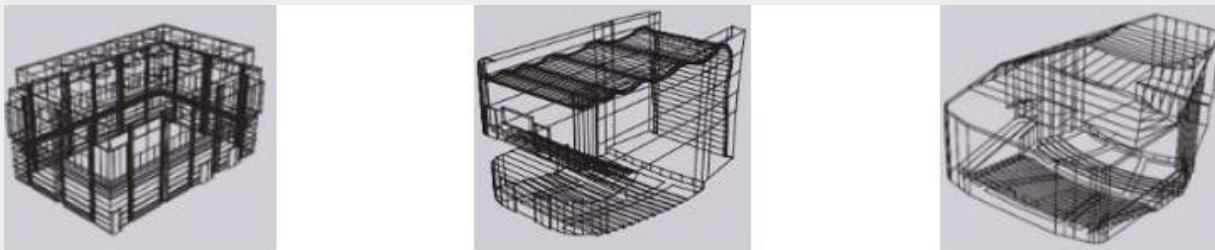


Figure 7 : exemple de prototypes de salle proposés par le système AFC du CNSMDP

équipement tous les 30cm. Ces 264 haut-parleurs sont à comparer aux 240 haut-parleurs nécessaires pour qu'un système classique offre une résolution spatiale proche de la discrimination perceptive. Cela donne le même ordre de grandeur.

Ces systèmes demandent des installations extrêmement complexes en capacité de traitement de signal et en nombre de haut-parleurs, ce qui limite aujourd'hui *de facto* leur usage.

Avec la technique HOA, sans rentrer dans les détails, il faut pratiquement le même nombre de haut-parleurs qu'avec la WFS pour une performance comparable. Dans les deux techniques, complexité et performance augmentent à peu près de la même manière avec le nombre des haut-parleurs.

Un avantage de la HOA est de pouvoir placer les haut-parleurs assez librement, en limitant les effets de proximité des enceintes que l'on constate avec la WFS. De plus, lorsqu'on diminue le nombre des haut-parleurs ou qu'on s'éloigne trop du point idéal d'écoute, la dégradation est acceptable : cela rajoute du flou à la perception de la localisation sans créer de catastrophe perceptive.

La WFS produit une scène sonore précise dans le plan horizontal et arrive même à la placer à l'intérieur de la zone d'écoute. En effet, le long de la ligne des

haut-parleurs, la WFS peut créer des ondes sphériques concaves (normalement elles sont convexes) qui viennent virtuellement de sources situées à l'intérieur de la zone d'écoute<sup>16</sup>.

Ces nouveaux systèmes placent les sources virtuelles dans l'espace avec une excellente résolution. Cependant ce n'est pas la différence principale avec les systèmes classiques. En effet, si l'on multiplie le nombre des haut-parleurs avec un système classique, on obtient la même résolution. Par ailleurs, tous ces systèmes ont besoin d'une acoustique virtuelle pour maîtriser le timbre, l'étendue et la profondeur. L'avantage exclusif des nouveaux systèmes devient évident quand l'auditeur peut se déplacer dans la zone d'écoute. Avec un système classique qui gère la profondeur, la position d'une source est perçue différemment lorsqu'on se déplace, tandis que, avec les nouveaux systèmes, elle est parfaitement stable partout sur la zone d'écoute<sup>17</sup>, quel que soit le point de vue.

### 3. Création d'une acoustique virtuelle

Autant il est facile de définir la position désirée d'une source dans l'espace grâce à un système de coordonnées géométriques, autant la définition de l'acoustique virtuelle dans laquelle on souhaite immerger les sources ne va pas de soi.

16

C'est une expérience assez spectaculaire et cela a été utilisé par l'IRCAM pour sonoriser la Tragédie du Roi Richard II, mise en scène J.-B. Sastre, Festival d'Avignon, juillet 2010. Les systèmes HOA peuvent aussi créer des sources internes mais avec un résultat moins convaincant.

17

Malgré tout, si c'est le mouvement de la source qui compte musicalement, les auditeurs perçoivent bien une trajectoire quel que soit le système. Avec un système classique, la trajectoire perçue est propre à l'auditeur. Avec un système nouveau, c'est le point de vue qui change avec l'auditeur.

## Approche architecturale

Une première solution est de définir les caractéristiques architecturales de la salle désirée, ses plans et ses habillages internes. On peut éventuellement piocher dans un catalogue de salles connues. La figure 7 montre les prototypes proposés pour un système AFC, Active Field Control, installé au CNSMDP<sup>18</sup>.

Cette première approche a l'avantage de parler directement au compositeur, qui sait par expérience comment va sonner à peu près tel ou tel type de salle. Cependant, si le compositeur désire modifier la salle pour améliorer sa réponse acoustique, il devra rentrer dans un programme de design architectural 3D qui n'a pas une interface homme-machine facile à manipuler. Il lui faudra de plus avoir des compétences acoustiques pour savoir quels paramètres architecturaux modifier pour obtenir l'effet perceptif voulu. Cela limite assez vite les possibilités de varier les acoustiques virtuelles en jouant sur ces seuls paramètres de commande.

Une fois que la salle est choisie et qu'on y a positionné une source sonore, un programme simple de simulation de la propagation du son associe une source image<sup>19</sup> à chaque réflexion apportée par les murs de la salle. En pratique on ne simule qu'un nombre limité de sources images et la réverbération, constituée des réflexions tardives, est créée par un processeur paramétré par les caractéristiques globales énergétiques, temporelles et fréquentielles du champ réverbéré.

Finalement, en plus de la source virtuelle initiale, le système créera une réflexion virtuelle pour chaque source image en respectant les contraintes vues plus haut sur les réflexions virtuelles. Par ailleurs, si la source initiale bouge, toutes les sources image bougent aussi et le programme doit calculer leurs nouvelles positions. Comme ces programmes ne fonctionnent généralement pas en temps réel, il faut calculer à l'avance les déplacements de ces sources images.

La réverbération virtuelle est assez facile à gérer, car elle dépend peu de la position des sources initiales<sup>20</sup> et l'on peut conserver les mêmes haut-parleurs pour la diffuser. En pratique les processeurs de réverbération sont bien connus aujourd'hui et permettent d'utiliser des sorties décorréélées pour éviter les interférences

Retard	Facteurs Perceptifs
0 -> 1ms	<b>Localisation fantôme</b> entre le son direct et la réflexion
0 -> 0,2ms	<b>Présence source</b> (sommation en signal)
0,2 ms -> 50 ms	<b>Présence source</b> (sommation en intensité) Risque de filtre en peigne s'il n'y a pas suffisamment de réflexions précoces Timbre : <b>Chaleur / Brillance</b>
10 ms -> 80 ms	<b>Etendue</b> de la source <b>Enveloppement sonore</b> si la réflexion arrive sur le côté Diminution de l'effet de réverbération
> 80 ms	<b>Effet de salle</b> et diminution de l'intelligibilité avec risque d'écho si des réflexions sont trop fortes <b>Réverbérance</b> <sup>21</sup> et sa couleur spectrale : lourdeur ou vivacité Profondeur de la source

**Tableau 1 : Contribution d'une réflexion à différents facteurs perceptifs (en gras) en fonction de son retard par rapport au son direct.**

**18**

AFC est un système d'acoustique virtuelle développé par YAMAHA. Sa dernière version AFC 4 a été installée dans un des studios du CNSMDP.

**19**

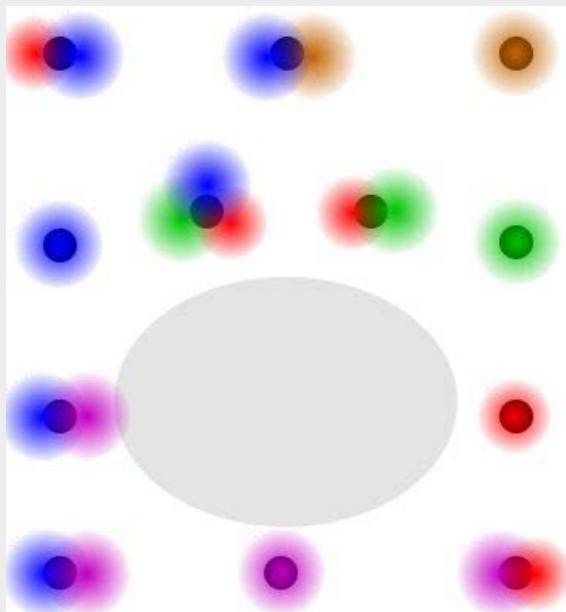
Une source image est une manière commode de simuler la réflexion acoustique d'un mur. Tout se passe comme s'il y avait une autre source de l'autre côté du mur, en miroir de ce mur.

**20**

Lorsqu'on prend soin de simuler les premières réflexions qui, elles, dépendent de la position de la source initiale, on peut se contenter de simuler une réverbération complètement indépendante de cette position.

**21**

La réverbérance est le facteur perceptif qui correspond à la perception de la durée de la réverbération.



la source, le champ réverbéré devient plus présent et l'intelligibilité du signal diminue ; lorsqu'il s'éloigne des murs latéraux, le sentiment d'enveloppement sonore se dégrade, etc. Même si le système de spatialisation fonctionne correctement, l'acoustique virtuelle reproduira ces variations non désirables sur la zone d'écoute. Il est malgré tout possible de s'en affranchir en partie avec d'autres types d'acoustique virtuelle.

### Approche perceptive

Dans cette démarche, le compositeur définit la valeur désirée pour chaque facteur perceptif et le système de spatialisation convertit ces valeurs en un jeu de paramètres techniques pour créer réflexions et

réverbération virtuelles. Le contrôle de la trajectoire et du rayonnement de la source est fait ailleurs.

Par exemple, l'interface utilisateur du programme Spat de l'IRCAM intègre les éléments suivants : "*warmth and brilliance; presence/proximity of the sound source; room presence; early and late reverberation, heaviness and liveness. Easy control over the radiation of sound sources (aperture and orientation)*"<sup>22</sup>. On y retrouve presque tous les facteurs perceptifs évoqués dans le tableau 1.

Ces facteurs sont plus ou moins connus des compositeurs et il faut

entre les haut-parleurs utilisés pour sa diffusion. Par ailleurs, lorsqu'un auditeur se déplace dans une salle, la qualité acoustique change sans pour autant s'améliorer à tout coup. Par exemple, lorsque l'auditeur s'éloigne de

éventuellement passer par une phase d'apprentissage pour mieux comprendre à quels aspects de la perception ils correspondent. Mais une fois cette compréhension acquise, il est possible d'agir

22

Informations disponibles sur le site de l'Ircam : <https://forum.ircam.fr/projects/detail/spat/>

librement sur chaque facteur, comme en bougeant un potentiomètre de table de mixage. Chacun peut alors contribuer à la structure temporelle de la composition.

Dans leur livre *Penser les sons*, Mc Adams & Bigand (1994, pp.26-39) analysent les bases multiples de ségrégation (fusion des éléments sonores et identification de formes) dans l'analyse des scènes auditives. Ils montrent que les régularités fréquentielles permettent de fusionner perceptivement des sources différentes. On reconnaît ici un des fondements de l'école spectrale (Grisey, 2008), passée maîtresse dans l'invention de timbres qui fusionnent des éléments sonores partageant une régularité harmonique de leurs partiels. Ils soulignent aussi que la plus ou moins grande cohérence des localisations et/ou des trajectoires de différents éléments sonores jouent aussi sur leur intégration perceptive. Le contrôle fin des paramètres de positionnement, de profondeur et d'étendue sur différents éléments sonores permet d'agir sur la/les forme(s) musicale(s) qu'ils produisent, ce qui apporte une spécificité spatiale à l'orchestration musicale.

### **Approche signal**

Cette dernière approche est directement liée aux éléments techniques d'un système de sonorisation : retards, filtrages, réverbération et mixage sur un ensemble de haut-parleurs. Elle est assez comparable à la précédente, mais sans la couche d'interprétation perceptive. Ici, il faut maîtriser directement les paramètres techniques pour obtenir les effets souhaités.

Dans les deux approches précédentes, la partition spatiale ne dépendait pas directement du système

de diffusion utilisé. Ici, elle est intimement liée au système : si on le change, la plupart des réglages sont modifiés.

L'intérêt d'une telle approche est qu'elle peut s'affranchir de toute règle *a priori* en exploitant toutes les ressources des outils nouveaux. Pour cela, Duchenne utilise ce qu'il appelle un « espace unifié » où le signal d'un objet peut être librement diffusé, avec des gains variables, sur chacun des haut-parleurs du système. Dans la figure 8, les haut-parleurs sont représentés par des points et le signal d'une source par une couleur. « Le système offre la souplesse nécessaire pour pouvoir juxtaposer et superposer des objets aux masses d'attributs variés et indépendants (techniquement), sans autre limite que les choix compositionnels » (Duchenne, 2021). Il propose alors de distinguer différents attributs de ce qu'il appelle une « masse spatiale ».

Que l'on parte d'une approche acoustique ou d'une approche musicale, la même conclusion s'impose : on ne peut pas se contenter de gérer uniquement la position spatiale des sources virtuelles. Les notions d'étendue et de profondeur sont indispensables pour décrire et maîtriser les attributs spatiaux d'un objet sonore.

D'une part, Jean-Marc Duchenne affirme que le simple contrôle du « site », c'est-à-dire de la localisation des objets sonores, ne présente qu'un intérêt musical limité. C'est la prise en compte des autres attributs de la masse spatiale, dont l'« aire » et la « densité », qui permet de construire des objets sonores intéressants du point de vue musical.

D'autre part, l'acousticien avance que l'apport d'une ►

acoustique virtuelle à un système de contrôle des trajectoires des sources virtuelles n'est pas simplement un luxe optionnel. Cela permet à la fois de supprimer certains défauts et d'apporter les attributs d'étendue et de profondeur aux objets sonores. Ces paramètres spatiaux permettent de configurer des formes musicales et de jouer sur d'éventuelles fusions perceptives de différents objets sonores.

Dans les deux cas, cela rajoute une certaine complexité au système de spatialisation. Cette complexité rajoutée peut être directement liée au nombre, à la disposition et aux gains associés des haut-parleurs utilisés pour diffuser un signal objet. Pour l'acousticien, cette complexité rajoutée est liée à la gestion de nouveaux objets sonores construits à partir d'un même signal source, les réflexions et la réverbération virtuelles.

La résolution spatiale, liée au nombre de haut-parleurs, reste dans tous les cas un paramètre déterminant de la performance. Les nouveaux systèmes comme la WFS ou la HOA représentent un nouveau standard d'excellence pour le placement des sources dans l'espace, mais ils ne sont pas fondamentalement différents des systèmes classiques pour gérer une acoustique virtuelle.

## Bibliographie

- > BLAUERT, J. (1996), *Spatial Hearing*, Cambridge, The MIT Press, Revised edition, 494 p.
- > BLAUERT, J., COBBEN, W. (1978), Some consideration of binaural cross correlation, analysis. *Acustica* 39, pp. 96-103.
- > CASTELLENGO, M. (2015), *Écoute musicale et acoustique*, Paris, Editions Eyrolles, 541 p.
- > CREMER, L., MÜLLER, H. A. (1982), *Principles and Applications of Room Acoustics*, Volume 1, Applied Science Publishers LTD, trad. SCHULTZ, T. J., 651 p.
- > DANIEL, J., NICOL, R., MOREAU, S. (2003), Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging, *114th Audio Engineering Society Convention*, Paper 5788, Amsterdam.
- > DUCHENNE, J.-M. (2021), *Des sons dans l'air*, [consulté en décembre 2021] : [http://sonsdanslair.free.fr/multiphonie/espace\\_compose.htm](http://sonsdanslair.free.fr/multiphonie/espace_compose.htm).
- > GRISEY, G. (2008), *Écrits ou l'invention de la musique spectrale*, édition de G. Lelong avec A.-M. Réby, Éditions MF, coll. « Répercussions », Paris (rééd. 2019), 388 p.
- > KUTTRUFF, H. (1979), *Room Acoustics*, 2nd edition, London, Applied Science Publishers LTD, 309 p.
- > LOCHNER, J. P. A., BURGER, J. F. (1964), The influence of reflections on auditorium acoustics, *Journal of Sound and Vibration*, 1 (4), pp. 426-454.
- > McADAMS, S., BIGAND, E. (1994), *Penser les sons*, Psychologie cognitive de l'audition, Paris, Presses Universitaires de France, 402 p.
- > THEILE, G. (1980), *On the localization in the superimposed soundfield*, Thèse de doctorat, Technische Universität Berlin, trad. T. Neher.

### JEAN-PASCAL JULLIEN

**Ingénieur du corps des Mines retraité, et actuellement professeur d'acoustique au Conservatoire National Supérieur de Musique et de Danse de Paris, Jean-Pascal Jullien a débuté sa carrière dans les années 1980 en tant que chercheur au Centre National d'Études des Télécommunications, maintenant Orange Labs, sur le traitement du signal numérique appliqué à la prise de son. En parallèle, il a travaillé dans l'équipe d'acoustique des salles de l'Ircam sur la perception de la qualité acoustique des salles. Il a ensuite été directeur scientifique de l'Ircam. Dans les années 2000, il a développé les services multimédias pour Orange. En 2011, il est devenu directeur de la formation Ingénieur de Télécom ParisTech et a rejoint l'équipe d'enseignants de la Formation Supérieure aux Métiers du Son du CNSMDP.**

*Retired engineer of the Corps des Mines, and currently professor of acoustics at the Conservatoire National Supérieur de Musique et de Danse de Paris, Jean-Pascal Jullien began his career in the 1980s as a researcher at the Centre National d'Études des Télécommunications, now Orange Labs, on digital signal processing applied to sound recording. He simultaneously worked in the acoustics team of Ircam on the perception of the acoustic quality of rooms. In 1991, he became scientific director of IRCAM. In the 2000s, he developed the multimedia services for Orange. In 2011, he became director of training engineering at Télécom ParisTech and joined the teaching team of the Formation Supérieure aux Métiers du Son at the CNSMDP.*

## Présentation comparée de la WFS et de la HOA (Daniel et al., 2003)

Jean-Pascal Jullien

La *Wave Field Synthesis* (WFS) et la *High Order Ambisonics* (HOA) sont des systèmes de spatialisation du son qui reconstruisent physiquement le champ acoustique. Dès l'origine, la WFS a été une solution pour servir une grande zone d'écoute, tandis que l'Ambisonics est un système surround avec une zone d'écoute limitée. Les années 1990 ont vu l'extension de ce dernier à des résolutions spatiales plus élevées (HOA), ce qui a permis d'élargir la zone d'écoute tout en apportant des propriétés intéressantes de scalabilité et de flexibilité.

### La WFS

Un ensemble instrumental crée un certain champ acoustique dans une zone d'écoute indiquée en grisé sur la Figure 1. L'objectif de la WFS est de reproduire ce champ acoustique en remplaçant l'ensemble instrumental par un ensemble de haut-parleurs distribués sur la surface qui entoure le volume de la zone d'écoute<sup>1</sup>.

L'intégrale de Kirchhoff-Helmholtz montre que cela est possible. Il faut diffuser sur des haut-parleurs omnidirectionnels et en dipôle les signaux respectivement de la pression et de la vitesse acoustique normale à la surface, captés par des microphones placés partout sur la surface. Chaque haut-parleur, dipôle ou omnidirectionnel, contribue donc à la reconstitution du champ sonore et ce sont les interférences entre tous

ces haut-parleurs qui permettent de reconstituer en chaque point les bonnes valeurs physiques du champ acoustique.

Lorsque le champ acoustique est une simple onde sphérique créée par une source sonore située au centre de cette onde, une simplification acceptable est de n'utiliser que des microphones de vitesse (pas de pression) et de rediffuser chaque signal sur un haut-parleur omnidirectionnel (pas de dipôle) placé au même endroit que le microphone utilisé.

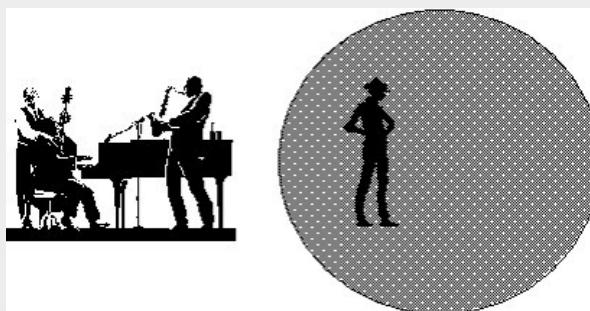


Figure 1 : auditeur à l'intérieur d'une zone d'écoute

Pour capter complètement une scène sonore, il faut donc un très grand nombre de microphones et de signaux captés. Ces microphones doivent couvrir toute la surface autour de la zone d'écoute. Et il faut autant de haut-parleurs pour reconstituer ensuite le champ acoustique. L'espacement entre les haut-parleurs détermine la précision spatiale de cette reconstitution : on ne peut reproduire exactement que les fréquences dont la longueur d'onde est supérieure au double de cet espacement. Pour les fréquences plus grandes, le phénomène d'aliasing spatial crée des interférences parasites. Il faut donc limiter cet

<sup>1</sup>

La WFS peut aussi traiter la configuration inverse où les sources sonores sont à l'intérieur du volume et la zone d'écoute à l'extérieur.

### Microphones captant la vitesse normale (bidirectionnel)

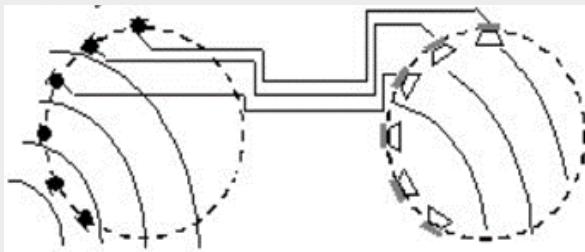


Figure 2 : simplification pour le cas d'une onde sphérique créée par une source sonore ponctuelle.

espacement à quelques dizaines de cm pour avoir une efficacité du système. Cela permet de se faire une idée du nombre de haut-parleurs qu'il faut utiliser lorsqu'on connaît les dimensions de la zone d'écoute.

Pour diminuer le nombre des microphones, on capte en proximité les différents instruments et on simule ce que captent les microphones virtuels autour de la zone d'écoute, en fonction de leurs distances et de leurs orientations par rapport à chaque instrument. C'est le système de diffusion qui fixe la géométrie de la zone d'écoute et l'emplacement où doivent se trouver ces microphones virtuels.

Si la zone d'écoute est couverte dans toutes les directions, plafond compris, les sources sonores reconstituées peuvent venir de n'importe où, et aussi d'en haut. Bien sûr, cela demande un nombre considérable de haut-parleurs. En pratique, on est surtout intéressé à reconstituer des scènes sonores qui sont dans le même plan que la zone d'écoute. Il suffit alors de placer les haut-parleurs sur une ligne qui entoure la zone d'écoute. Cela réduit fortement le nombre des haut-parleurs, mais on perd la possibilité de faire entendre des sources qui viendraient d'en haut.

## La HOA

La HOA s'appuie sur le principe mathématique que la connaissance précise du champ acoustique en un seul point de l'espace, appelé *sweet spot*, permet d'extrapoler le champ acoustique autour de ce point par un développement limité qui utilise des fonctions particulières et qui sont en nombre d'autant plus élevé que l'on s'éloigne du *sweet spot*.

Pour cela, il suffit de connaître, pour chaque fréquence, la fonction « directionnelle » qui décrit comment ce champ de pression varie localement lorsqu'on s'écarte du *sweet spot* dans n'importe quelle direction. Si l'on sait donc reproduire *au seul point du sweet spot* un champ acoustique qui ait exactement les mêmes caractéristiques directionnelles, on reproduit en même temps *partout dans l'espace* le champ acoustique désiré.

On peut montrer mathématiquement que les « harmoniques sphériques » forment une base orthonormée des fonctions directionnelles, c'est-à-dire que n'importe quelle fonction directionnelle peut être écrite comme une combinaison linéaire de ces harmoniques sphériques. Elles sont classées par ordre et sont de plus en plus complexes spatialement lorsqu'on monte dans les ordres. L'ordre 0 correspond à la fonction constante dans toutes les directions, soit omnidirectionnelle. La Figure 3 montre à quoi ressemblent ces fonctions harmoniques dans les 4 premiers ordres. Un premier indice donne l'ordre et les deux autres permettent de distinguer les différentes harmoniques de même ordre.

Pour une fréquence donnée, la fonction directionnelle a donc une coordonnée complexe (amplitude et phase) pour chaque harmonique sphérique. Si, pour

une harmonique donnée, on prend les valeurs des coordonnées pour toutes les fréquences, on obtient la représentation en fréquence du signal que capterait un microphone qui aurait exactement les mêmes caractéristiques directionnelles que cette harmonique. À chaque harmonique correspond donc un signal et l'ensemble de ces signaux représente l'intégralité du champ acoustique dans tout l'espace. Cet ensemble est appelé codage HOA de la scène sonore. Il existe maintenant commercialement des systèmes de microphones qui donnent ces signaux HOA<sup>2</sup>. Il faudrait idéalement que ces microphones soient tous au même point de l'espace, le *sweet spot*, ce qui est impossible. Cela crée un phénomène d'aliasing spatial à la prise de son. On peut aussi simuler ces microphones de manière analogue à ce qui est fait pour la WFS, ce qui affranchit complètement de ce phénomène d'aliasing spatial.

Pour calculer le champ acoustique dans l'espace autour du *sweet spot*, on part de la même décomposition sur les harmoniques sphériques, mais on modifie les coordonnées en les multipliant par un facteur qui dépend de la fréquence, de la distance d'éloignement et de l'ordre de l'harmonique. Lorsqu'on s'éloigne ou lorsqu'on monte en fréquence, cela donne plus de poids aux harmoniques d'ordre élevé. Autrement dit, pour bien représenter le champ acoustique loin du *sweet spot* ou pour des fréquences élevées, il faut monter dans les ordres et donc utiliser plus d'harmoniques sphériques<sup>3</sup>. À l'origine, l'Ambisonics n'utilisait que les harmoniques d'ordre 0 et 1, ce qui était bien adapté à une écoute individuelle sur une zone limitée. Aujourd'hui, le système installé dans l'ESPRO de l'IRCAM, dont la zone d'écoute fait environ 300 m<sup>2</sup>, monte jusqu'à l'ordre 15, ce qui donne 256 harmoniques sphériques !

Un des avantages de la HOA est que l'on est libre de placer

les haut-parleurs aux emplacements les plus pratiques. Chaque haut-parleur va avoir une contribution au *sweet*

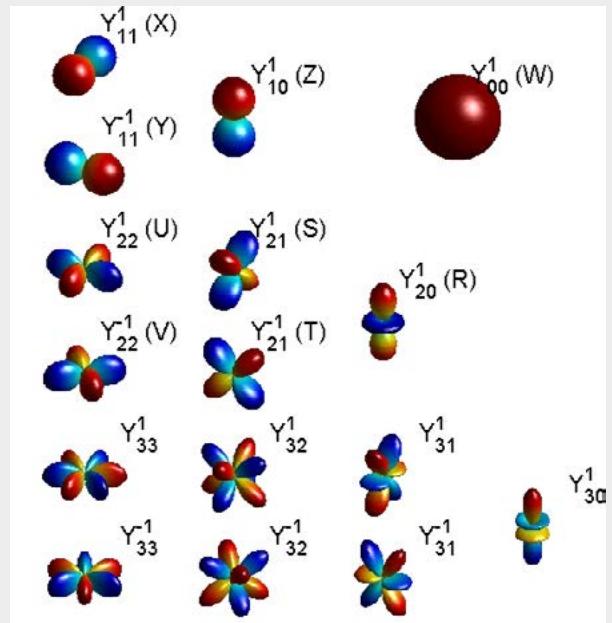


Figure 3 : les 16 harmoniques sphériques correspondantes aux ordres 0 à 3.

*spot* sur chacune des harmoniques. La reconstitution est correcte lorsque la somme des contributions de tous les haut-parleurs sur chaque harmonique est bien le signal HOA désiré sur cette harmonique. C'est un problème inverse classique qui a une solution lorsque le nombre des haut-parleurs est supérieur ou égal au nombre des harmoniques. Une fois ce réglage effectué, qui consiste à filtrer et mixer correctement chaque signal HOA sur la matrice des haut-parleurs, on peut utiliser n'importe quel jeu de signaux HOA en entrée du système. Si le système de diffusion utilise moins de haut-parleurs, par exemple parce que la zone d'écoute

<sup>2</sup>

Par exemple mh acoustics commercialise l'Eigenmike qui donne les 25 signaux d'ordre inférieur ou égal à 4 à partir d'une sphère de 32 microphones.

<sup>3</sup>

Il y a  $(M+1)^2$  harmoniques sphériques d'ordre égal ou inférieur à M. Lorsqu'on se limite à une reproduction dans un plan, il n'y a que  $2M+1$  harmoniques d'ordre égal ou inférieur à M.

est plus petite, on peut utiliser les mêmes signaux HOA en ne gardant que ceux dont l'ordre est compatible avec le système de diffusion. Le codage HOA a donc les propriétés intéressantes d'être indépendant du système de diffusion et scalable pour des systèmes de diffusion d'ordre inférieur.

### Comparaison des performances entre la WFS et la HOA

Les figures 4 à 6 comparent les performances de ces deux systèmes pour une configuration en deux dimensions avec 32 haut-parleurs espacés de 30 cm, placés sur un cercle de 3 m de diamètre. Elles donnent le résultat de la reconstitution, par chacun des systèmes, d'ondes planes de fréquence pure. L'intensité du gris représente l'amplitude de la pression à un instant donné et une reconstitution parfaite se traduirait par des lignes de même intensité parfaitement rectilignes et parallèles. On mesure une tolérance de 20 % dans la zone entourée d'une ligne bleu foncé, et une tolérance de 50 % dans celle entourée d'une ligne jaune.

L'espacement des haut-parleurs permet de reconstruire le champ acoustique sans aliasing spatial jusqu'à 600 Hz. On constate sur la figure 4 que la reconstitution est effectivement excellente pour les deux systèmes. Certains défauts subsistent sur les bords à cause de la simplification faite pour les ondes sphériques (sans dipôles de haut-parleurs).

En théorie, il faudrait que les haut-parleurs soient omnidirectionnels. En pratique, ce n'est pas le cas et de moins en moins lorsqu'on monte en fréquence. Ce défaut apparaît surtout à proximité des haut-parleurs, où le front d'onde n'est plus correctement reconstitué,

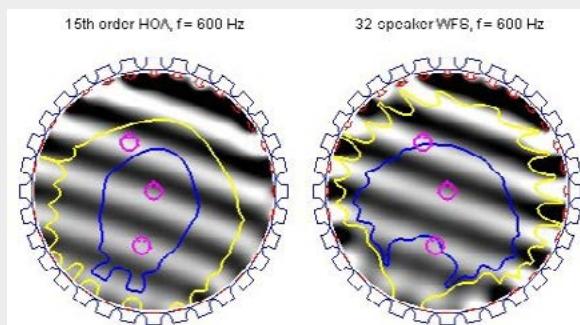


Figure 4 : comparaison de la HOA et de la WFS pour reconstruire une onde plane de 600 Hz

et surtout de ceux qui émettent le plus fort, c'est-à-dire ceux qui sont dans la direction d'où vient l'onde acoustique, ici en haut de la figure.

Les figures 5 et 6 montrent que les dégradations apportées par la HOA et la WFS deviennent, comme attendu, de plus en plus fortes lorsqu'on monte en fréquence. Pour la HOA, c'est surtout la zone d'écoute correcte qui se réduit autour du *sweet spot* placé au centre du cercle. Pour la WFS, la zone d'écoute correcte s'éloigne vers le fond où l'effet de directivité des haut-parleurs est moins marqué.

Enfin, on voit sur la figure 6 que l'effet d'aliasing spatial est très fort pour la WFS tandis que la HOA restitue partout de manière cohérente l'onde de 2kHz même si la direction de provenance est modifiée pour la faire tendre vers le haut-parleur P. La HOA est pratiquement insensible à l'effet d'aliasing spatial tout en gardant une liberté dans le placement des haut-parleurs.

On voit que les deux systèmes donnent des résultats assez comparables à complexité égale. Ils sont excellents dans leur domaine de validité pour les fréquences en deçà de l'aliasing spatial. C'est à proximité des

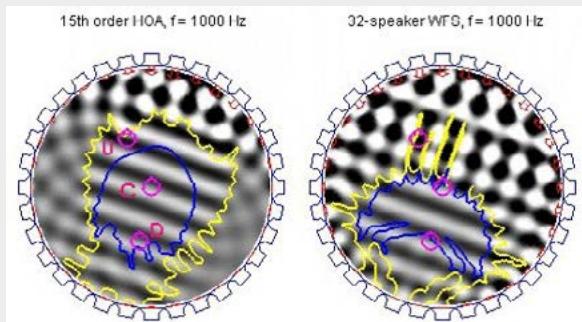


Figure 5 : comparaison de la HOA et de la WFS pour reconstruire une onde plane de 2kHz

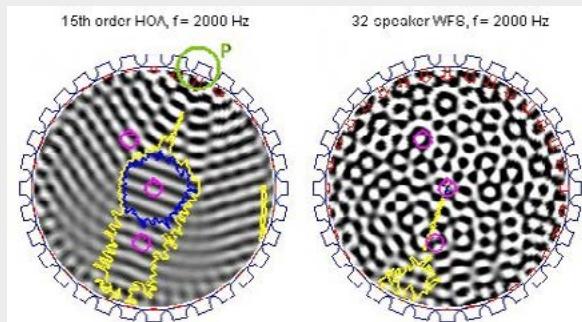


Figure 6 : comparaison de la HOA et de la WFS pour reconstruire une onde plane de 1kHz

haut-parleurs que l'on trouve les défauts les plus marqués. Pour les fréquences plus élevées, la HOA est moins sensible que la WFS au phénomène d'aliasing spatial. Si l'on considère l'intégralité du spectre audible, 50 Hz-20 kHz, on voit que ces systèmes ne peuvent reconstruire le champ acoustique que sur la partie basse du spectre audible. Il faut donc que notre système perceptif fasse abstraction, d'une manière ou d'une autre, des défauts de reconstruction sur le haut du spectre pour garder les bénéfices obtenus avec l'exactitude de la reconstruction dans la partie basse du spectre.

## Méthodes de prise de son en multicanal / Objets

Jean-Christophe Messonnier,  
Jean-Marc Lyzwa et Alexis Ling

### Résumé

Depuis les années 1990, le service audiovisuel du Conservatoire de Paris travaille sur les systèmes de captation et de restitution de l'espace sonore dédiés à l'enregistrement de la musique. Ces recherches ont abouti à des méthodes adaptées à l'audio 3D, comme l'intégration du mode multicanal/objet, le contrôle du mixage via un rendu binaural avec *headtracker*, ou des techniques transaurales. En parallèle, nous avons pratiqué sur de nombreuses productions des systèmes de captation permettant de post-produire sous ces formats. L'évolution récente des moyens de diffusion audio et audiovisuelle comme le développement du format Dolby Atmos pour la musique donnant une actualité à ces méthodes, il a semblé opportun de faire un point sur celles-ci, en décrivant les différentes raisons de cette démarche.

### Abstract

*Since the 1990s, the audiovisual department of the Conservatoire de Paris has been working on music specific systems for recording and diffusing the sound space. This research has resulted in methods particular to 3D audio, such as the integration of multichannel/object mode, mixing control via binaural rendering with headtracker, or transaural techniques. At the same time, we have put to test, on many productions, capture systems allowing post-production in these formats. The recent evolution of the means of audio and audiovisual diffusion, such as the development of the Dolby Atmos format for music, gives a topicality to these methods, and it seemed appropriate to take stock of them, by describing the different reasons for this approach.*

Depuis les années 1990, le service audiovisuel du Conservatoire de Paris travaille sur les systèmes de captation et de restitution de l'espace sonore. Ses recherches ont conduit à la mise en œuvre de techniques de prise de son et de post-production permettant d'améliorer, de préciser et de stabiliser la sensation d'enveloppement et d'immersion tout en renforçant la lisibilité et la compréhension de l'œuvre restituée. Travailler la prise de son et la post-production musicale, c'est attacher une importance toute particulière à la restitution spectrale, à la dynamique, à l'espace et à son ressenti, à la définition des sources pour une bonne intelligibilité et compréhensibilité des différents instruments ou groupes instrumentaux constituant l'œuvre. L'industrie cinématographique a très largement contribué aux développements et à la popularisation des techniques de restitutions spatiales et multicanales sans toutefois définir de canon esthétique. Après avoir expérimenté en multicanal depuis l'avènement du 5.1 et intégré le groupe Bili (<http://www.bili-project.org/le-projet/>) consacré au binaural, nous pratiquons depuis de nombreuses années des prises de son musicales, diffusées sur le site du Conservatoire de Paris, intégrant ces techniques.

Cet article propose une synthèse de notre démarche. Nous avons développé plusieurs outils (moteurs de rendu transauraux et binauraux avec *headtracker*, base de données de HRTF) nous permettant de progresser dans la qualité de nos productions et nous décrivons leur utilisation, ainsi que la manière dont nous intégrons les techniques classiques d'enregistrement.

## **Stéréophonie à deux canaux et multicanal**

La stéréophonie à deux canaux permet, à partir de deux enceintes disposées à  $\pm 30^\circ$  et d'une position précise de

l'auditeur, de donner à percevoir des sources fantômes en continu dans l'espace, d'une enceinte à l'autre, en fonction de la différence d'informations entre les deux oreilles perçue par l'auditeur (différence d'intensité et différence de temps). Pour plus d'informations sur ce phénomène, on peut se référer à Blauert (1983) et Plenge & Theile (1987). Elle permet également de percevoir les sources en profondeur, mais cette propriété est conservée en multicanal et elle ne constitue donc pas une différence entre les deux formats. On peut optimiser le signal à deux canaux, à la prise et en post-production. Cette attitude est largement dominante en musique, à la fois avec l'emploi d'un système principal microphonique à deux canaux par beaucoup d'ingénieurs du son en musique classique, et par la pratique du *mastering* sur le signal à deux canaux. Plus généralement, le *monitoring* de référence est dans la plupart des cas un système à deux canaux, sur enceintes ou au casque.

Pendant, dans la production du son pour le cinéma, un autre modèle a émergé et est maintenant largement dominant. Le signal *master* est ainsi constitué d'un ensemble de signaux décrivant la scène sonore, s'adaptant à toute une gamme de systèmes, allant de la très grande salle de cinéma équipée de dizaines d'enceintes, à l'écoute domestique pour laquelle un *downmix* deux canaux est réalisé. Nous reviendrons sur les caractéristiques de ces signaux, mais nous devons auparavant poser le débat.

Premièrement, le modèle de la production cinéma est-il exportable en partie pour la musique ? Y-a-t-il un intérêt à pratiquer le son multicanal pour la musique ? La question mérite actuellement d'être posée, car l'industrie musicale a commencé récemment à produire au format Dolby Atmos, et que la société Apple transforme ce format en binaural avec *headtracker* vers ►

ses écouteurs nomades, afin que l'auditeur perçoive la scène sonore comme étant fixe, lorsqu'il bouge la tête. Nul ne connaît l'avenir socio-économique de ces systèmes lorsqu'ils sont appliqués à la musique, mais ils impliqueraient alors une production en multicanal qui changerait radicalement les pratiques si elle s'imposait.

L'expérience que nous avons acquise depuis trois décennies dans le domaine de la captation et de la restitution en multicanal, mais également dans le développement d'outils spécifiques aux techniques en audio 3D, nous permet de donner un avis esthétique (pourquoi le faisons-nous ?) et technique (comment le faisons-nous ?). L'apport du multicanal est simple : il permet une immersion sonore beaucoup plus perceptible que la stéréophonie à deux canaux, l'immersion étant définie comme la sensation d'être à l'intérieur de l'espace restitué plutôt que devant cet espace. Cette différence d'immersion n'est pas forcément sensible dans des lieux d'écoute de petits volumes, où la réverbération de la salle crée l'essentiel de cette immersion et où, par conséquent, l'apport du multicanal n'est pas décisif. En effet, le signal émis par les enceintes est diffusé à travers l'acoustique du lieu d'écoute et l'auditeur est alors en présence du son direct des enceintes (entre  $\pm 30^\circ$ ), mais également du son réverbéré de la salle (en plus de celui diffusé par les enceintes sur  $360^\circ$ ). La perception des plans sonores qui en résulte dépend donc de ces deux facteurs (Messonnier & Moraud, 2011). Mais, dès que le volume de la salle augmente assez pour que le niveau de la réverbération soit suffisamment bas à la position d'écoute, l'immersion est créée par le signal direct du système d'écoute et le multicanal est dans ce cas un atout décisif. Cette immersion va de pair avec une possibilité accrue d'analyse du champ acoustique due au démasquage binaural (Moore, 1989) et les systèmes multicanaux sont donc plus riches en quantité

d'information perçue. L'immersion est liée à une utilisation de l'espace sonore qui peut être réalisée, soit par les sons directs, soit par l'ambiance, soit par la réverbération, et une infinité d'esthétiques peuvent être développées à partir de cette possibilité d'utiliser toutes les directions de l'espace sonore.

Le multicanal peut donc être, à notre sens, un atout. La question est donc la suivante : peut-on pratiquer la stéréophonie en multicanal pour la musique en ayant exactement la même qualité que celle de la stéréophonie à deux canaux ? Les systèmes de diffusion du son ont évolué. Indépendamment des dispositifs multicanaux, deux systèmes ont vu leur popularité exploser : l'enceinte *Bluetooth* avec un retour à la monophonie, et le casque audio pour la lecture de médias en mobilité. Pour les enceintes *Bluetooth*, la compatibilité mono est donc un enjeu. L'usage du casque ouvre aux techniques binaurales, avec ou sans *headtracker*, et cela nécessite un contenu multicanal. En parallèle avec la question précédente : continuer à produire comme avant n'est-il pas un problème ? Pour permettre cette présentation, nous devons tout d'abord rappeler quelques définitions.

### **Les modes de spatialité : multicanal ; mode objet ; ambisonics (Messonnier et al., 2016)**

Comme décrit dans la norme EBU (2014) spécifiant les modes de spatialité pouvant être inscrits dans les signaux audio (*Audio Definition Model*, ADM), on dénombre au moins trois manières de spatialiser les sons. Ces données sont stockées dans la partie réservée aux metadata des fichiers *Broadcast Wave File* (BWF). Le même type de données est utilisé dans les systèmes propriétaires Dolby Atmos DTSX et MPEGH, permettant

de plus de diffuser le signal à un débit raisonnable par des systèmes de compression de données. Ces trois modes de spatialité sont les suivants : le multicanal, le mode objet et l'ambisonie.

Le multicanal (*Channel Based Audio*), dont la stéréophonie à deux canaux fait partie, définit des canaux normalisés (L, R, C, etc.) correspondant à des directions de l'espace EBU, 2014). Ces canaux reçoivent des signaux répartis entre eux, soit par les systèmes de prise de son (un couple AB réparti par exemple le son entre L et R dans la stéréophonie à deux canaux), soit par les systèmes de post-production (potentiomètre panoramique). Si un signal n'est envoyé que sur un seul canal, ce signal se comporte comme un objet fixe. Un objet (*Object Based Audio*) est un signal défini par sa position spatiale, par exemple  $-15^\circ$  azimut,  $0^\circ$  d'élévation, à un temps (*time code*) donné. Les positions des objets peuvent donc évoluer avec le time code et donc avec l'automatisation.

L'ambisonie (*Scene Based Audio*) est basée sur une analyse régulière de l'espace sonore avec une précision donnée par un paramètre appelé ordre (voir aussi ce numéro, Jullien, chapitre 5, pp. 109-113). Chaque ordre renvoie donc à un ensemble de signaux complémentaires entre eux et représentant l'espace sonore, indépendamment du système de monitoring (en son 3D cela implique un minimum de canaux au 1er ordre : 4 canaux, 2ème ordre : 9 canaux, 3ème ordre : 16 canaux, 4ème ordre : 25 canaux, 5ème ordre : 36 canaux, 6ème ordre : 49 canaux...)

Ces différents modes de spatialité sont combinables entre eux dans l'ADM (*Audio Definition Model*, cf. ci-dessus), mais les formats propriétaires n'admettent qu'une base multicanale (appelée *Bed*) et un certain

nombre d'*objets* (le nombre dépendant des systèmes). Lorsque nous parlons du multicanal dans la partie précédente, il s'agissait d'une combinaison de multicanal et de mode objet.

## Les moteurs de rendu

En ambisonie, on a donc l'idée d'une scène sonore idéale, que l'on va devoir adapter au système de monitoring que l'on utilise. Lorsque l'on réalise une scène complexe mélangeant multicanal et mode objet, on est confronté à une situation similaire : la réalisation de la spatialité de la scène dépend du système de monitoring que l'on utilise et on doit avoir un dispositif permettant d'adapter l'information spatiale contenue dans le mixage (dans les metadata) à ce système de monitoring : c'est la fonction du moteur de rendu. Par extension, nous appellerons aussi "une scène" toute combinaison de multicanal, mode objet et ambisonie, nécessitant un moteur de rendu pour être réalisée.

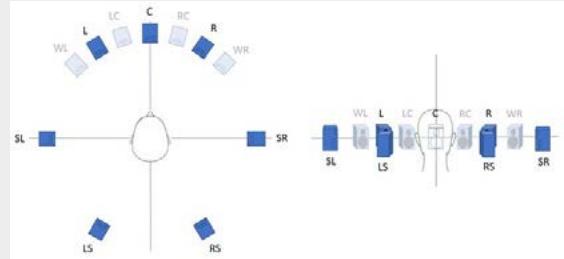
## Downmix et Upmix

Le moteur de rendu permet d'adapter un contenu multicanal/objet d'un système à un autre. Si le canal prévu dans le mixage existe, le signal sera diffusé sur l'enceinte correspondant à ce canal. Si le canal prévu n'existe pas, le signal prévu sera diffusé comme une source fantôme entre les deux enceintes correspondant aux canaux existants les plus proches, ou sera projeté dans le plan horizontal s'il n'y a pas d'enceintes en élévation dans la zone d'azimut concernée. S'il y a plusieurs plans d'élévation, le moteur réalisera une interpolation entre ces différents plans. ►

## L'optimisation du signal deux canaux

La méthode la plus simple du *downmix* deux canaux consiste à utiliser des panoramiques d'intensité classiques pour tous les angles compris entre  $-30^\circ$  et  $+30^\circ$  et pour la zone symétrique à l'arrière (entre  $-150^\circ$  et  $+150^\circ$ ), qui est repliée sur l'avant. Les objets et les canaux dont les angles sont compris entre  $-30^\circ$  et  $-150^\circ$  sont fixés sur le canal de gauche, les objets et les canaux dont les angles sont compris entre  $+30^\circ$  et  $+150^\circ$  sont fixés sur le canal de droite. On obtient ainsi un signal stéréophonique pouvant être lu classiquement au casque comme sur enceinte, mais la scène est tronquée au-delà de  $\pm 30^\circ$ .

On peut optimiser la scène sonore au casque ou sur enceinte en utilisant les techniques binaurales/transaurales. C'est l'objet des outils *Transpan* et *MyBino*, des moteurs de rendu transauraux et binauraux, dont le développement est issu d'une collaboration entre le service audiovisuel du Conservatoire et différents partenaires. Les techniques binaurales, pour une écoute au casque, utilisent des filtres reproduisant l'effet de la tête et des pavillons des oreilles sur le champ acoustique pour chacune des directions de provenance du son, nommés HRTF (*Head Related Transfer Function*). Les techniques transaurales, pour une écoute sur haut-parleurs, ajoutent à ces HRTF des filtres annulant les trajets croisés des enceintes (signal sortant du haut-parleur gauche allant sur l'oreille droite de l'auditeur, et signal sortant du haut-parleur droit allant sur l'oreille gauche), afin de retrouver ces HRTF aux oreilles de l'auditeur. Il y a un enjeu spectral dans le choix de ces HRTF. Pour *Transpan* comme pour *MyBino*, un travail important a été réalisé pour que la sonorité des filtres se rapproche le plus possible de la sonorité du signal, l'idée étant de garder l'équilibre spectral du signal



**Figure 1 : Format 7.1 avec sources fantômes (en pointillés). Exemple : Un canal Lc ou l'objet correspondant (Left Center,  $-15^\circ$  azimut ;  $0^\circ$  élévation) sera diffusé sur l'enceinte Lc si elle existe, et entre les enceintes L ( $-30^\circ$  ;  $0^\circ$ ) et C ( $0^\circ$  ;  $0^\circ$ ) si elle n'existe pas.**

stéréophonique tout en optimisant la spatialité de la scène, pour le casque ou pour le système d'enceintes.

Le moteur de rendu prend en entrée les différents signaux et affecte à chacun une voie de traitement qui va réaliser le filtre (HRTF + annulation des trajets croisés si transaural) pour une direction donnée. Il faut donc faire correspondre les réglages des directions à la configuration que l'on utilise. Pour prendre un exemple, on peut alimenter le moteur de rendu par un bus multicanal 11.1.4 et le régler sur les angles correspondants :

L ( $-30^\circ$  ;  $0^\circ$ ) R ( $+30^\circ$  ;  $0^\circ$ ) C ( $0^\circ$  ;  $0^\circ$ ) LFE ( $0^\circ$  ;  $0^\circ$ )  
 WL ( $-45^\circ$  ;  $0^\circ$ ) WR ( $+45^\circ$  ;  $0^\circ$ ) SL ( $-90^\circ$  ;  $0^\circ$ ) SR  
 ( $+90^\circ$  ;  $0^\circ$ ) LS ( $-150^\circ$  ;  $0^\circ$ ) RS ( $+150^\circ$  ;  $0^\circ$ )  
 TL ( $-45^\circ$  ;  $45^\circ$ ) TR ( $+45^\circ$  ;  $45^\circ$ ) LC ( $-15^\circ$  ;  $0^\circ$ ) RC  
 ( $+15^\circ$  ;  $0^\circ$ ) TRL ( $-135^\circ$  ;  $45^\circ$ ) TRR ( $+135^\circ$  ;  $45^\circ$ )

Idéalement, chacun des sons de la scène doit être traité par une seule HRTF pour que l'on ait une direction précise. Si un son est situé entre deux canaux, par exemple entre LC et C, il en résulte un flou de localisation, puisque la somme des HRTF ( $0^\circ$  ;  $0^\circ$ ) et ( $-15^\circ$



**Figure 2 : Représentation des directions des canaux sur trois moteurs de rendu différents : Transpan, MyBino2, MyBino1 (Note : l'ordre du bus correspond à la modification d'un bus 9.1.6 dans Pyramix, qui existe en natif et permet, de façon extrêmement pratique, une affectation des sons aux différents canaux, ce qui explique ce choix).**

; 0°) ne correspond pas exactement à la HRTF (-7,5° ; 0°). On a par conséquent le choix : utiliser plus d'objets, pour chacun des sons à spatialiser, ou accepter ce flou sur certains sons, pour toujours garder le même schéma de travail pour tous ses mixages.

Le signal ainsi obtenu peut être utilisé comme un signal stéréophonique traditionnel, même si la spatialité est optimisée soit pour le casque, soit pour les enceintes. Le moteur de rendu *MyBino2* est conçu pour générer un signal binaural optimisé pour une écoute au casque, qui puisse être entendu comme un signal stéréophonique traditionnel sur enceintes, l'image sonore étant alors anamorphosée.

Il est également possible d'utiliser un *headtracker* avec le moteur binaural (*Hedrot* pour *MyBino1* (<https://alexisbaskind.net/fr/hedrot-head-tracker/>), *Feichter* pour *MyBino2*). Ce dispositif envoie les coordonnées d'orientation de la tête de l'auditeur au moteur de rendu, afin que celui-ci change les HRTF produites pour que les sources paraissent à l'auditeur comme étant fixes dans l'espace lorsqu'il bouge la tête, comme dans

un espace sonore réel. Le contrôle de la stabilité de la scène passe par la maîtrise de l'information fournie par le *headtracker* et par l'individualisation des HRTF, tous les individus ne percevant pas la même localisation pour une HRTF donnée. Ce point constitue une direction de recherche encore en cours, mais *MyBino* apporte des débuts de réponse. Ce dispositif permet d'améliorer le réalisme des sensations sonores (Hendrickx *et al.*, 2017), notamment pour les sources situées dans la zone frontale. On peut formuler l'hypothèse qu'il existe une ambiguïté pour ces sources, le cerveau ayant des difficultés à déterminer leur position, avant ou arrière, puisque les différences de phase et d'intensité perçues par nos deux oreilles sont identiques dans ces deux cas. Cette ambiguïté est levée par le dispositif car, lorsqu'on tourne la tête, les sources se déplacent en sens inverse si elles sont à l'avant ou à l'arrière. ►

Nous utilisons ces différentes réalisations du mixage pour le contrôler. Chacune des réalisations est vérifiée sur un système particulier (mono sur une seule enceinte, deux canaux sur enceintes, deux canaux au casque, multicanal dans différents formats) et renseigne sur un aspect particulier du mixage, les localisations et le déploiement des réverbérations dans l'espace conditionnant les relations de masquage entre les différents sons qui constituent la scène. On peut passer d'un mode de monitoring à un autre en fonction de son environnement de travail. Ces modes de monitoring sont complémentaires.

## La spatialité des scènes

Les considérations précédentes ne font pas d'hypothèses sur les signaux correspondants aux différentes directions de l'espace. Dans le contexte de la production de musique en général, et de notre approche en particulier, une partie de ces signaux correspond aux sons directs des différentes sources constituant la scène sonore, et une autre partie de ces signaux correspond à la réverbération de la scène. Cette réverbération n'est pas forcément homogène : elle peut être constituée de plusieurs réverbérations ayant chacune des caractéristiques spécifiques, comme par exemple différents temps de réverbération, pour différents groupes de sources, ayant chacune une répartition spatiale différente. Cette spatialité de la réverbération est une des possibilités nouvelles offertes par ces nouveaux formats et elle ouvre de nouveaux champs esthétiques.

On peut considérer le format 11.1.4 comme un format 7.1.4 dont on a augmenté la définition dans la partie avant par l'ajout de sources supplémentaires WL, WR,

LC et RC. Le fait de matérialiser autant que possible la direction des sources par leur présence sur un canal (ou un objet) unique, et non pas comme une source fantôme entre deux canaux, permet d'une part, d'avoir une scène plus stable en multicanal (si l'auditeur se déplace, les sources fantômes se déplacent mais les sources réelles restent à leur place) et d'autre part, d'avoir des scènes binaurales plus définies en évitant de mélanger plusieurs HRTF pour une direction de source, ce qui est inévitable si un signal est présent sur deux canaux en

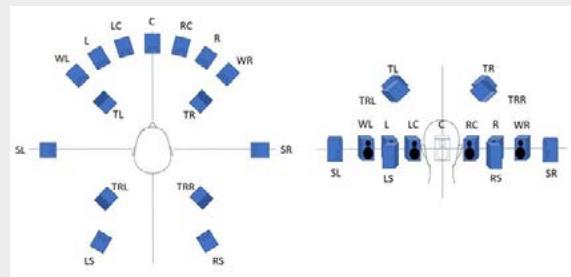


Figure 3 : Format 11.1.4

même temps. Il est donc nécessaire de disposer d'un certain nombre de canaux, ou d'objets, afin d'obtenir une bonne définition pour les sons directs, mais il n'est pas indispensable de tous les utiliser pour diffuser la réverbération de la scène.

Un des avantages des formats immersifs est de permettre une différenciation plus forte des directions spatiales de la réverbération. Dans une acoustique naturelle d'une salle de grand volume, la réverbération qui vient de l'avant n'est pas la même que celle qui vient des côtés, de l'arrière ou du haut, car le jeu des réflexions sur les parois de la salle est différent suivant la position des sources, leur directivité et l'acoustique même de la salle. On peut utiliser les canaux L, R, SL, SR, LS, RS, TL,

TR, TRL et TRR pour répartir la réverbération sur des directions occupant tout l'espace d'écoute de manière homogène, car on a ainsi 6 bases équiréparties de 60° dans le plan et 4 bases de 90° en élévation (les angles des canaux étant réglés comme sur la Figure 1).

Cette richesse de l'acoustique est très bien perçue en acoustique naturelle mais est limitée et altérée en stéréophonie à deux canaux. Il ne s'agit pas de prétendre que l'on peut rendre cette complexité telle qu'elle est à l'origine, car le multicanal implique un couplage avec la salle de diffusion qui va forcément modifier le rendu de la scène, mais il sera possible d'enrichir l'écoute en bénéficiant d'un plus grand nombre de ressources pour restituer cette complexité.

Chacun est libre de jouer ou non sur le naturel de la représentation, et les options de répartition de la réverbération dépendent bien sûr du type de musique que l'on produit, mais, encore une fois, la nouveauté réside dans la possibilité de différencier la provenance spatiale des sons.

## **Les systèmes de captation et de post-production**

Un des aspects fondamentaux de notre démarche est de rechercher une continuité entre les méthodes que l'on utilise pour la stéréophonie à deux canaux et celles que l'on peut utiliser en multicanal ou en mode objet. Il nous semble important par exemple de continuer à exploiter les mêmes microphones que nous utilisons pour la stéréophonie à deux canaux car ils constituent pour nous un des éléments centraux de l'esthétique que nous cherchons à développer, quel que soit le format pour lequel on travaille.

Les méthodes sont un sujet sensible car elles représentent l'approche personnelle de chacun et chacune. Le fait de partager notre expérience permet de respecter ces approches et peut donner à réfléchir sur l'évolution de ses méthodes personnelles pour aborder ces formats ou offrir des éléments de comparaison si on les pratique déjà. De grandes lignes peuvent être partagées, qui dépendent notamment des styles musicaux enregistrés. Concrètement, chacun des objets de la scène est capté par un ou plusieurs microphones (cas du piano par exemple) et la méthode de base de captation des objets est donc identique à celle des microphones d'appoint en stéréophonie. Il est même usuel, pour nous, de construire une scène de jazz uniquement à partir de microphones d'appoint et de réverbérations artificielles.

On trouve des réverbérations artificielles multicanales, comme le SPAT, que nous utilisons beaucoup. On peut construire l'espace sonore à partir d'un ensemble de réverbérations 4 canaux ou 2 canaux réparties dans l'espace. La société AudioEase propose avec l'Altverb des réverbérations 4 canaux à convolution ; TC Electronic, avec son System 6000, propose des moteurs de réverbérations multicanales ; et on peut également citer la DRE 777 de Sony, dont la production a été abandonnée, mais dont les réponses impulsionnelles sont excellentes et toujours utilisées. Pour les réverbérations à deux canaux, tous les systèmes disponibles sur le marché sont utilisables. On peut construire une réverbération multicanale avec plusieurs réverbérations à deux canaux possédant des critères communs (même temps de réverbération, par exemple). Toutes les techniques des musiques actuelles amplifiées sont utilisables, il suffit donc juste d'ajouter à ces traitements des choix plus larges de répartition spatiale qu'en stéréophonie à deux canaux. ►

Ces éléments ont en commun d'être tous suffisamment décorrélés et donc de pouvoir être additionnés sans problèmes de phase. La constitution de la scène va dépendre du positionnement spatial de chacun de ces éléments et la réduction stéréophonique peut très bien correspondre à ce que l'on aurait choisi en stéréophonie à deux canaux. Les réverbérations seules peuvent être immersives en multicanal et redevenir des réverbérations stéréophoniques à deux canaux lors de la réduction. L'emploi des techniques binaurales et transaurales nous permet de définir des scènes sur plus ou moins  $45^\circ$ , plutôt que plus ou moins  $30^\circ$  pour les sons directs, afin de les rendre plus immersives. En transaural, la scène sonore sera plus large que le cadre des enceintes, et le format binaural, lorsqu'il est écouté au casque, permettra une meilleure externalisation des sources.

On peut adjoindre à ces éléments des ambiances (captation de la réverbération de la salle) ou des ailes (captations latérales de la scène sonore et captation de la réverbération de la salle) qui elles aussi seront décorrélées. Il existe de nombreux choix et de nombreuses nuances pour capter les ambiances, en fonction de la directivité et du placement des microphones. On peut, par exemple, utiliser des ambiances omnidirectionnelles situées au-delà de la distance critique, placées, lors du mixage, à plus ou moins  $90^\circ$  dans la scène, et des ambiances directives (cardioïdes ou autres) orientées vers le fond de la salle, placées à plus ou moins  $150^\circ$ . Pour les canaux d'élévation, on peut utiliser des ambiances diffuses si l'acoustique de la salle est intéressante, ou utiliser des réverbérations artificielles si l'on souhaite modifier l'impression acoustique de l'auditeur. On peut bien évidemment mélanger ambiance naturelle et réverbération artificielle sur un même canal.

Si l'on place ces réverbérations, naturelles ou artificielles, avec des gains similaires, dans des directions

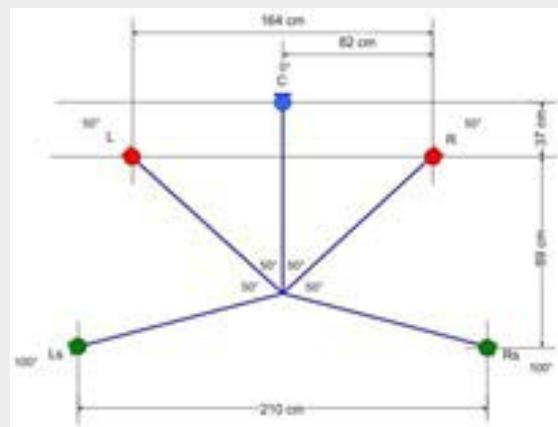


Figure 4 : Exemple Antenne pour un 5.1 – 5 microphones DPA 4041

régulièrement réparties de l'espace (dans notre cas  $\pm 30^\circ$ ,  $\pm 90^\circ$ ,  $\pm 150^\circ$  dans le plan horizontal, et  $\pm 45^\circ$ ,  $\pm 135^\circ$  à  $45^\circ$  d'élévation), on obtient une réverbération immersive. La partie la plus délicate va concerner les secteurs où il y a aussi du son direct (traditionnellement l'avant-scène).

Une première solution consiste à associer des appoints et des ambiances avant (par exemple des microphones cardioïdes orientés vers la scène, mais distants entre eux). La localisation sera alors donnée par la direction de l'appoint. On peut aussi souhaiter capter une source, ou un ensemble de sources, d'un point plus éloigné, d'une part, parce que l'on ne souhaite pas placer d'appoints sur chacune des sources (cas d'un orchestre symphonique par exemple) et d'autre part, lorsqu'un éloignement du microphone est préférable pour une question de qualité de timbre. Les sources musicales n'ont pas la même qualité de son direct à différentes distances car les points de rayonnement s'y additionnent différemment. La qualité de leurs timbres (indépendamment de la réverbération qui y participe) est optimisée pour une certaine distance de perception. C'est donc à chacun de décider, en fonction

de l'ensemble de ses contraintes, de ses conceptions et de ses goûts propres, s'il préfère, selon les cas, utiliser un appoint ou enregistrer la source par un système placé plus loin et captant une partie de la scène sonore, voire toute la scène sonore.

La première différence d'un système frontal, utilisé en multicanal, comparé à un système principal, utilisé en stéréophonie à deux canaux, est qu'il ne capte qu'une partie de la réverbération, les autres directions de celle-ci devant être réparties sur d'autres directions de la scène afin d'avoir un résultat immersif. Il devra donc être placé plus près ou être constitué de microphones plus directifs. La deuxième différence est qu'il devra résister aux différentes réductions (*downmixs*) et que la sommation des éléments du système ne devra pas poser de problème de phase. La troisième est que, si l'on veut utiliser une version binaurale du mixage, on a tout intérêt à matérialiser chacun des signaux que l'on veut ensuite traiter par des HRTF pour avoir une meilleure définition de la scène sonore. L'interpolation entre les HRTF  $\pm 30^\circ$  de deux signaux issus d'un couple stéréophonique ne conduira pas au même résultat que la perception de sources fantômes à partir de deux enceintes situées à  $\pm 30^\circ$  diffusant ces signaux. Dans le deuxième cas, le cerveau dispose des informations issues des mouvements de la tête et travaille avec ses propres informations correspondant aux HRTF, ce qui n'est généralement pas réalisé dans le premier cas.

Les systèmes que nous avons adoptés et qui répondent aux critères précédents sont des antennes acoustiques, dont la théorie et l'application aux systèmes multicanaux a été décrite de manière exhaustive par Mike Williams (2000). Les antennes acoustiques destinées à l'enregistrement multicanal sont initialement prévues pour capter tout l'espace et nous avons, en première approche, utilisé des antennes à 5 canaux omnidirectionnelles pour les systèmes 5.1.

Dans la mesure où une antenne destinée à un grand nombre de canaux est très encombrante, on peut aussi remplacer les parties de l'antenne correspondant aux directions de l'espace où il n'y a pas de sources sonores directes par des ambiances.

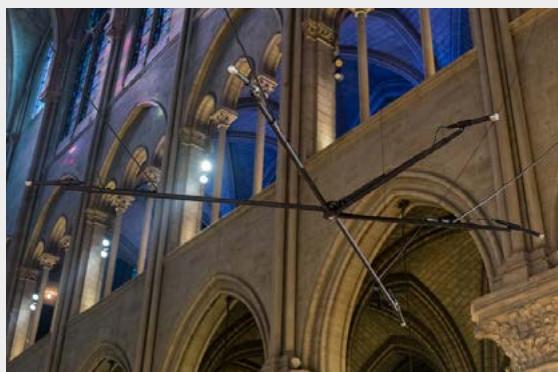


Figure 5 : Installation à Notre-Dame de Paris

Une autre de nos influences, importante, vient des systèmes que Kimio Hamazaki a développé pour les systèmes de restitution multicanaux en 22.2 (Hamazaki, 2006). Ces systèmes sont constitués d'un mélange d'antennes et de systèmes d'ambiance et nous avons travaillé en ce sens en essayant de rendre ces systèmes plus discrets visuellement. Les ambiances peuvent être placées par exemple sur les côtés afin de ne pas être trop visibles par le public ou par les caméras.

On peut choisir une antenne constituée de microphones de directivités différentes afin que la réverbération captée par l'antenne soit moins présente sur les canaux centraux que sur les côtés, ce qui permet de contrôler sa répartition spatiale.

Ces possibilités permettent de construire dans chacune des situations des systèmes différents, en fonction du ►

but sonore recherché et des contraintes de production (présence de la vidéo ou du public et donc contraintes visuelles, contraintes dues au temps d'installation, à des limitations techniques, etc.). L'important est de définir un signal pour chacune des directions de la scène. Mais ce signal peut être défini par un appoint, une ambiance, un élément d'antenne ou de la réverbération artificielle, ce qui offre un grand nombre de possibilités.

Ces systèmes ont des fondements très classiques, dans la possibilité de les réaliser. Nous gardons d'une part, la maîtrise de la microphonie. D'autre part, nous pouvons capter le champ acoustique de près ou de loin en utilisant un système principal ou des appoints et des ambiances, comme nous le faisons avant. Le changement consiste uniquement à réaliser la direction du son différemment, en définissant quel signal correspond à toutes les directions d'une scène, plutôt qu'en définissant quelles relations (différences d'intensité et/ou différences de temps) il y a entre les sons présents sur deux canaux à la fois. Ceci ne définit pas une esthétique.

L'esthétique commence dans la mise en œuvre de ces moyens pour occuper l'espace. En partant du modèle cinématographique, on peut observer que cette esthétique peut être très différente suivant les films. Dans *Birdman* (2014) de A.G. Iñárritu, Michael Keaton fait le tour de la salle via les enceintes *surround*, ce qui est loin d'être habituel. Dans *The French Dispatch* (2020) de Wes Anderson, la musique est en mono à droite et les dialogues au centre, ce qui crée un parti-pris très fort. Chaque film a en fait ses spécificités et la manière de répartir le son dans l'espace fait partie de son style, comme les choix de lumière, le format d'image, etc. Le champ d'expérimentation est à notre avis tout aussi ouvert en musique, avec des possibilités de spatialisation beaucoup plus importantes qu'auparavant.

## Bibliographie

> Blauert, J. (1997), *Spatial hearing*, London, The MIT Press.

> HAMASAKI, K., HIYAMA, K. (2006), *Development of a 22.2 Multichannel Sound System*, Broadcast Technology n°25, <https://www.nhk.or.jp/str/publica/bt/en/fe0025-2.pdf>

> Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B. F. G., de Boishéraud, C. (2017), *Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis*, J. Acoust. Soc. Am., vol. 141, pp. 3678–3688.

> Messonnier, J.-C., Lyzwa, J.-M., Devallez, D., de Boishéraud, C. (2016), *Object-based audio recording methods*, AES 140<sup>th</sup> convention.

> Messonnier, J.-C. & Moraud, A. (2011), *Auditory distance perception: criteria and listening room*, AES 130<sup>th</sup> convention, London.

> Moore, B. C. J. (1989), *An introduction to the psychology of hearing*, Academic press,

> EBU TECH 3364 *Audio Definition Model. Metadata specification*, Genève, édit. 2014.

> Plenge, G., & Theile, G. (1986), *Überlegungen zur Leistungsfähigkeit verschiedener stereofoner Verfahren*. Institut für Rundfunktechnik München.

> Williams, M. (2004), *Microphone Arrays for stereo and Multichannel sound recording*, Vol II, Editrice Il Rostro.

## JEAN-CHRISTOPHE MESSONNIER

Après une formation d'ingénieur du son à l'ENS Louis-Lumière et parallèlement à une formation d'ingénieur en acoustique au CNAM, terminée en 1994, Jean-Christophe Messonnier est entré au service audiovisuel du Conservatoire de Paris (CNSMDP) en 1988. Dans cette longue pratique au service du Conservatoire, il a cherché à comprendre comment on pouvait formaliser la transmission du son et comment l'acoustique pouvait aider à comprendre certains des problèmes reliés à cette transmission. C'est ce qu'il enseigne à la Formation Supérieure aux Métiers du Son du Conservatoire depuis 1992.

*After training as a sound engineer at the ENS Louis Lumière while studying acoustics engineering at the Conservatoire National des Arts et Métiers, completed in 1994, Jean-Christophe Messonnier joined the audiovisual department of the CNSMDP in 1988. In this long practice serving the Conservatoire he tried to understand how to formalize the transmission of sound and how acoustics could help to understand some of the problems related to this transmission. This is what he has been teaching at the Formation Supérieure aux Métiers du Son of the Conservatoire since 1992.*

## JEAN-MARC LYZWA

Jean-Marc Lyzwa exerce la fonction d'ingénieur du son depuis 1990 au service audiovisuel du Conservatoire de Paris et enseigne à la Formation Supérieure aux Métiers du Son les techniques de prise de son et de post-production. Investi depuis 1995 dans la restitution et la création de l'espace sonore en développant des techniques de captation et de post-production 3D en stéréophonie à deux canaux et en multicanal, il a collaboré avec l'équipe espaces acoustiques et cognitifs de l'Ircam à la réalisation d'outils de traitement binaural / transaural, spécifiquement adaptés à la production musicale, notamment avec Transpan. Il participe au projet de recherche BiLi (Binaural Listening) de 2013 à 2017 et au projet de recherche PHEND (The Past Has Ears at Notre-Dame) depuis 2021.

*Jean-Marc Lyzwa has been working as a sound engineer since 1990 at the audiovisual department of the Conservatoire de Paris (CNSMDP) and teaches the Advanced Training in sound recording and post-production techniques at the Formation Supérieure aux Métiers du Son. Involved since 1995 in the restitution and the creation of the sound space by developing 3D capture and post-production techniques in two-channel stereophony and in multichannel, he has collaborated with the team Acoustic and Cognitive Spaces (EAC) of Ircam for the realization of processing tools for binaural / transaural treatment, specifically adapted to music production, notably with Transpan. He participates in the research project BiLi (Binaural Listening) from 2013 to 2017 and in the research project PHEND (The Past Has Ears at Notre-Dame) since 2021.*

## ALEXIS LING

Alexis Ling est responsable du service audiovisuel du Conservatoire de Paris depuis 2020, mettant à profit les connaissances qu'il a acquises lors de son parcours : licence EEA à la faculté d'Orsay, Maîtrise de Sciences et Techniques image et son à l'université de Brest, Formation Supérieure aux Métiers du Son au Conservatoire de Paris, puis activité d'ingénieur du son en freelance dans le secteur de la musique classique, puis en broadcast pour le groupe Canal+. Au sein du Conservatoire, il enseigne et mène une réflexion sur la façon dont les outils et les activités audiovisuelles, en production et en recherche, peuvent accompagner le travail pédagogique et valoriser le travail des étudiants et des étudiantes.

*Alexis Ling has been in charge of the audiovisual department of the Conservatoire de Paris (CNSMDP) since 2020, using the knowledge he acquired during his career: EEA degree at the Orsay University, Master's degree in Sciences and Techniques of Image and Sound at the University of Brest, Advanced Training in Sound Design at the Conservatoire de Paris, and then activity of sound engineer in freelance in the classical music sector and in broadcast for the Canal+ group. Within the Conservatoire, he teaches and leads a reflection on the way in which audiovisual tools and activities, in production and research, can accompany the pedagogical work and enhance the value of students' work.*

## Raconter en son binaural : entretien avec Léa Chevrier, ingénieure du son et créatrice, le 28 novembre 2021

| Alan Blum

**Alan Blum : Bonjour Léa, pourrais-tu nous présenter brièvement ton parcours et ton secteur d'activité professionnelle actuel ?**

Léa Chevrier : J'ai fait l'école Louis-Lumière, en section son. J'en suis sortie en 2015 et depuis je travaille surtout avec du son binaural : en partie sur du podcast - fiction pour enfants et documentaire pour adultes - et en partie sur de la muséographie - parcours sonore ou son dans les musées pour des expos. Je fais d'autres choses, du côté du spectacle vivant, mais qui n'ont rien à voir avec le son binaural.

**AB: C'est vrai que nous allons nous centrer sur le travail avec les techniques binaurales qui font partie intégrante de tes outils et constituent presque ta carte de visite. Pourquoi avoir choisi cette technologie plutôt qu'une autre ?**

LC : Ce que j'aime bien dans le binaural, c'est que c'est hyper simple en fait. Je ne suis pas très technophile, donc ça ne me branche pas trop d'avoir les supers derniers logiciels et cinquante amplis, cinquante enceintes. Du coup je trouve le binaural intéressant parce qu'il faut juste, soit une tête, qui coûte cher, mais qu'on peut aussi

fabriquer avec trois fois rien et on peut aussi mettre des micros dans ses propres oreilles et aller faire des prises de son, soit mettre un plug-in sur un logiciel<sup>1</sup>. Ce n'est pas très compliqué et je trouve ça chouette qu'on puisse le faire de façon assez démocratique ; et surtout que les gens puissent écouter facilement le résultat. Bien sûr, les gens n'ont pas forcément un super casque, mais à peu près tout le monde peut écouter le mixage que tu as fais, alors que, quand tu fais un mixage en Dolby Atmos ou en un autre format multicanal, personne n'a le bon système son correspondant installé chez soi, personne n'a même une stéréo correcte installée chez soi. Avec le binaural je trouve que c'est moins déprimant quand tu fais du mix, car tu sais qu'il sera *écouté* à peu près dans des conditions similaires à ce que tu as fait.

**AB : Et du point de vue de l'auditeur justement, dans ton travail, selon toi, qu'est-ce qui crée l'illusion immersive avec ces procédés ?**

LC : Tu veux dire techniquement ?

**AB : Pas forcément que techniquement. Je veux dire en quoi cette illusion marche-t-elle ? À quel endroit as-tu l'impression que cette technique, simple, mais qui est quand même une technique,**

<sup>1</sup> Il y a deux manières de créer un contenu binaural : à la prise de son, en captant directement la scène sonore en binaural, on parle alors souvent de « binaural natif », ou bien a posteriori, à l'aide de traitements, généralement logiciels et sur de l'audio ne contenant pas nécessairement d'information spatiale au départ, on parle alors de « synthèse binaurale ». Dans les deux cas le contenu binaural est destiné à être écouté au casque stéréophonique. En effet, chaque canal, gauche et droit, doit être délivré associé aux oreilles correspondantes de l'auditeur, sans autres interférences acoustiques qui nuiraient à la bonne interprétation des indices présents dans le signal - ces indices acoustiques étant justement ceux que notre système auditif exploite pour percevoir les sons dans l'espace. Ci-après nous explicitons un peu plus en détail les deux types de procédés :

> Binaural natif : consiste en un enregistrement réalisé directement en binaural, de la même manière que l'on peut capter un son en stéréo avec un couple microphonique. La prise de son peut s'effectuer à partir d'une tête artificielle ou bien d'une tête humaine équipée de capsules microphoniques au niveau des oreilles. Les microphones utilisés sont des capteurs de pression omnidirectionnels car étant les plus proches du fonctionnement de notre appareil auditif. La spatialisation est alors fixée au moment de la prise de son et les sources seront perçues par un auditeur en fonction de leur position relative à la captation.

> Binaural de synthèse : consiste à spatialiser un signal monophonique, comme on peut le faire avec le panoramique d'intensité en stéréo par exemple, à l'aide d'un outil logiciel, souvent un plugin, intégrant des filtres binauraux. Ces filtres sont appelés des HRTF, pour Head-Related Transfer Function, c'est-à-dire des « fonctions de transfert relatives à la tête ». Un couple de HRTF filtre ainsi le son des canaux gauche et droit de manière à restituer la « signature acoustique » du chemin d'un point de l'espace jusqu'aux oreilles d'un auditeur. À l'écoute, l'auditeur peut ainsi avoir l'impression que le son traité vient de ce point de l'espace, et on parle souvent alors de « source virtuelle ».

**fonctionne ? Et pourquoi dans le discours sonore, y a-t-il cette notion d'immersivité qu'on peut y associer ? Ou pas d'ailleurs ?**

**LC :** Si, quand même (rires) ! Je trouve que ce qui marche vraiment bien c'est la reconstruction des acoustiques par exemple, ou même juste la reconstruction des ambiances sonores. Il faut que ce soit un peu bruyant pour qu'on s'en rende compte, mais même sur des enregistrements comparatifs que j'ai pu faire, du métro par exemple, c'est beaucoup plus immersif en binaural qu'en stéréo, juste parce que c'est beaucoup plus large, et tu entends l'arrière. Même si tu n'entends pas toujours distinctement que c'est à l'arrière, tu as quand même une sensation d'immersion qui est beaucoup plus importante. Le binaural, les limites que ça a, c'est que ce n'est pas toujours précis, mais tu as cette sensation quand même très différente d'avec la stéréo : d'un coup, ça s'ouvre. Ça s'ouvre, et même si tu n'entends pas si le son est à 120 degrés ou à 130 degrés, c'est quand même plus ou moins à l'arrière.

Un autre aspect, c'est comme quand tu passes de la mono à la stéréo : ça démasque. Là c'est pareil, ça démasque parce que tu passes en 360 degrés. Et même si ce n'est pas très précis, ça élargit énormément et tu as quand même une sensation d'immersion beaucoup plus grande. Et en effet, je trouve que ça marche bien avec des endroits un peu bruyants, le métro, la ville, tout ça : ça démasque et ça s'ouvre ; ou dans des acoustiques qui sonnent, donc des églises, des châteaux, des lieux comme ça : tout à coup la réverbération tu l'entends, elle s'ouvre et c'est ça qui fait que tu es immergé... Sur les réverbs et les ambiances surtout je dirais.

**AB : Et alors, entre cette promesse technologique,**

**avec les exemples que tu donnes, et les besoins narratifs d'une œuvre, comment se fait l'accord pour toi ?**

**LC :** En fait, je trouve que c'est un peu une erreur quand les gens essayent d'écrire une narration avec l'idée que les choses soient très précises dans l'espace. Par exemple là, je bosse sur un parcours sonore à Vaux-le-Vicomte, et pour l'autrice qui écrit le texte c'est la première fois qu'elle faisait du binaural et on lui avait dit : « C'est du son 3D, il faut donner des choses dans l'espace, etc. » Et en fait, elle cherchait des choses trop précises qui ne marchaient pas. Par exemple, il fallait absolument qu'on sache que la fenêtre était dans tel angle, que la porte était là précisément, et ça ne marche pas. Déjà, parce qu'il n'y a pas de *head tracking*<sup>2</sup>, donc on ne sait pas comment le spectateur va se positionner et donc, que la fenêtre soit à 30 ou à 40 degrés on s'en fiche. Et de toute façon ça va être différent pour chaque personne. Mais par contre, ce qui va être intéressant dans la narration, c'est de jouer juste avec le fait que, quand même, il y a cette sensation d'immersion qui marche à peu près pour tout le monde. Elle sera aussi différente pour tout le monde. Il faut le prendre en compte, savoir que ce n'est pas forcément précis, mais, en termes de narration, se débrouiller pour que cette sensation rentre dans le fil narratif. Alors, je ne sais pas, par exemple on a fait un parcours où il y avait un écureuil volant qui se balade autour de nous. Souvent, on l'entend juste, il est là, il se promène un peu autour de nous et on se fiche de savoir exactement où il est. Et si chaque personne l'entend à un endroit qui est un peu différent, ce n'est pas grave. Par contre, d'un coup il vit cet écureuil, il est autour de nous. Alors que si c'était en stéréo, on n'aurait pas du tout l'impression qu'il est avec nous. Là, on a une externalisation<sup>3</sup> qui joue, qui fait qu'on a l'impression qu'il est autour de nous et qu'il se ►

<sup>2</sup>

Head tracking : système de suivi des mouvements de la tête de l'auditeur à l'aide d'un capteur, dans le but de les compenser par rotation inverse de la scène sonore diffusée. De cette manière les sources perçues paraissent rester positionnées au même endroit de l'espace, quelque soit l'orientation de l'auditeur avec son casque sur les oreilles.

<sup>3</sup>

Externalisation - relatif à l'écoute au casque en binaural : fait de ressentir les sources diffusées comme émanant de l'extérieur du casque, là où en stéréo classique les sons sont perçus de manière intra crânienne, comme un espace s'étalant entre les deux oreilles

balade en volant, et ça, c'est chouette. C'est typiquement ce genre de petites astuces qui marche.

Dans une autre fiction, il y avait un fantôme : c'est pareil, il se balade autour de toi et puis il flotte, et ça, tu l'entends. Et il n'y a pas besoin d'être précis spatialement. Par contre, tu as la sensation d'externalisation et d'immersion qui est intéressante à utiliser. Après, je trouve que c'est un peu une erreur de se dire : on va entendre la cheminée pile à tel endroit, la fenêtre pile à tel endroit. « Pile à tel endroit » ce n'est pas vrai ; ce n'est pas vrai parce qu'il y a trop de problèmes d'individualisation, c'est trop galère. Il faut plutôt miser sur une sensation plus globale et écrire des scénarios qui utilisent le fait qu'on a besoin d'être en 3D. C'est la même chose pour le cinéma 3D : si c'est juste pour « mettre du 3D », ça ne sert à rien. Il faut que la narration s'empare de ça. Mettre n'importe quel film en 3D ce n'est pas intéressant ; il faut utiliser cette possibilité comme quelque chose de créatif au cœur de la narration. Cette idée du fantôme ou de *l'écureuil volant, en stéréo ça ne marcherait pas*, alors que là, en binaural, *ça fonctionne*. Tout ce qui est parcours sonore marche bien aussi, parce que de base, on enregistre à chaque fois dans les pièces où les visiteurs vont écouter et donc, on reconstitue la vraie acoustique de là où sont les gens. De fait, ce concept d'un parcours sonore marche mieux en binaural qu'en stéréo, simplement car tu es dans la pièce, tu entends l'acoustique comme elle est dans la « vraie vie » et du coup, ton cerveau fait le lien avec comment la pièce est visuellement : est-ce qu'il y a un haut plafond ? Est-ce qu'on est dans une pièce qui réverbère beaucoup ?... Et ça marche beaucoup mieux en binaural. Comme tu externalises, il n'y a pas ce truc un peu bizarre de la stéréo où tu entends dans ta tête, qui fait que ce n'est pas évident de poser les sons sur ce que tu vois. Là, tu externalises, donc ça marche bien. Et quand c'est accompagné avec du visuel, ça marche encore mieux, parce que ton cerveau « colle » les deux infos.

**AB : Et peut-être qu'avec le binaural on a moins la sensation que le casque nous coupe de l'environnement dans lequel on est ?**

**LC :** Oui c'est ça, ça te coupe moins. Encore plus si tu peux avoir un casque ouvert. Après malheureusement, c'est rarement possible parce que, dans tout ce qui est muséographie, il faut prendre en compte le fait qu'il y a du public et qu'il y a du bruit. Donc malheureusement, on fait la plupart du temps le choix d'un casque fermé ou semi-fermé. Souvent je milite pour avoir au moins un semi-fermé parce que, quand même, ça change tout pour le binaural. Mais c'est vrai qu'il y a des problématiques de bruit, de groupes, de visites, de guide, donc là, il vaut mieux un casque fermé. Mais, plus on a un casque ouvert, plus l'immersion et l'externalisation fonctionnent. Reste que c'est important d'essayer de l'intégrer dans la narration parce que, s'il s'agit simplement d'enregistrer n'importe quoi en binaural, si ça n'a pas été pensé en amont, ce n'est pas très intéressant, et c'est trop souvent ça !

**AB : Tu disais tout à l'heure que l'autrice avec qui tu travaillais avait peut-être une pensée trop visuelle, à vouloir placer les choses très géométriquement : plus globalement, est-ce que le fait d'utiliser un « son immersif » et des technologies pas forcément connues quand c'est toi qui les amènes, cela change quelque chose à la collaboration avec tes interlocuteurs sur un projet de création ?**

**LC :** Oui. À chaque fois, c'est bien de pouvoir en parler avant, c'est bien de pouvoir écouter des choses. Ce n'est pas toujours possible, mais c'est bien en effet de pouvoir faire une mini-initiation au binaural dès le début du projet, justement pour y penser dès l'écriture, pour essayer d'intégrer ça tout de suite dans le récit, et de ne pas faire cette erreur d'être dans quelque chose

de trop visuel. Sauf si on est en *head tracking*. Là, c'est autre chose. Mais si on est en binaural natif, c'est bien de savoir un peu quels sont les limites et les intérêts de cette technique et de réfléchir à ce qu'on peut faire avec : qu'est ce qu'on peut faire avec cette histoire d'immersion ? Qu'est-ce qu'on peut faire avec cette histoire d'acoustiques mieux rendues ? Qu'est-ce qu'on peut faire avec la proximité aussi, les approches qui fonctionnent beaucoup mieux, les chuchotis à l'oreille ? Prendre un peu tous les trucs qui fonctionnent et essayer de les intégrer dans la narration pour que ce soit intéressant. Parce que parfois, des choses peuvent être un peu plates simplement parce qu'elles ont été écrites et pensées en stéréo, et puis on le fait en binaural, mais ce n'est pas si différent et incroyable que ça si on n'utilise pas les choses qui marchent bien, comme des chuchotis, des choses qui tournent, des acoustiques intéressantes... Encore une fois, l'exemple de l'écureuil volant qui te tourne autour, qui s'approche ou s'éloigne, et qui parfois vient te chuchoter à l'oreille, ça marche très bien et ça, c'est étonnant pour le visiteur qui n'a pas l'habitude d'écouter du son binaural. En plus, il faut réfléchir en termes de timbre : il y a des choses qui marchent et d'autres moins.

En gros, tout ce qui est dans l'aigu, avec des transitoires, on le spatialise beaucoup mieux. Pour cet écureuil, on en avait fait un magique : il faisait un petit son de poussière d'étoiles et il avait des ailes de libellule dans la fiction. Donc, il fait un petit vrombissement « vrrrrr », avec plein de transitoires, plus un son un peu dans l'aigu avec la poussière d'étoiles ; tout ça se spatialise bien. De la même manière, tu peux faire voler un moustique, ça marche super bien. Alors que, par exemple, pour le fantôme, au début j'avais mis des choses un peu plus graves et ça ne fonctionnait pas très bien parce qu'on ne les spatialisait pas. Il faut donc penser à quel type de son, quel timbre, et que faire avec pour obtenir un résultat probant.

**AB : Et as-tu aussi parfois le rôle de la réalisation ?**

**LC :** C'est rare. En général, je suis en position d'ingénieure du son, mais comme souvent les réalisateurs ne connaissent pas le binaural, il y a une espèce de flou là-dessus, et du coup je prends un peu plus de place dans la réalisation. Il peut s'instaurer une sorte de « conseil binaural » où je vois les réalisateurs en amont, quand c'est possible bien sûr, et on discute de binaural, on va écouter des sons avant d'écrire ; mais ce n'est pas toujours le cas. Et après, je suis « pseudo réal » parce que, sur le tournage, le réal va surtout s'occuper de la direction des comédiens et on aura parlé un peu de la spatialisation en amont ; mais c'est quand même plutôt moi qui vais dire : « He bien là, le personnage va arriver de tel endroit, il va se positionner de telle manière par rapport au micro parce que c'est intéressant comme ci ou comme ça ». Le réal n'aurait pas forcément de réflexe de comment placer les comédiens par rapport à la tête binaurale, donc les rôles sont un peu plus flous. Là, je peux suggérer : « Ce serait bien de le faire en chuchoté », par exemple, je propose des choses qui influent un peu plus sur la mise en scène.

**AB : C'est une collaboration intéressante du coup ?**

**LC :** Oui c'est hyper intéressant. Alors ça dépend avec qui : il y a des gens qui ne sont pas très ouverts à ça. Mais si on tombe sur quelqu'un qui accroche, il y a des gens qui captent très vite ce qu'on peut faire avec du son binaural. Là, c'est chouette.

**AB : Et au regard de ton parcours, quelles expériences liées à l'emploi du son immersif ont été les plus marquantes ? Que ce soit positif ou négatif, d'ailleurs.**

**LC :** Je crois qu'une des expériences des plus marquantes, ► c'est ce parcours de visite pour les enfants dans le

château de Vaux-le-Vicomte, avec cet écureuil. Du fait d'être en fiction, et en fiction pour enfant qui plus est, on est en général plus libre sur l'écriture. C'est plus facile je trouve que lorsque c'est une création pour adulte, où on essaye de faire quelque chose de réaliste, et moins inventif en général. Alors que pour les enfants, on met plus facilement des personnages fantastiques, des *éléments imaginaires*, et ça permet de plus jouer avec le binaural. Sur ce projet, en effet, c'était chouette cette idée qu'a eu la réalisatrice de mettre un écureuil volant. Au moment du tournage, j'avais placé la tête binaurale au niveau du visage du comédien principal et une comédienne interprétait le petit écureuil qui se baladait autour de l'auditeur. Donc, il fallait qu'on la mette sur un tabouret pour qu'elle puisse être en hauteur, puisque l'écureuil était censé voler. Il fallait qu'on l'entende en haut et on ne pouvait pas toujours baisser la tête binaurale. On se retrouvait dans Vaux-le-Vicomte la nuit : il faut imaginer qu'il faisait cinq degrés parce que j'avais fait éteindre tous les chauffages. J'avais fait éteindre les lumières aussi car ça faisait des *buzz*. Donc, on était avec des lampes frontales dans le château avec la comédienne sur son tabouret, et ce qui était compliqué c'était qu'il fallait lui faire faire des mouvements sans qu'on entende du tout son corps, parce qu'on voulait juste sa voix pour ensuite refaire en *sound design* un petit son d'écureuil qui vole. Elle était entièrement en jogging et en polaire qui ne faisaient pas de bruit et elle faisait des mouvements sans faire de pas. Et Éléonore Mallo, qui faisait le bruitage sur ce parcours, mémorisait tous les mouvements que faisait la comédienne autour de la tête et les reproduisait en bruitage juste après. Il y avait des bruitages de « *woosh* » avec un bâton et des bruitages de *petite fourrure d'animal avec des poils pour avoir une présence*. Et derrière, on faisait encore en post-prod une couche de *sound design* de « *poussière d'étoile et petit vrombissement d'ailes* » et on « *collait* » tout ça. Et ça marchait bien ! C'était rigolo et je n'avais jamais fait ça.

**AB : Pour ce design en post production il s'agissait donc de synthèse binaurale et plus de binaural natif ?**

**LC :** Synthèse oui, et je refaisais le mouvement en logiciel en essayant de suivre ce qui avait été fait. Là, c'était un peu compliqué parce qu'il fallait que la voix et le bruitage soient parfaitement au même endroit, sachant qu'on pouvait rarement les faire en même temps ; parce qu'en plus, le bruitage *était très fin* et j'avais besoin de monter les gains par rapport à la voix. Il fallait donc le faire en deux prises en refaisant exactement, ou la mieux possible, la même spatialisation. Après ça, je faisais le *sound design*, et il fallait que je refasse encore la même spatialisation avec le logiciel. Donc, c'était un peu difficile, mais au final ça *marche* bien. C'était assez ambitieux pour un projet où on n'avait pas spécialement d'argent pour faire du *sound design*, ni spécialement de temps en plus pour faire un bruitage en studio. *C'était juste la réal qui avait* cette idée de : « Ok je vais créer un personnage, ça va être un écureuil volant », et c'est parti quoi. Et après, en truc chiant (rires)... je ne sais pas !

**AB : Ce n'est pas obligé ! Ou alors ça pourrait être une expérience où l'immersion n'a peut-être pas fonctionné, car non prise en compte dans la narration justement ?**

**LC :** Oui si, je pense à un truc : c'était du documentaire pour le coup et c'était un parcours dans les rues de Paris qui racontait des histoires pendant la Shoah. Le principe était de raconter ce qui *c'était passé dans des rues*, que le visiteur se balade dans des rues et qu'on lui raconte que, dans telle rue, dans tel immeuble, il y a telle personne qui a été déportée, qu'on raconte des histoires intimes des gens, en se baladant. Et l'idée était qu'on fasse l'enregistrement dans la rue, ce qui est quand même très compliqué, et que l'historienne se balade avec toi, qu'on ait la sensation que l'historienne se balade avec

toi et te raconte l'histoire. La réalisatrice voulait du coup qu'on entende tout le temps l'historienne sur notre droite, comme si la personne était à côté de toi et te racontait l'histoire. Ça typiquement, *ça ne marche pas*, parce qu'en fait, à entendre tout le temps une voix sur la droite, on croit juste qu'il y a un problème. Tu ne te dis pas : « Il y a quelqu'un qui est sur ma droite et qui m'accompagne » ; tu te dis juste : « *il y a un bug et mon oreillette gauche ne marche pas, mon casque ne marche qu'à droite* ». Du coup, je leur ai dit : « Non, entendre toujours quelque chose sur la droite, *ça va être pénible*, on n'est pas habitué à ça et ça ne fonctionne pas. Ce n'est pas comme dans la rue car dans la rue, quand quelqu'un nous parle, on a tendance à tourner notre visage vers cette personne. Donc au final, on l'entend en face, on a rarement quelqu'un qui nous parle à droite tout le temps. J'ai donc dit qu'il fallait qu'on entende la voix en face ; *même si on veut avoir la sensation que cette personne est avec nous, il faut qu'on l'ait en face*. Mais c'est resté hyper compliqué à faire, car elle ne voulait quand même pas faire les prises en studio, parce que ce n'était pas une comédienne, mais vraiment l'historienne qui parlait. Et c'est vrai qu'elle était bien meilleure, en diction, en jeu, quand elle était en situation, quand elle voyait vraiment la fenêtre, l'immeuble, le magasin, que quand elle était en studio et qu'on lui demandait de lire le texte où c'était plat et ça ne marchait pas du tout. On a essayé et ça ne marchait pas, donc on l'a vraiment fait dans la rue. Mais c'était un casse-tête parce qu'elle marchait et moi, j'enregistrais avec ma propre tête : il fallait que je recule pour être en face d'elle. C'était assez casse-gueule et, en plus, on capte tous les bruits, des voitures et tout, et il faut être raccord quand on fait les coupes, donc c'était une vraie galère.

Au final, ça ne marche pas trop mal, mais c'était vraiment hyper compliqué à faire et pour le coup, pour une production, je pense que c'est difficile budgétairement.

La mise en place était complexe par rapport à ce qu'on voulait faire. Avec le recul, je ne sais pas comment on pourrait faire mieux si ce n'est peut-être de faire jouer l'historienne par une comédienne qui nous fait croire qu'elle est l'historienne ; mais comme c'était un documentaire, ils voulaient vraiment que ce soit la réalité, et une non-professionnelle a dû mal à *vous faire sentir qu'elle est dans la rue et qu'elle voit les magasins si elle n'y est pas. Mais enregistrer dans la rue c'est compliqué*, particulièrement en binaural où c'est assez précis malgré tout spatialement. Pour faire un raccord avec une moto qui passe par exemple, c'est encore plus compliqué qu'en stéréo de faire passer la coupe. J'ai dû faire beaucoup d'enregistrement de rue, « *à blanc* », pour avoir un bus qui passe, une moto qui passe, et pouvoir faire les raccords. Et au montage je me suis bien pris la tête pour que... ça marche finalement ! Mais *c'était fastidieux*. Et dans ce cas, l'apport du binaural au final, je ne sais pas... Alors si, quand même, parce que l'immersion dans la rue c'est chouette et puis, ça permet de démasquer la voix de l'ambiance de rue. C'est moins désagréable qu'il y ait tout le temps des bruits de rue que si c'était en stéréo, je pense ; et en même temps, c'est vraiment compliqué en rapport à l'objectif de la narration : nous raconter des histoires de la Shoah, au final c'est ça qui était important. Est-ce que c'était vraiment un plus ce dispositif mis en place ? Parce qu'on ne jouait pas tellement avec la spatialisation au final, c'était juste pour se sentir immergé et pour que, quand on écoute dans la rue, on entende ces sons en binaural et qu'on ne sache plus si la voiture qui passe est dans la réalité ou dans l'enregistrement. Mais au final, on ne joue pas avec la spatialisation, il n'y a pas de sons qui tournent, on ne joue pas avec des chuchotis, c'est juste la sensation d'immersion, et du coup c'est quand même très compliqué pour relativement peu de chose, je pense. Dans ce cas typiquement, je ne sais pas si c'était très justifié de le faire de cette manière avec le recul. ►

**AB : Ca illustre un peu l'attente d'un « réalisme », dans ce cas de quelqu'un qui parlerait à côté de nous, alors que ça reste quand même une prise de son.**

**LC :** Oui c'est ça. Ca reste une prise de son, avec quand même des codes, auxquels on est habitué ou pas. Bon, au final ça a fonctionné, c'est quand même externalisé, l'apport se situe là. Ensuite, les traductions en anglais ont été faites. J'ai gardé des ambiances en binaural, des sons de pas que j'ai laissés : *ça permet de ne pas être coupé du lieu dans lequel on est.* Mais là, on enregistrerait la voix en studio, et même si ne je suis pas « pro studio », c'était quand même *plus simple !* Et particulièrement pour le montage, parce que, pour du documentaire, monter avec des ambiances de rue en binaural ce n'est pas évident !

**AB : Et, sur le travail de la spatialisation sonore, est-ce qu'il y a d'autres artistes ou ingénieurs du son qui t'ont influencée ?**

**LC :** Forcément. Il y a plusieurs créations sur Arte Radio. Il y a une qui s'appelle *Mutt Dogs*<sup>4</sup> et ça, c'est génial : ils ont mis des micros dans les oreilles d'un chien et fait toute une création sonore comme ça. Et c'est fou parce que, je ne sais pas à quoi à cela tient, mais on a vraiment la sensation qu'on est proche du sol. Est-ce que c'est dû aux réflexions du sol, je ne sais pas. Et est-ce que le binaural apporte quelque chose parce qu'on entend la hauteur des sons et que, du coup, tous les sons sont un peu hauts ? Il faudrait faire un test comparatif avec de la stéréo. En tout cas, c'est génial, car on sent ça. Et puis, il fait sa vie de chien, il va voir d'autres chiens dans la rue, on entend d'autres aboiements... Cela aurait pu marcher aussi en stéréo ou en mono, mais là, ça apporte vraiment quelque chose, on est vraiment dans la tête d'un chien, ça m'avait vraiment marqué.

**4**  
Mutt Dogs, de Félix Blume et Sara Lana :  
[https://www.arteradio.com/son/61659422/mutt\\_dogs](https://www.arteradio.com/son/61659422/mutt_dogs)

**5**  
Dans les dents, de Sara Monimart :  
[https://www.arteradio.com/son/616535/dans\\_les\\_dents](https://www.arteradio.com/son/616535/dans_les_dents)

**6**  
Virtual Barber Shop : <https://www.youtube.com/watch?v=IUDTlvagjJA>

**7**  
Cf. 100% flûtes binaurales, de Jean-Mathias Petri (flûtes) et Lucie Hardoin (réalisation son binaural) :  
<https://www.jeanmathias-petri.com/programmes/flutes-en-solo/>

Il y en a une autre sur Arte Radio, chez le dentiste<sup>5</sup> : ça, c'est horrible à écouter, mais c'est génial ! Et vraiment, le dentiste, ça marche très bien parce que ce sont des sons qui sont très proches de notre tête. C'est comme la séance chez le coiffeur hyper connue<sup>6</sup>, ça marche très bien. Ce sont des sons en proximité et en plus, pour le dentiste, ce sont que des sons aigus, des vibrations, avec des transitoires horribles et donc, ça marche hyper bien ! Mais c'est dur à écouter tellement on a l'impression qu'il y a vraiment quelqu'un qui vous met des outils dans la bouche. C'était malin, il y a des idées comme ça malines : les trucs qui sont proches du visage ça marche, donc le coiffeur, le dentiste ça marche, c'est intelligent. Après, il y a Lucie Hardoin qui fait pas mal de choses aussi qui sont chouettes. Elle avait fait une création plus en musique, une création immersive avec des flûtes<sup>7</sup>. C'est abstrait, mais elle a fait ça en binaural natif, donc elle enregistre des flûtes comme ça, dans des lieux, dans l'espace, et c'est vraiment chouette.

**AB : Une dernière question : comptes-tu poursuivre ton travail sur cette technologie ? Et comment vois-tu l'avenir des technologies immersives ?**

**LC :** Alors oui, je compte continuer (rires) et je continue actuellement. Comment je vois l'avenir des technologies immersives ? Je ne sais pas... Si je me mets dans un contexte politique, je trouve que le binaural peut faire sens en termes de problématiques énergétiques. Si jamais on a encore assez d'énergie pour faire des créations sonores, ça fait plus sens que de développer des choses avec cinquante enceintes, cinquante amplis : c'est beaucoup moins coûteux en énergie et en matériel. Après, j'imagine bien qu'on va quand même, ne serait-ce que pour le cinéma, continuer à développer plein de systèmes comme l'Atmos qui sont très coûteux, en nombre d'enceintes. Mais je crois que le binaural commence

à prendre pas mal de place aussi, dans des petites structures qui ne peuvent pas se permettre de faire des choses en multicanal avec beaucoup de matériel. Il y a de plus en plus de radios qui font du binaural en fait, même des radios associatives, des radios où il n'y a pas un centime. Moi, j'habite à Rennes et j'ai déjà discuté un peu avec les étudiants qui sont à C-Lab, qui est la Radio Campus<sup>8</sup> de là-bas, et il y en a que ça intéressait par exemple. Et en fait, j'ai fabriqué une tête pour quinze euros : avec une tête de coiffeur, tu lui remplaces les oreilles, tu lui colles des oreilles un peu mieux, tu achètes deux micros - mais tu n'es pas obligé d'acheter des DPA non plus - et hop, tu peux faire du binaural ! Et quasiment tout le monde à un casque aujourd'hui en France, ou en Europe en tout cas. C'est plus facilement *démocratique* et je trouve ça intéressant. Je pense que ça va continuer à se développer, surtout avec l'explosion des podcasts. De plus en plus de gens écoutent des podcasts au casque, et ça permet de faire écouter du binaural beaucoup plus facilement qu'il y a 50 ans quand tout le monde allumait son poste de radio. D'ailleurs, il y a beaucoup de boîtes de podcast qui m'appellent pour faire du binaural parce que c'est vendeur et que les gens écoutent avec un casque.

Malheureusement, ça ne va pas toujours avec une logique de créativité et de création sonore, mais dans une logique purement marketing. Je me rends bien compte que, par rapport à il y a six ans quand j'ai commencé à travailler, ça s'est développé très vite du côté des podcasts et on m'appelle souvent pour ça. Et les grosses boîtes commencent à *vouloir faire des créations* en binaural, malheureusement pour une logique commerciale... Donc, il y a Audible qui t'appelle, il y a Spotify qui t'appelle, parce qu'il faut faire du binaural. Je ne sais pas si c'est bien, mais en tout cas c'est ce qui se passe. Et j'imagine que ça va continuer à se développer parallèlement avec, dans le cinéma et d'autres milieux, l'expansion de gros

systèmes, comme l'ambisonie, le mixage objet, etc. Je pense que ça co-existe et ça se co-développe dans des secteurs très différents.

Après, avec la Covid, ça complique un peu les choses. Parce que, par exemple dans les musées, j'ai un parcours sonore qui a dû s'arrêter, car ils n'avaient pas assez de personnel pour nettoyer les casques entre chaque visiteur. Et comme les casques on les manipule avec les mains, « covidiquement » ce n'est pas pratique, alors que pour des expos qui utilisent des enceintes qui sont camouflées, ça ne pose pas de soucis, car le spectateur ne manipule rien avec les mains. Le casque, ça pose donc un problème avec la Covid. Il y a des musées, ou des structures qui sont plus importantes, qui passent un coup de désinfectant à chaque fois. Mais la Covid, ça met un peu le bazar dans les parcours sonores en binaural.

**8**

Radio Campus France fédérant un réseau de radios étudiantes associatives : <https://www.radiocampus.fr>

## ALAN BLUM

Alan Blum. Ingénieur du son et musicien. Co-coordonateur du Master Son à l'ENS Louis-Lumière, il enseigne les « techniques audio », incluant un module d'initiation aux techniques de spatialisation sonore. Il est le co-fondateur de l'association OmniHead, regroupant des preneurs de son autour de l'expérimentation des techniques binaurales. Ses travaux de recherches, entre l'IRCAM et le LIMSI-CNRS, se sont intéressés à l'étude de la plasticité du système auditif en localisation sonore dans un contexte d'écoute au casque de synthèse binaurale<sup>9</sup>.

*Alan Blum is a sound engineer and musician. Co-coordinator of the Master Son at ENS Louis-Lumière, he teaches audio techniques, including a module of initiation to sound spatialization techniques. He is the co-founder of the OmniHead association, bringing together sound recorders to experiment with binaural techniques. His research work, between Iracam and LIMSI-CNRS, has focused on the study of the plasticity of the auditory system in sound localization in a binaural synthesis headphone listening context.*

## LÉA CHEVRIER

Léa Chevrier. Ingénieure du son et créatrice sonore spécialisée en son binaural. Intervenante à l'ESRA pour des conférences autour du son binaural dans le podcast et les parcours sonores. Avec la société de production Narrative<sup>10</sup>, elle travaille sur des parcours sonores en binaural pour le patrimoine : Abbaye aux Dames de Saintes, Château de Vaux-le-Vicomte, Abbaye de Saint-Savin, Théâtre de Mogador. Elle a réalisé en 2015 son mémoire de Master à l'ENS Louis-Lumière sur l'expérimentation des techniques binaurales appliquées au documentaire radiophonique. Site internet : leachevrier.fr

*Léa Chevrier is a sound engineer and sound designer specialized in binaural sound, and also a speaker at ESRA for conferences about binaural sound in podcast and sonic walks. With the production company Narrative<sup>9</sup>, she works on binaural sound paths for heritage: Abbaye aux Dames de Saintes, Château de Vaux-le-Vicomte, Abbaye de Saint-Savin, Théâtre de Mogador. In 2015, she completed her Master's thesis at ENS Louis-Lumière on the experimentation of binaural techniques applied to radio documentaries. Website : leachevrier.fr*

<sup>9</sup>

Voir par exemple : Blum, A., Katz, B. F. G., Warusfel, O. (2004), Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training presence. CFA-DAGA, Mar 2004, Strasbourg, France.

<sup>10</sup>

Narrative : <http://www.narrative.info/>

## Le multicanal étendu au cinéma : entretien avec Cyril Holtz, mixeur cinéma, le 29 novembre 2021

| Sylvain Lambinet

**Sylvain Lambinet : Bonjour, peux-tu te présenter, ainsi que ton parcours ?**

Cyril Holtz : Bien sûr. Alors je m'appelle Cyril Holtz, je suis mixeur de film cinéma. Mon parcours est assez simple avec un premier pas de côté puisque j'ai fait des études d'économie. Bon étudiant à la sortie du Bac, je ne savais pas trop quoi faire donc je me suis orienté vers des études, puis j'ai développé une passion, j'allais de plus en plus au cinéma. Donc, au moment où il a fallu choisir ce que j'allais faire de ma vie, je me suis dit que ce qui m'intéresse, c'est vraiment le cinéma et notamment le son au cinéma, donc j'ai tenté le concours de La Fémis parce que ça combinait les deux choses de la meilleure manière, et je l'ai eu. J'ai rapidement su que je voulais faire du mixage, parce qu'humainement, c'était un cadre qui me convenait mieux, c'est-à-dire des équipes plus réduites, un aspect un peu plus intime. Et puis, on parlera beaucoup, je pense, de l'outil, parce qu'il y a ce côté un peu fétichiste par rapport à l'outil, et les outils qu'on manipule en mixage me semblaient plus corrélés avec mes désirs et mes souhaits, et surtout cette construction qu'on opère au mixage me fascinait. Donc très tôt, j'ai su que c'est ce que je voulais faire et j'ai commencé par mixer énormément de courts-métrages. Et naturellement peu de temps après, j'ai eu un poste technique comme *recorder* dans un studio pendant environ une année. Et puis j'ai fait mon premier long-métrage, et voilà, je suis devenu mixeur !

**SL : Parmi les nombreux bouleversements du son ces quinze-vingt dernières années - le passage**

**au numérique, l'ouverture au 5.1 discret, etc... - le plus récent est celui des technologies dites « son 3D ». Comment ces questions ont-elles commencé à apparaître dans ton métier, jusqu'à la concrétisation dans des mixages entièrement faits dans ces formats-là ?**

CH : C'est presque implicite dans la question, le terme « son 3D » est peut-être un peu abusif. C'est un abus de langage parce que, quand on pense à « son 3D », on pense à une représentation holographique du son, ce qui n'est toujours pas le cas, et qui reste pour l'instant un fantasme, en tout cas dans les systèmes de diffusion conventionnels. La question se pose peut-être plus dans le cadre du binaural. Quoiqu'il en soit, les systèmes disponibles jusqu'à maintenant au cinéma, c'est-à-dire Dolby Atmos et consorts, sont immersifs peut-être, c'est-à-dire englobants. En tout cas, la démarche date d'il y a assez longtemps puisque c'est arrivé dès les premiers formats Dolby multicanaux, et même avant puisqu'il y avait déjà eu des expériences, de plein de façons différentes. Mais la préoccupation était d'englober les spectateurs, comme les premiers termes le décrivaient, de l'ambiophonie, avec tout ce que ça comporte d'approximations. Donc, pour moi on n'est pas dans un « son 3D » proprement dit, pour plein de raisons : il y a des facteurs de qualité - on est loin d'une qualité optimale pour faire ce genre de choses, loin de pouvoir mettre en œuvre des dispositifs suffisamment complexes pour le faire. Et la démarche, en tout cas celle des constructeurs, a finalement toujours été tournée vers le spectaculaire. ►

Et après, on pourrait en parler, mais il y a aussi la question du marketing, de la vente, enfin la mise en avant de la qualité du son ou de son aspect spectaculaire pour vendre des films tout simplement. Donc ma réflexion par rapport à ça, je dois l'avouer, s'est toujours faite avec pas mal de mesure, voire de doute parfois, parce que pour moi cette dimension spectaculaire était l'arbre qui cache la forêt, et j'avais longtemps espéré qu'ils avançaient surtout en termes de qualité, parce qu'on est resté toujours sur les mêmes standards. Ça a l'avantage de tout normaliser, ce qui est un bien, mais en même temps d'entraver les progrès sur l'aspect qualitatif de la diffusion du son au cinéma, tout simplement. On a le même type de haut-parleurs, même s'il y a des avancées, mais le même type d'obstacle justement à la qualité du son, de franchir le cap du fameux écran... Après, au-delà de cette dimension spectaculaire, c'est vrai que ces formats permettent d'être plus, réemployons ce terme, « immergés » dans l'action, mais permettent aussi des choses qui ne sont pas mises en avant par les constructeurs eux-mêmes.

Par exemple, au-delà de l'aspect spectaculaire et très frontal de ces dispositifs, de pouvoir vraiment bénéficier d'un effet d'aération. En fait, c'est ce qui me frappe plutôt que le reste, qui est vraiment intéressant, notamment sur les musiques, et sur toutes les ambiances, c'est-à-dire que ça permet d'avoir, contrairement à ce qu'on pourrait penser, en tout cas c'est ce qui m'a frappé dans mes premières utilisations du système, des bandes sons plus aérées, moins denses, moins chargées. Ça permet de disposer d'une répartition plus agréable et qui laisse mieux respirer les bandes son, et ça je l'ai constaté, de ne pas être toujours à fond en faisant des bandes sonores hyper denses. Alors, c'est vrai que ce n'est pas exactement ce qu'on remarque sur les bandes sons traditionnelles, notamment des gros films américains, qui restent très

denses. Mais voilà en tout cas, c'est l'une des qualités de ces systèmes. Après, si je suis vraiment honnête, je n'ai jamais trouvé que c'était de véritables révolutions. Ce sont des évolutions, ce ne sont pas des révolutions, et la mise en œuvre est relativement lourde quand même, on sent qu'ils sont partis d'un système, mais ils sont rentrés un peu au chausse-pied comme ils pouvaient, en ne voulant pas trop révolutionner et mettre un gros coup de balai sur les normes qui étaient déjà en vigueur et qui, c'est vrai, nous permettaient d'avoir des systèmes assez homogènes et uniformes.

**SL : Ces considérations sur l'aération de la bande son me font penser à Claude Bailblé, qui disait que le haut-parleur est un hublot dans lequel venaient se tasser les sons. C'est exactement cette différence d'aération par rapport à l'écoute naturelle, et là tu parles des musiques et des ambiances comme étant les premières bénéficiaires d'un gain de « qualité », comme tu l'as dit ? Ni « immersion », ni « réalisme », mais « qualité ».**

**CH :** Tu as raison, et c'est marrant que tu remarques ça parce que c'est ce qui me frappe, c'est que même presque malgré moi je continue d'utiliser ces mots parce que ça devient des obsessions. Et c'est vrai que le réalisme en est une autre, parce que j'ai toujours cette obsession, ces deux mots, « qualité » et « réalisme », et « réalisme » ça devient un peu un serpent de mer en cinéma, parce que le cinéma n'est pas réaliste, on le sait très bien, et il ne permet pas le réalisme, ou alors vraiment dans une moindre mesure. Et même dans les formats immersifs, je ne trouve pas le réalisme. Enfin, j'y trouve une grammaire du réalisme et une grammaire de convention, mais le réalisme, c'est très difficile d'y accéder. C'est fou à quel point. Par exemple, une réflexion que je me faisais toujours : quand tu

vas dans un endroit et que tu entends de la musique jouée *live*, même si elle n'est pas acoustique, même si elle passe à travers une sono, etc., tu arrives toujours à dire qu'elle est jouée *live*. Même si c'est très bien joué, tu sais très bien que si c'était un CD avec le même titre joué par les mêmes gens qui passe à travers des enceintes, tu n'aurais pas la même impression. Il y a donc cette notion de réalisme qui, une fois passée à travers le prisme des systèmes audio du cinéma, est de toute façon rompue, et tout le reste marche sur des conventions. Donc pour moi, même ces nouveaux formats, qui pourtant franchissent un pas en termes de spectacle et d'amplification des émotions au cinéma, n'accèdent toujours pas à ce réalisme. Et que ce soit dans l'enveloppement, dans l'immersion... Et se demander si un jour on arrivera, dans une salle de cinéma par exemple, à éprouver les mêmes choses que si on était vraiment dans un environnement qui est mis à l'image, c'est presque devenu un vœu pieux. Et pourtant, je pense que beaucoup d'entre nous, de gens qui travaillent dans le domaine du son à l'image, ont cette idée en tête : comment faire en sorte que ce soit comme si c'était vrai ? On n'en est toujours pas là, ça, c'est sûr, et le binaural s'en approche peut-être parfois plus, même si j'ai moins d'expérience là-dedans, il y a des choses qui sont spectaculaires de réalisme sur certains aspects, et d'autres complètement navrants. Et c'est marrant ce mélange, le binaural c'est plus les extrêmes pour moi, alors que le cinéma a trouvé un langage, disons plus uniforme dans l'irréalisme, c'est un irréalisme qu'on connaît, c'est une représentation du réel qu'on connaît et donc qu'on accepte, peut-être encore un peu mieux avec ces nouveaux formats. Après, il faut savoir les utiliser et les mettre en œuvre, mais ce réalisme-là auquel on voudrait accéder coûte que coûte, on en est loin encore, je pense.

**SL : Musiques et ambiances sont donc les grandes gagnantes, tu as parlé à leur propos d'aération, de démasquage. Je note que ce sont les deux éléments qui n'ont pas ou quasiment pas d'incarnations à l'écran, alors que tu parles de la grammaire du réalisme ou de réalisme de convention du cinéma. Comment réconcilies-tu ça avec le « fantasme » du son holographique, du vœu pieux que tu évoques ?**

CH : La réponse la plus simple, ce serait de dire par la sensation, c'est-à-dire peut-être par quelque chose de moins factuel. Et qui, encore une fois, n'a pas vraiment de rapport avec l'objet filmé. Et tu as complètement raison, c'est très bien formulé, c'est marrant que les choses qui ont vraiment franchi un stade avec le format immersif sont des choses qui sont hors cadre, presque pourrait-on dire, hors film, hors histoire. Et donc, pour moi l'axe principal, là où ces formats immersifs nous apportent beaucoup, c'est l'aspect sensation. Pas simplement sensationnel, mais sensation, c'est-à-dire qu'une fois que tu fais ton deuil du réalisme, ce qui est mon cas, parce que je sais que je serai mort au moment où peut-être on y arrivera, là où on arrive à faire des choses bien, et vraiment ce qui dans le mixage me fascine, ce qui m'intéresse énormément, c'est cette histoire de sensations, c'est-à-dire d'arriver à générer des émotions et des choses peut-être moins concrètement liées à ce qu'on voit et qu'on entend, forcément, mais qui ne sont pas en rapport direct avec l'objet filmé, à créer des choses qui remuent le spectateur, pas simplement remuent au sens spectaculaire, qui procurent des émotions tout simplement et pour moi, ça passe beaucoup par la sensation.

C'est pour ça que je suis très admiratif des réalisateurs qui arrivent à utiliser ça, en s'écartant peut-être parfois ►

même de la narration. Et sans parler en plus de choses très intellectuelles, comme chez Lucie Lassilovitch, qui est une réalisatrice que je trouve absolument formidable parce qu'elle utilise beaucoup la sensation. Elle utilise le son là-dedans. Et, pour moi, c'est peut-être le biais le plus direct pour toucher les spectateurs, parce que c'est quelque chose qu'ils ne peuvent pas maîtriser, et que nous, nous pouvons maîtriser. Et je ne parle même pas des effets spectaculaires pour créer des effets de surprise, mais quelque chose de plus profond, de plus enfoui, le frissonnement quand on entend la pluie ou le mélange de sons, des mélanges incongrus ou au contraire des choses qui te galvanisent. Et ça pour le coup, plus les formats sont ambitieux, et c'est quand même le cas d'un Dolby Atmos, plus tu peux le faire, et le faire plus facilement. Pour moi, finalement, c'est ça l'immersion, c'est ça la seule *immersion*. En fait, c'est la sensation, c'est-à-dire qu'il ne s'agit pas de l'espace que tu crées autour des spectateurs, mais de l'espace que tu crées à l'intérieur d'eux, et peut-être que là on n'y arrive plus facilement. Après, à quel degré et avec quelle efficacité ça permet de le faire par rapport au 5.1 par exemple, je n'en sais rien, mais j'ai l'impression quand même que ça nous laisse une marge de manœuvre plus importante.

**SL : Puisque tu as fait plusieurs films en Atmos, quel rôle, quelle place prend cet aspect dans ton dialogue avec les réalisateurs et quelle est son attitude dans ce qui se passe au mixage, sachant qu'il est dans un dispositif nouveau, quasi expérimental.**

CH : Alors je vais reprendre des termes que j'ai déjà utilisés. En tout cas, ça les fait tous kiffer, ça c'est sûr. Mais il y a beaucoup de cet enthousiasme qui reste lié à des fantasmes. Et il y a beaucoup de suggestion,

d'autosuggestion, presque comme tout le monde, on est victime un peu de cette mise en avant de l'aspect technologique. Et ça marche, je cite toujours cet exemple du THX : personne à l'époque ne savait ce que c'était le THX, c'était purement et simplement un cahier des charges et effectivement, ça permettait d'accéder à une certaine qualité. Mais finalement, les spectateurs ne savaient pas ce qu'ils entendaient. Il suffisait de dire que c'était du THX et c'était mieux, et je pense qu'on est tous un peu victimes aussi de ça. C'est-à-dire que les constructeurs maintenant savent valoriser un produit - parce que c'est un produit - et les réalisateurs savent à peu près ce que c'est, mais dans l'à-peu-près, on y met beaucoup de choses, et on se dit parfois un peu à tort qu'on peut faire des choses incroyables, mais qui relèvent du fantasme pour moi. Cela étant, on trouve toujours des solutions, et peut-être que ce sont des solutions qu'on mettait déjà en œuvre quand on n'avait que du 5.1, et cetera.

Après, encore une fois, ça permet de franchir un cap, mais ce que je veux dire, c'est qu'ils ne sont pas tous particulièrement informés et éduqués de ce qu'on peut faire vraiment, de quelles sont les avancées de ce dispositif par rapport au multicanal traditionnel. Donc, peu savent que l'on peut avoir des surrounds pleine bande. Peu savent que finalement tu as des traitements - je ne parle même pas du mixage objet, ça, ça reste du chinois pour le commun des mortels. Donc, derrière ce terme, on se figure qu'on peut tout faire, et que tout d'un coup on va entendre des choses qu'on ne pouvait absolument pas entendre ou faire. Et pour moi c'est un écueil quand même, parce qu'on vend quelque chose, il y a un demi-mensonge. Après, ça permet quand même de donner le change sur certaines choses. En tout cas, jamais je ne me mettrai à parler technique très précisément, de ce que ça permet ; à la limite je

préfère que le fantasme reste parce que ça permet de faire, il y a un vrai enthousiasme, mais c'est important cet enthousiasme. Donc dans mon rapport avec le réalisateur, c'est à moi finalement de faire en sorte qu'il entende ce qu'il fantasmat.

**SL : Ni plus ni moins que dans un mixage traditionnel.**

CH : Exactement. Sauf que là, l'exigence devient supérieure et elle peut légitimement l'être, c'est-à-dire qu'un réalisateur choisissant ce format se dit : « Je vais faire des choses que je ne pouvais pas faire avant ». Et c'est vrai dans une certaine mesure, mais en fait, ils ne savent pas bien dans quelle mesure. Alors ils savent que, par exemple, on a le droit, on peut être encore plus spectaculaire. C'est vrai. Ils savent qu'on va avoir peut-être plus de précision dans les localisations, ce n'est pas faux, quoique...

**SL : Peux-tu élaborer sur ce « quoique » ?**

CH : Non, mais voilà, il y a ce fameux truc des « objets ». Alors pour moi la précision « pinpoint laser » de la localisation des objets dans l'espace, c'est valable pour certaines choses très particulières. Il faudrait poser la question à un psychoacousticien, mais de la même façon que je te citais l'exemple du concert où on arrive dans un endroit et on entend la musique jouée *live*, ça sera toujours le cerveau qui comprendra et fera la différence, et ce n'est pas parce qu'on va mettre un objet dans un seul haut-parleur en haut très précisément que le cerveau va se dire : « Ah, c'est le vrai objet qui est au-dessus de moi » ; non, pas du tout. Et encore une fois, on retombe sur une nouvelle convention, c'est-à-dire que : oui, c'est spectaculaire d'entendre une bombe passer au-dessus de soi. C'est spectaculaire, ça ne veut

pas dire que les gens se diront que c'est vrai. Donc, on est encore loin de ce que le son dit « 3D » prétend faire pour moi. Après, ce n'est pas une critique, c'est une constatation, c'est-à-dire que : certes c'est un progrès, et un progrès énorme, mais on n'atteint pas le réalisme, et ce « quoique », c'est ça. C'est-à-dire qu'on peut faire de nouvelles choses ; on ne peut pas faire des choses qui changent la donne de manière hallucinante, en tout cas sur la localisation.

**SL : Donc c'est un nouveau territoire qu'on a exploré, plutôt que l'avènement de ce qu'on avait toujours espéré.**

CH : C'est exactement ça.

**SL : sur la question de ces codes cinématographiques, il y a pour l'image cette fameuse ligne des 180°, qui permet au cerveau de comprendre un raccord qui est au fond une expérience paradoxale, sensoriellement. As-tu déjà pensé ou expérimenté une semblable grammaire du sonore ?**

CH : Oui, alors je n'ai pas attendu les nouveaux formats pour le faire évidemment, mais je l'ai fait parfois un peu à mes dépens en voulant aller jusqu'au bout. Et pourquoi ? Parce que c'est fait à la petite semaine, mais pas uniquement de mon côté, c'est-à-dire que ces problèmes ne sont pas mis en scène, donc, lorsqu'au mixage tu cherches à les mettre en scène alors qu'ils n'ont pas été pensés dès le début par le metteur en scène, tu te confrontes au fait que, autant ton cerveau accepte un champ-contrechamp sans problème, autant le mien et je pense le cerveau de la plupart des gens, vont être très gênés si, lors d'un champ-contrechamp, ce qui s'entend à droite dans le champ, passe à gauche ►

ou à l'arrière-gauche dans le contrechamp.

**SL : D'où peut être ce que tu disais à propos des musiques et des ambiances qui restent en dehors de cette grammaire ?**

CH : Exactement, qui restent floues. Et d'ailleurs, il suffit de voir ce fameux paradoxe : qu'est-ce qu'on fait au jour le jour ? En tant que mixeur, on s'en démerde, c'est-à-dire qu'on arrange les bidons vaguement, qu'on va le jouer à certains moments et pas à d'autres, que si ça a une importance en début de scène de localiser une source sonore, on va le faire, puis, dès qu'on va passer dans un mode de film plus conventionnel, un champ, contrechamp, etc., ils vont tout d'un coup rester à peu près à leur place. On est toujours dans une espèce de compromis. Le cas le plus intéressant quand on veut vraiment respecter, c'est le fameux plan-séquence, où effectivement c'est un rêve pour un mixeur, les plans-séquences, non seulement parce que ça te permet – je prêche vraiment pour ma paroisse ! - d'aborder une séquence en te laissant l'occasion de vraiment faire du mixage, c'est-à-dire d'avoir un vrai geste dans la continuité et de ne pas être obnubilé par le découpage, mais aussi d'être dans une cohérence de localisation des objets. Et là, tu peux faire ce que tu veux, c'est-à-dire que tu peux vraiment t'amuser à la respecter, et c'est cohérent avec le film lui-même et la construction des plans. Et là, pour le coup ça devient vraiment super intéressant de travailler dans ces formats. Et l'autre aspect, que moi en tout cas j'adore explorer, c'est qu'en termes de sensations tu peux vraiment faire des choses incroyables, tu peux faire tourner les sons comme tu veux, tu peux essayer ce que ça donne. Après, il faut que le film s'y prête, mais ça, c'est extrêmement intéressant.

**SL : Dans un plan séquence en mouvement,**

**avec des sources en mouvement autour du point caméra, qui est en même temps le point d'écoute, il y a un débordement des sources sonores qui viennent envahir la salle. Que penses-tu de cette sortie hors de l'écran, dans le cas des dispositifs immersifs ?**

CH : J'emploie la même prudence. Après, il y a là une question plus personnelle et individuelle, et toujours dans ce souci, peut-être un peu utopique, de garder une certaine modestie. C'est-à-dire que je ne veux pas, même dans ce genre de cas, que le son détourne l'attention de ce qu'on est en train de voir. De même, si on est la caméra, je ne veux pas, c'est un cliché, je ne veux pas que le spectateur se retourne ; je veux qu'il ait envie de continuer à suivre ce qui se passe. Donc, j'ai toujours la même mesure. En revanche, ça permet non seulement une meilleure fluidité, ça c'est certain, parce qu'il y a une disparité moins grande entre les différents haut-parleurs, du fait des extensions de basses à l'arrière et du nombre de haut-parleurs, donc on peut s'autoriser à aller plus loin dans ce genre de choses, et surtout avec une amélioration nette du résultat. Donc, autant on est plus aéré globalement quand ce sont des objets, disons flottants, mais fixes, statiques, autant pour des objets en mouvement tu as aussi un vrai gain quand même, ça, c'est certain.

**SL : Donc finalement, comme c'est revenu plusieurs fois, ce sont moins les questions de localisation, de spatialisation, qui ont l'air de t'intéresser, que la possibilité d'avoir un signal « pleine bande » tout autour de la salle.**

CH : Oui, que ce « pleine bande » tout autour de la salle, et la faculté aussi d'avoir des objets où tu peux faire du coup des mouvements très précis, puisque tu

peux les attribuer à un haut-parleur très précisément. Donc, les passages se font de manière très fluide entre les différents haut-parleurs avec les objets. Mais tu résumes bien ma pensée, c'est-à-dire que c'est plus sur des choses annexes, et qui ne sont pas forcément mises en avant dans les qualités du système, que je trouve qu'il y a un vrai progrès. Après, c'est vrai que, par exemple, encore une fois, le « pleine bande » à l'arrière permet enfin d'avoir des effets un peu puissants, que tu peux sans arrière-pensée décaler de la façade, alors qu'avant tous les trucs avec du -6 dBFS à 20 Hz dans les arrières ce n'était même pas la peine d'essayer, ou alors dès que tu décalais un peu les musiques de la façade, alors qu'elles étaient très riches en basses, tu perdais, même en précision. Ça, ça reste toujours le cas parce que ce sont toujours les mêmes haut-parleurs à l'arrière ; à part l'extension de basse, ça a un peu évolué, mais ce n'est pas non plus la panacée. Donc encore une fois, amélioration nette, mais ça ne change pas drastiquement la donne. C'est bien, mais ce n'est pas révolutionnaire.

**SL : As-tu une idée de l'appréhension du public, lorsqu'il écoute un film dans un format différent, as-tu l'impression qu'au-delà du label Dolby Atmos, il y a une réception différente ?**

CH : Oui, mais je ne sais pas bien mesurer ce qui est lié à cet aspect commercial, et ce qui est lié à la réalité. Je pense quand même que, lorsque les gens vont voir un film dans une salle Atmos, le système a de fortes chances d'être d'une meilleure qualité globale que dans une salle conventionnelle. Pourquoi ? Parce que c'est du matériel plus récent, que c'est peut-être mieux vérifié, ça a été réglé, donc ils accèdent forcément, selon moi, à une meilleure qualité, ne serait-ce que par ça, comme c'était le cas déjà du THX. Ils ont de plus

grandes chances d'écouter un film dans de meilleures conditions que dans une salle avec une installation qui n'a pas forcément été vérifiée récemment. Donc pour ces raisons, il va peut-être se produire quelque chose, pour des gens qui avaient l'habitude de voir des films dans de moins bonnes conditions, ils vont accéder plus facilement à cette dimension spectaculaire qui est quand même le maître mot dans le discours de Dolby en l'occurrence. Donc forcément, ils risquent de noter quand même une différence par rapport à une expérience plus traditionnelle de cinéma. Par ailleurs, les films qui sont mixés en Dolby Atmos, en tout cas en France, mais aussi aux États-Unis, sont souvent des films avec cette dimension toujours spectaculaire, donc forcément ça va impressionner, et ça, ça ne se dément pas. Après, il y a toujours cette histoire de niveau sonore. Je pense que ça fait belle lurette qu'on a franchi le Rubicon sur les niveaux, et désormais encore plus en Atmos, puisqu'on peut se permettre des choses qu'on ne pouvait pas se permettre, notamment sur les arrières, et je pense que ça fait longtemps qu'on va trop loin. Alors, c'est vrai que c'est le meilleur moyen, on arrive à ce paradigme qui est bien triste : « Plus c'est fort, mieux c'est ». J'espère qu'on en reviendra, mais je pense que ces formats n'aident pas forcément. Alors que, par exemple, dans les films que j'ai faits ou en tout cas dans une partie des séquences qui ont été mixées, j'avais en tête d'essayer de faire puissant et vraiment impressionnant, sans que l'on ait envie de se boucher les oreilles. Et pour ça, l'Atmos c'est très efficace, c'est-à-dire justement par cette aération à laquelle on peut accéder, un meilleur démasquage des éléments, on peut se permettre sans être hyper fort d'arriver à un truc déjà super impressionnant, sans que tu t'explodes les oreilles. Donc c'est toujours la même chose, l'outil en lui-même n'est pas critiquable, la façon dont on l'utilise après, peut-être, est à revoir. ►

**SL : À chaque fois qu'on mentionne l'outil, on dit « Atmos »... Ce sont les seuls à proposer un outil crédible ?**

CH : Non, ça veut dire que Dolby a réussi son coup, c'est-à-dire qu'ils sont tellement mis en avant, ils ont une telle hégémonie qu'on ne parle plus que d'eux, alors que DTS :X c'est la même chose. Sur le *Eiffel*, le dernier que j'ai fait en Atmos, on a fait a posteriori une transposition en DTS:X. Résultat : c'est exactement identique, aucune différence. Mais les auditoriums sont équipés en Atmos, pas forcément en DTS:X. Je connais pas mal de salles maintenant en France qui diffusent en Dolby Atmos, moins en DTS :X. Voilà, c'est toujours la même chose, c'est un peu le plus gros qui a raison et je suis le premier à le déplorer. Mais après ça ne veut pas dire que l'outil est mauvais.

**SL : Ça met en lumière que le grand bénéfice, c'est le meilleur équipement en termes de diffusion dans la salle, l'augmentation de la qualité de la calibration, de l'homogénéité des systèmes, pas forcément la technologie de spatialisation de tel ou tel fabricant.**

CH : Exactement, et encore une fois, pour illustrer mon propos sur l'amélioration de la qualité, par exemple, lorsque tu vas au Pathé Beaugrenelle, le fameux Dolby Vision, c'est spectaculaire, c'est étonnant, étonnant. Les capacités du système, l'endurance, la tenue en puissance c'est impressionnant, et pour le coup, même pour un professionnel rien que sur l'aspect technique, c'est fort quand même. Après, j'y ai vu un film qui était diffusé à un niveau bien trop fort à mon avis. C'était vraiment insoutenable, mais par contre tu es époustouflé par l'endurance du truc et la capacité à supporter des niveaux pareils ! Là, on a vraiment un grand progrès

parce que, pour arriver à bout du système, tu as une grande marge, une grande réserve, ça, c'est chouette. Mais encore une fois, c'est une débauche de moyens, une débauche technique, mais est-ce qu'on la met vraiment à profit ? C'est marrant, on parlait de cette espèce de course entre la technique et ce que tu en fais. Et là, on arrive à un point où on ne sait pas qui guide l'autre ? Quand même, quand Dolby sort ces systèmes, je n'ai pas l'impression que si tous les metteurs en scène de bonne volonté et tous les professionnels du son se mettaient à une table et se demandaient « quel est notre système de rêve ? », je n'ai pas forcément l'impression qu'ils prendraient de l'Atmos. Ou peut-être que je me trompe, hein !

**SL : Et bien justement, quel avenir prévois-tu aux systèmes actuels, quels développements espèrerais-tu sur la voie ouverte ?**

CH : Alors, revenons aux essentiels ! Un système qui permet d'accéder à un degré de réalisme supérieur, ce serait génial. Encore une fois, une amélioration de la qualité de diffusion parce que, à une époque où on a des systèmes de concerts qui sont dingues, en bande passante, en dynamique et cetera, je ne comprends pas pourquoi en cinéma, quand on fait venir un musicien, même dans un auditorium super bien réglé, il tombe toujours de haut en disant : « Oh là là ! Mais alors, et les aigus ?! » – « Ben oui, mais les aigus bye à l'écran, désolé... » - « Ah oui, mais alors les basses c'est quand même un peu mou... » - « Et bah, oui, mais désolé, ce sont les haut-parleurs qui sont comme ça, même avec les nouvelles enceintes. »... Et c'est vrai qu'on a progressé, avec des systèmes qui se sont quand même pas mal améliorés, mais on reste dans quelque chose d'un peu déceptif. Et encore une fois, je ne parle pas de volume sonore, je parle de qualité perçue. Alors, si on

parle de spectacle en soi, même si je ne suis pas du tout partisan du cinéma 3D en relief, etc., je me souviens de l'une des premières petites démos que j'avais écoutées en binaural : le *Virtual Haircut*, où le ciseau te passe à 2 cm, j'avais des frissons, parce que tu sens vraiment les ciseaux te passer contre l'oreille. Et ça, c'est l'aspect plus ludique disons, mais le son ce n'est pas que des trucs dramatiques, ça peut être aussi ça, d'avoir ce genre de trucs, de nouvelles sensations, j'adorerai. Je rêve d'un système pareil.

Là encore, c'est un peu un fantasme, mais peut-être qu'il mélangerait deux types de diffusion. Une diffusion binaurale au travers d'un casque ouvert pour certains sons, et une diffusion traditionnelle ? Enfin tu vois, un truc où tu te dises : « On aura des choses qui changent,

qui bouleversent un peu les conventions et la grammaire du son telle qu'on la connaît jusqu'à maintenant. » En tous cas, il y a des choses qui m'intéressent énormément dans le binaural, c'est sûr. Après, il y a ce truc de porter un casque qui est toujours un petit peu... Moi je suis un grand défenseur du cinéma, je suis le premier à croire fermement que cette expérience de la salle est une expérience collective où on partage quelque chose, avec des étrangers en plus, ce qui est de moins en moins commun et qui est pour moi absolument essentiel. On partage des émotions avec des gens qu'on ne connaît pas dans un même lieu, et tout ce qui vient l'entraver est dommageable et délétère. Est-ce qu'on trouvera un système qui permette de garder ça, qui est essentiel, et d'améliorer cet aspect perceptif ? J'espère et j'aimerais.

#### **SYLVAIN LAMBINET**

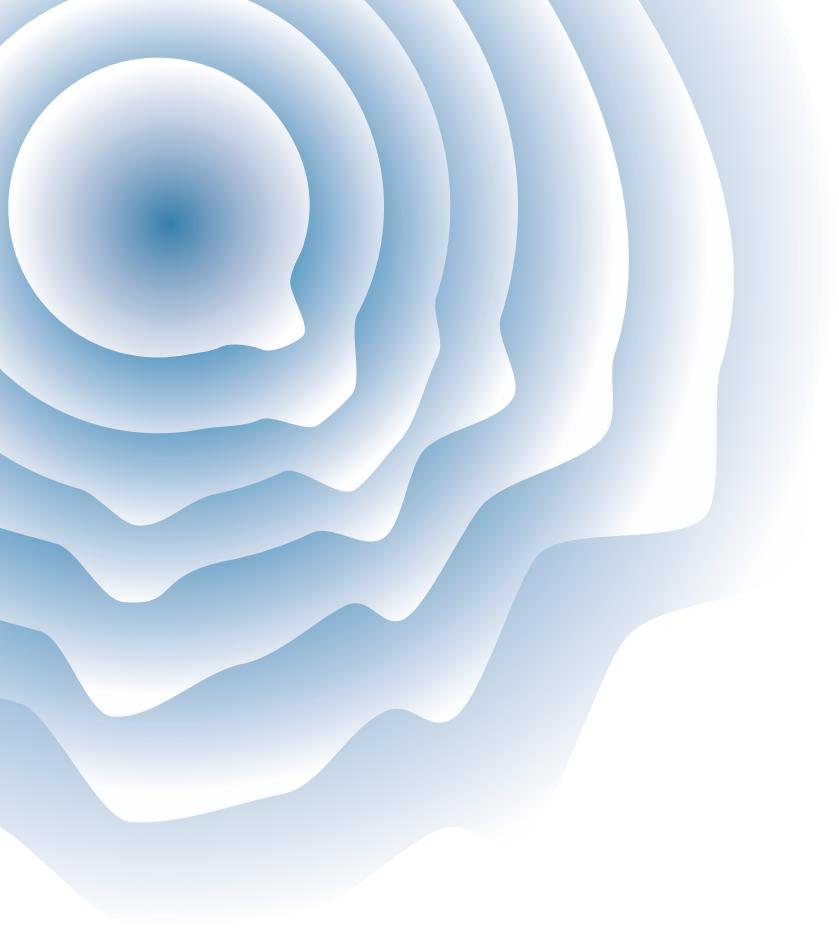
**Après des études de philosophie et un cursus en son à la Fémis, Sylvain Lambinet se tourne vers l'enseignement au sein de l'ENS Louis-Lumière. Il y dispense les cours sur les fondamentaux du son à l'image, et poursuit le développement d'outils de spatialisation.**

*After studying philosophy and graduating from La Fémis as a sound engineer, Sylvain Lambinet opt for a career in teaching, within ENS Louis-Lumière, where he is in charge of the fundamentals of sound for films. He also pursues the development of audio tools, especially for sound spatialization.*

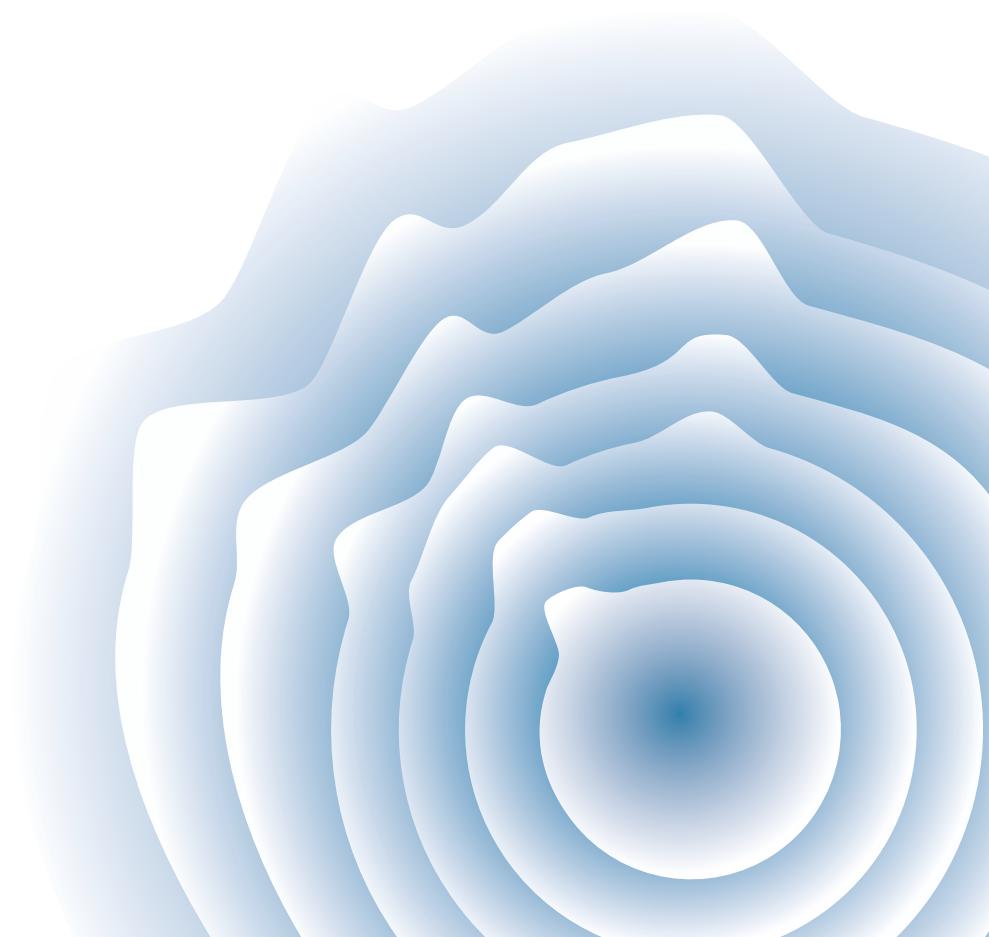
#### **CYRIL HOLTZ**

**Après des études en ingénierie du son à la Fémis, Cyril Holtz est mixeur pour le cinéma. Nominé huit fois aux Césars du meilleur son, et trois fois récompensés, (pour *Sur mes lèvres*, de Jacques Audiard en 2002, *Gainsbourg vie Héroïque* de Joann Sfar en 2011 et *les Frères Sisters* de Jacques Audiard en 2019). En parallèle de son activité de mixeur, il fonde HAL Audio, un regroupement d'ingénieur du son, dont les activités s'étendent à la création sonore au-delà du cinéma, ainsi qu'à la recherche et au développement d'outils au service des praticiens.**

*Since graduating from La Fémis as a sound engineer, Cyril Holtz is a mixer for cinema. Eight times nominated at the Césars for Best Sound, and three times winner ( *Sur mes lèvres*, by Jacques Audiard in 2002, *Gainsbourg vie Héroïque* by Joann Sfar in 2011 and *The Sisters Brothers* by Jacques Audiard in 2019). While pursuing his career as a mixer, he is also a founding member of HAL Audio, a group of sound engineers whose activities go beyond sound for films and extend to other forms of sonic art, as well as research and development of tools for the practitioners.*



**Varia**



## Alternative aux analyses de Fourier globale et glissante pour les signaux temporels : Analyse-Resynthèse par Intégration de Densité Spectrale

| Laurent Millot

### Résumé

Dans cet article, nous proposons une alternative aux analyses de Fourier globale et glissante qui sont largement utilisées dans l'étude des signaux audio et plus largement dans celles des signaux temporels. Après avoir pointé des limitations ou problèmes rencontrés avec les analyses de Fourier lors de leur utilisation en analyse, nous verrons les principes d'analyse et de resynthèse de l'analyse-resynthèse par Intégration de Densité Spectrale (ou analyse-resynthèse IDS) qui permet notamment une écoute des resynthèses partielles à totales des signaux ou scènes sonores étudiés. Un panorama non exhaustif des applications déjà réalisées de cet outil permet de constater que cette analyse-resynthèse est efficace pour traiter d'un large éventail de problèmes de nature physique quelconque.

### Abstract

*In this article, we propose an alternative to the global and sliding Fourier analyses that are widely used to analyse audio signals and more broadly temporal signals. After pointing out the limitations or problems encountered with Fourier analyses when using them in analysis, we explain the principles of analysis and resynthesis of the Spectral Density Integration analysis-resynthesis (or IDS analysis-resynthesis), which allows, in particular, the listening of partial resyntheses to totals of the signals or sound scenes studied. A non-exhaustive panorama of the applications already carried out with this tool is proposed and it shows that this analysis-synthesis makes it possible to treat a broad range of problems of any physical nature.*

Les principaux outils de traitement du signal utilisés pour réaliser une analyse de signaux temporels en audio sont les analyses de Fourier globale ou à court terme (encore appelée transformée de Fourier glissante), ou des déclinaisons de l'analyse de Fourier à court terme comme les transformées de Fourier en cosinus discrètes, préférées dans les outils audio *nécessitant des latences de calcul faibles*, car elles ne nécessitent pas la mise en œuvre de calculs avec des nombres complexes.

On trouve aussi quelques équipes utilisant les transformées en ondelettes (Mallat, 2000 ; Meyer, 1992 ; Truchetet, 1998), notamment rapides. Les transformées en ondelettes rapides utilisent un seul jeu de paires complémentaires de filtres d'analyse et de resynthèse, et un sous-échantillonnage (respectivement sur-échantillonnage) systématique après chaque étage de décomposition en sous-octaves (respectivement de resynthèse ou rassemblement de sous-octaves). Si cette astuce permet de ne conserver pour chaque signal de sous-octave qu'un signal échantillonné à un échantillonnage minimal avec une fréquence deux fois supérieure à la largeur de la sous-octave, elle ne permet une resynthèse sans pertes et sans distorsion que si on prend en compte tous les signaux pour toutes les sous-octaves et, de plus, sans introduire de modifications pour ces signaux. À l'inverse, si on calcule une resynthèse partielle avec des modifications des signaux de sous-octaves retenus, le signal resynthétisé subit systématiquement un repliement spectral en proportion conséquente, ce qui rend *de facto* impossible toute écoute. En effet, les deux jeux de paires de filtres d'analyse et de resynthèse sont conçus de manière à éliminer le repliement spectral introduit par les décompositions en sous-octaves à condition que tous les signaux de sous-octaves repassent par toutes les étapes de sur-échantillonnage et de filtrage nécessaires sans qu'aucune modification de ces signaux ne soit introduite (Mallat, 2000 ; Meyer, 1992 ; Truchetet, 1998).

C'est probablement une des raisons qui a conduit au remplacement des étages de décomposition en sous-octaves par des analyses par transformées de Fourier en cosinus discrètes dans les outils de compression audio (notamment pour les formats audio suivants : MP3, AC3, Ogg Vorbis, WMA, ATRAC, AAC, LDAC, Dolby AC-4, MPEG-H 3D Audio). Et, c'est pour cette raison, principalement, que la transformée en ondelettes ne constitue pas un outil intéressant dans le contexte d'étude proposé.

Aussi, dans cet article, nous présentons une alternative aux analyses de Fourier, globale comme à court terme, pour l'étude des signaux temporels. Et, l'outil proposé n'est pas seulement une analyse, puisqu'il permet une resynthèse des signaux étudiés, partielle à complète, comme on le verra par la suite.

Dans la deuxième section, on passe en revue une sélection de limitations, voire de problèmes liés à l'utilisation des analyses de Fourier globales à court terme en illustrant certains des défauts pointés grâce au tracé de spectres d'amplitude. Dans la troisième section, on décrit le principe d'analyse et de resynthèse de l'outil proposé. Dans la quatrième section, on recense une partie des utilisations déjà opérées de cet outil dans le cadre de travaux de recherche comme de mémoires de Master de l'ENS Louis-Lumière (Master Son) qui permettent de constater que cet outil peut être utilisé pour étudier de nombreux cas de figure. Dans la cinquième section, on présente des informations sur les outils réalisés et en cours de développement, ainsi que les conditions pour y accéder et les aides disponibles pour leur prise en main et leur utilisation. Enfin, dans la dernière section, on dresse l'inventaire d'utilisations futures possibles de cet outil, pour certaines dans un avenir proche et on propose une synthèse des conclusions quant à l'utilisation de cet outil. ►

## Pourquoi une analyse-resynthèse supplémentaire ?

Dans la suite de cette section, nous ne considérons donc plus que le cas de l'analyse de Fourier, globale ou à court terme, pour expliquer pourquoi il semble intéressant de proposer une analyse-resynthèse supplémentaire pour les signaux temporels. Nous allons seulement esquisser une argumentation restreinte des raisons pour lesquelles il semble intéressant de proposer un outil alternatif sans entrer dans le détail des explications mathématiques. Le lecteur intéressé à connaître ce détail pourra, par exemple, se reporter à la lecture de (Mallat, 2000), (Millot, 2007), (Max et Lacoume, 1996) ou (Bellanger, 2006). Dans la suite, on désigne par spectre le signal fréquentiel obtenu après application de la transformée de Fourier considérée au signal temporel étudié.

Un des arguments les plus importants en faveur de la recherche d'un nouvel outil d'analyse repose sur le constat que les analyses de Fourier ne permettent pas une resynthèse partielle facile des signaux étudiés et ne permettent donc pas une écoute de resynthèses partielles avec ou sans modifications des éléments d'analyse. En fait, on peut proposer que ces analyses se traduisent essentiellement par le tracé des spectres d'amplitude le plus souvent, voire des spectres de phase ou même de retard de groupe, principalement dans le cas de l'analyse de filtres temporels.

D'un point de vue mathématique, la transformée de Fourier correspond à la projection (produit scalaire) du signal dans une base d'exponentielles complexes, non pas avec toutes les fréquences présentes en théorie dans le cas de la transformée de Fourier analogique (TF), mais juste avec un nombre limité de fréquences choisies, souvent réparties de manière régulière, dans le cas de

la transformée de Fourier (numérique) discrète (TFD). Et, à partir des informations pour une fréquence donnée et la fréquence opposée, on peut, au besoin, en déduire une décomposition en sinusoides, mais seulement pour un nombre de fréquences limitées dans le cas de la TFD. La transformée de Fourier rapide (*fast Fourier transform* en anglais, d'où l'acronyme FFT couramment utilisé en traitement du signal) correspond à une optimisation du calcul de la TFD le plus souvent pour des TFD calculées sur des signaux numériques de longueur égale à une puissance de 2 ; au besoin, si le signal à étudier n'a pas une durée égale à une puissance de 2, on le complète par l'ajout du nombre adéquat de zéros (technique dite du *zero-padding*), ce qui n'ajoute aucune information ou distorsion pour le calcul du spectre. Ainsi, lors des utilisations pratiques, les spectres des TFD ne sont donc calculés que pour un nombre limité de fréquences, avec un compromis à trouver entre un nombre de fréquences calculées important, ce qui s'inscrit dans une optique de recherche de précision pour la description, et un nombre aussi faible que possible de fréquences pour limiter les temps de calcul de ces spectres et obtenir une description pas trop lourde.

Dans le cas de la transformée de Fourier globale, on calcule et on trace le spectre pour la durée totale du signal, ce qui conduit à une description statistique du signal puisque le spectre d'amplitude fournit alors la probabilité de présence de chacune des composantes fréquentielles pour lesquelles il a été calculé, à chaque instant dans le signal temporel. En effet, la transformée de Fourier correspond à une projection du signal étudié dans une base d'exponentielles complexes ce qui se traduit par la concentration des évolutions temporelles dans les exponentielles complexes avec une description fréquentielle du signal, son spectre, qui est indépendante du temps. Ainsi, lorsqu'on considère le spectre d'énergie, on obtient le poids relatif de

chacune des composantes fréquentielles qui ont été calculées<sup>1</sup>. Les variations temporelles ne sont donc retrouvées, en théorie, que si on prend en compte toutes les composantes fréquentielles pour la resynthèse du signal temporel. On dit, par abus de langage, que les informations temporelles se « cachent » dans les phases (ou le retard de groupe) du spectre. Cela signifie que l'on ne dispose d'aucune information sur la présence effective d'une composante fréquentielle calculée à un instant donné, donc d'aucune information sur l'évolution temporelle du signal étudié. Mais, si on prend en compte beaucoup de fréquences pour le calcul du spectre, on peut avoir une précision et/ou une discrimination fréquentielle assez fine, au prix d'une description très lourde.

Dans le cas de la transformée de Fourier à court terme, on procède au calcul de spectres successifs par TFD (FFT en général) sur une portion limitée du signal temporel, en déplaçant, après calcul d'un spectre, la fenêtre de calcul (d'analyse), encore appelée horizon temporel, pour considérer une nouvelle portion, de même durée, du signal qui fait l'objet du calcul d'un nouveau spectre par TFD. Il est alors nécessaire de trouver un compromis entre la taille des portions de signal pour lesquelles on calcule un spectre par TFD et le pas temporel dont on translate la fenêtre d'analyse entre deux calculs de spectres successifs. De cette manière, on peut essayer de conserver une trace de l'évolution temporelle d'une fenêtre d'analyse à l'autre, mais il faut rester conscient que, pour chaque fenêtre, on ne dispose toujours que d'une information statistique sur la présence éventuelle des fréquences prises en compte pour le calcul du spectre. Le tracé du spectre d'amplitude de tous les spectres successifs calculés correspond à ce que l'on nomme un *sonagramme* quand on utilise une représentation bidimensionnelle

(Figure 1, gauche), et à un *waterfall* (« cascade », ou « chute d'eau », littéralement) quand on utilise une représentation tridimensionnelle (Figure 1, droite). Dans le cas des *waterfalls*, le troisième axe correspond aux niveaux d'amplitudes exprimés en général en décibels. Dans le cas du sonagramme, les deux axes des tracés correspondent au temps et aux fréquences, et, on utilise un codage des couleurs des *échelles de niveaux des amplitudes généralement en décibels* : codages en niveaux de gris ou avec des couleurs chaudes, froides ou allant des couleurs chaudes aux froides. Ainsi, dans le cas des sonagrammes, l'utilisateur doit aussi définir la dynamique des niveaux qui seront représentés et choisir le jeu des couleurs (*color map*) qu'il va utiliser. On voit donc se dessiner la nécessité de prédéfinir beaucoup de paramètres d'analyse et donc d'apprendre à lire tant les tracés des spectres globaux que des sonagrammes. En pratique, chaque utilisateur arrête un choix limité de paramètres qu'il utilise ensuite de manière quasi systématique pour essayer d'interpréter les données figurant dans les tracés des spectres ou des sonagrammes.

Mais, ce ne sont pas les seuls défauts ou problèmes associés à l'analyse de Fourier, globale ou à court terme. Un autre problème réside dans la mise en œuvre des outils numériques de la TFD : il faut opter pour un nombre de fréquences limité. Ceci signifie que l'on ne dispose, en pratique, que des informations pour les seules fréquences calculées qui, de plus, sont réparties de manière régulière puisque l'on procède au découpage en intervalles fréquents égaux de l'intervalle allant de 0 Hz à la fréquence d'échantillonnage utilisée. Or, avec des fréquences réparties régulièrement suivant ce schéma, la probabilité que ces fréquences correspondent aux fréquences effectivement présentes, à un moment ou un autre, dans le signal étudié est très proche de ►

<sup>1</sup> Toutes les fréquences pour une transformée de Fourier analogique ou à temps discret, uniquement un nombre limité de fréquences pour une transformée de Fourier discrète

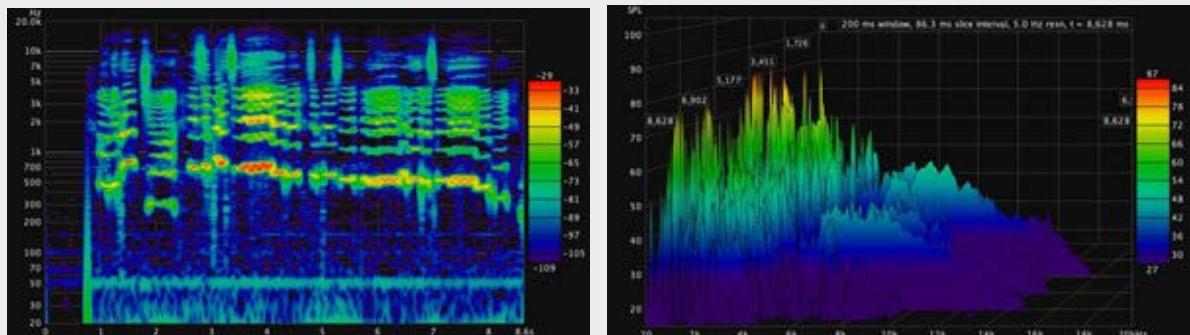


Figure 1 : Tracé du sonagramme (à gauche) et du waterfall (à droite) pour un signal de chant enregistré en chambre anéchoïque.

zéro. Les spectres ou sonagrammes tracés ne donnent qu'une indication partielle et ne garantissent pas que l'on ne passe pas à côté de la présence d'une fréquence présente dans le signal et associée à une largeur spectrale trop fine pour générer une contribution sensible des deux fréquences calculées qui l'encadrent. Ainsi, on peut proposer que ces spectres ne donnent qu'un aperçu du nombre de fréquences effectivement présentes et qu'il faut mettre en œuvre une recherche supplémentaire par analyse paramétrique supposant un nombre donné de fréquences présentes pour obtenir une caractérisation précise de l'amplitude, de l'amortissement et de la phase des composantes fréquentielles qui auront été effectivement identifiées par une analyse paramétrique supplémentaire. Les analyses des tracés des spectres ou des sonagrammes constituent donc plutôt une analyse préliminaire du contenu spectral potentiel d'un signal.

Un autre problème encore plus gênant tient à l'utilisation d'une durée limitée tant pour le calcul d'un spectre global pour l'analyse de Fourier globale que dans le cas du calcul d'une succession de spectres locaux pour l'analyse de Fourier à court terme. En effet, l'analyse

de Fourier analogique comme numérique repose sur l'utilisation de composantes fréquentielles supposées présentes depuis le « début des temps » et subsister jusqu'à la « fin des temps », ce qui se traduit par un support temporel ou un horizon temporel infini pour les exponentielles complexes et, par voie de conséquence, pour les sinusoides, constituant la base dans laquelle on projette puis on resynthétise le signal temporel étudié. On peut considérer que l'étude d'un signal temporel de durée limitée correspond, en pratique, au calcul de la TF ou TFD pour un signal temporel de durée infinie multiplié par une fenêtre d'observation rectangulaire de la durée effective du signal dont on dispose : fenêtre temporelle valant 1 sur l'horizon temporel d'observation (ou d'étude) et 0 en-dehors. Or, les transformées de Fourier analogiques ou numériques d'un produit de signaux correspondent à la convolution des spectres de Fourier de ces deux signaux (Bellanger, 2006 ; Mallat, 2000 ; Millot, 2007).

Pour la fenêtre rectangulaire analogique, le spectre associé est un sinus cardinal multiplié par une exponentielle complexe de phase égale à  $-\pi fT$  où  $T$  correspond à la durée de la fenêtre d'observation

ou d'étude et  $f$  à la fréquence considérée. Pour le spectre du signal de durée infinie, on aurait autant de raies fréquentielles (pics de Dirac fréquentiels) que de fréquences présentes dans le signal et localisées chacune en une de ces fréquences. Aussi, le spectre analogique calculé (par TF) est donc le résultat de la convolution du sinus cardinal déphasé et du spectre de raies de Dirac idéal. Mais, cela se traduit par l'apparition d'une occurrence d'un sinus cardinal centré sur chacune des raies spectrales, situées en  $f_i$ , constituant le spectre idéal avec chacune de ces occurrences de sinus cardinal affectée d'un déphasage différent pour chaque fréquence puisque qu'égal à  $-\pi (ff_i) T$ . Ainsi, en pratique, chacune des répliques de sinus cardinal va interférer avec les autres répliques, conduisant à un spectre effectif très différent du spectre auquel on aurait aimé accéder. De plus, cela se traduit par des masquages

potentiels de la présence de composantes fréquentielles de niveau faible par les lobes secondaires des sinus cardinaux associés aux composantes fréquentielles de fort niveau.

Pour les signaux numériques et leur analyse numérique par TFD, les conséquences sont encore plus gênantes, car on ne calcule pas un spectre continu, mais un spectre discret pour un nombre limité de fréquences. Or, le calcul du spectre associé à une fenêtre rectangulaire numérique s'écarte d'autant plus du sinus cardinal que l'on monte haut en fréquence : le spectre de la fenêtre rectangulaire numérique ne tend vers le sinus cardinal que pour les basses fréquences, car on a un rapport de sinus en lieu et place du rapport entre le sinus et son argument. De plus, puisque le spectre numérique calculé par TFD est discret, les spectres d'amplitude

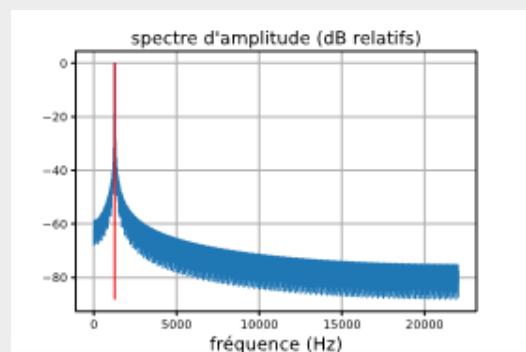
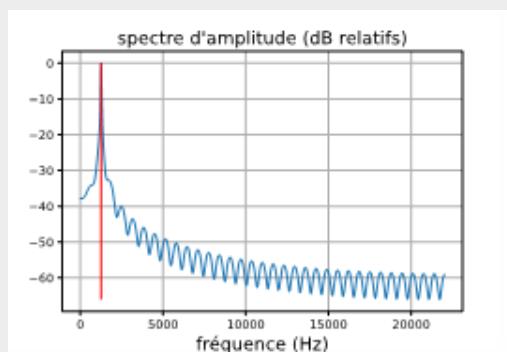


Figure 2 : Illustration de l'influence du nombre de fréquences calculées sur le tracé du spectre d'amplitude de Fourier global pour un signal numérique monochromatique à la fréquence 1261 Hz avec une échelle linéaire pour les fréquences : calcul avec respectivement 1024 (à gauche) et 4096 fréquences (à droite). La ligne verticale rouge correspond à la position effective de l'unique fréquence constituant le signal monochromatique

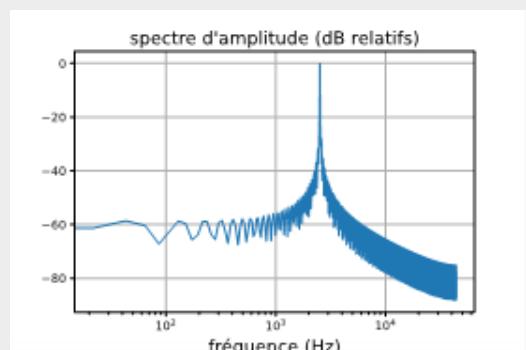
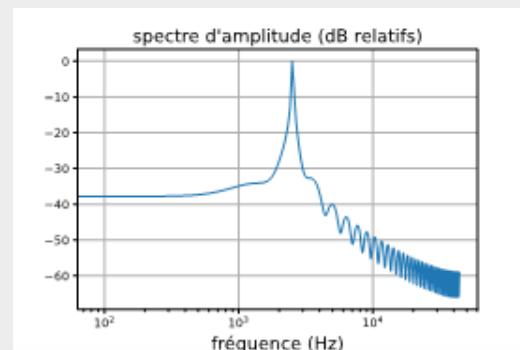


Figure 3 : Illustration de l'influence du nombre de fréquences calculées sur le tracé du spectre d'amplitude de Fourier global pour un signal numérique monochromatique à la fréquence 1261 Hz avec une échelle logarithmique pour les fréquences : calcul avec respectivement 1024 (à gauche) et 4096 fréquences (à droite).

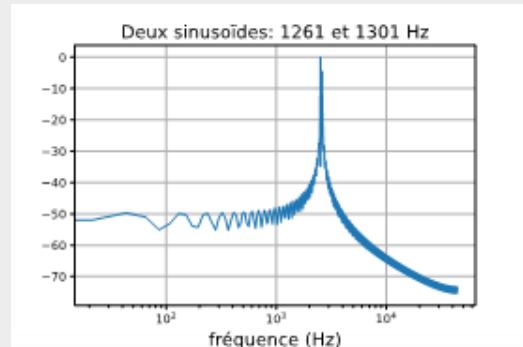
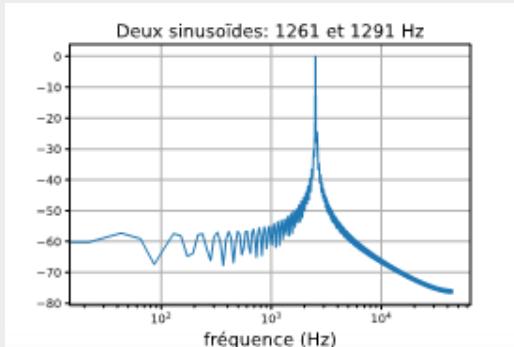
numériques associés sont tracés en reliant les points par des segments de droite, ce qui se traduit par des tracés qui ressemblent d'autant moins à un sinus cardinal que l'on calcule le spectre pour un nombre limité de fréquences.

En Figures 2 et 3, on illustre ce problème de « pollution spectrale » dans le cas d'un signal numérique monochromatique dont on calcule le spectre par TFD avec respectivement 1024 et 4096 fréquences. Pour la Figure 2, les tracés sont effectués en utilisant une échelle linéaire pour l'axe des fréquences et, pour la Figure 3, une échelle logarithmique. On trace de manière systématique des spectres d'amplitudes normalisés, ce qui signifie que l'on trace en décibels

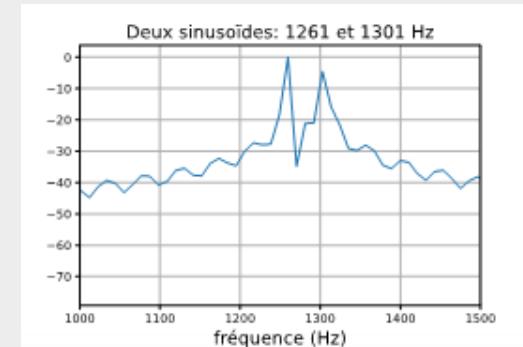
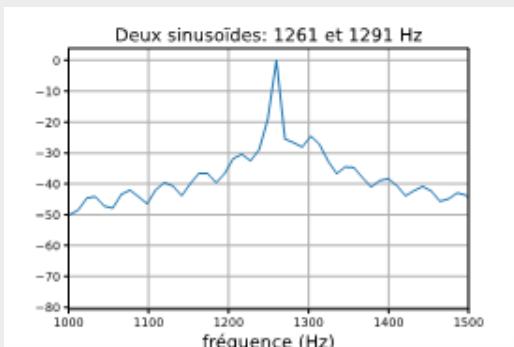
les amplitudes relatives correspondant à la division de l'amplitude par l'amplitude maximale.

La raie tracée en rouge sur les deux spectres de la Figure 1 correspond à la raie théorique centrée sur la fréquence du signal monochromatique, raie que l'on aurait aimé voir apparaître en lieu et place des occurrences de sinus cardinal. Ceci montre qu'il faut apprendre à décrypter les informations présentes dans les spectres de Fourier globaux. Et, pour les Figures 2 et 3, il ne s'agit que d'un signal vraiment très (trop ?) particulier, puisque composé d'une unique fréquence.

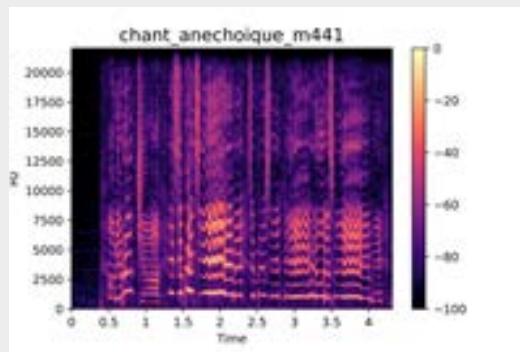
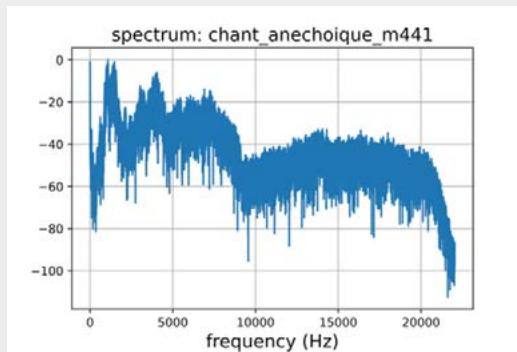
En Figure 4, on illustre le problème de difficulté à déterminer la présence de composantes fréquentielles



**Figure 4 :** Illustration des difficultés potentielles de détection de la présence de deux sinusoïdes dans un signal numérique avec le tracé du spectre d'amplitude de Fourier global (échelle logarithmique pour les fréquences, 4096 fréquences calculées) : pas de détection des deux fréquences séparées de 30 Hz (à gauche) ; indice sur la présence des deux fréquences séparées de 40 Hz (à droite).



**Figure 5 :** Illustration des difficultés potentielles de détection de la présence de deux sinusoïdes dans un signal numérique avec le tracé du spectre d'amplitude de Fourier global (échelle logarithmique pour les fréquences, 4096 fréquences calculées, zoom sur les fréquences comprises entre 1 et 1,5 kHz) : pas de détection des deux fréquences séparées de 30 Hz (à gauche) ; indice sur la présence des deux fréquences séparées de 40 Hz (à droite).



**Figure 6 : Tracé du spectre global d'amplitude de Fourier (à gauche) et du sonagramme (à droite) pour un signal de chant anéchoïque.**

si elles sont trop proches l'une de l'autre, alors que ces deux composantes ont le même poids dans le signal (même amplitude), ce qui correspond à un cas particulier simple. Dans les deux cas, on a utilisé 4096 fréquences pour le calcul des spectres numériques, mais on considère respectivement un premier écart de 30 Hz et un deuxième écart de 40 Hz, et des tracés des spectres globaux pour les fréquences allant de 0 Hz à 22050 Hz (soit la moitié de la fréquence d'échantillonnage ou fréquence de Nyquist). On peut constater que l'observation des spectres globaux permet juste de supposer que l'on a bien deux composantes fréquentielles uniquement dans le cas de l'écart de 40 Hz.

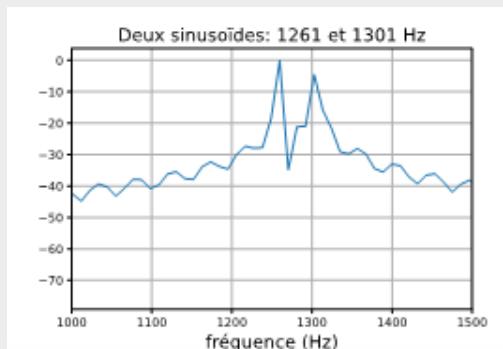
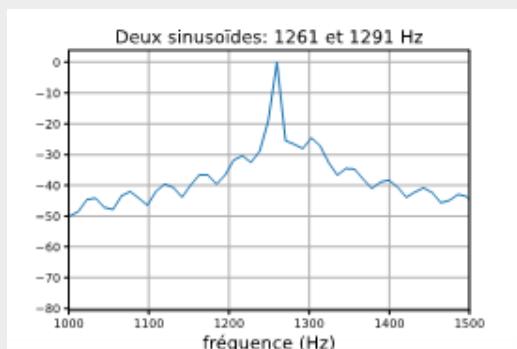
En Figure 5, on a tracé un zoom des spectres de la Figure 4 pour les fréquences allant de 1 à 1,5 kHz. On constate que le spectre global ne permet pas d'identifier clairement la présence de la fréquence la plus élevée dans le cas de l'écart de 30 Hz, puisque le pic associé à cette deuxième fréquence est quasiment masqué. Dans le cas de l'écart de 40 Hz, on voit bien émerger deux pics spectraux, mais il faut être habitué à utiliser ce genre d'outil pour éventuellement affirmer que deux composantes fréquentielles seulement sont effectivement présentes dans la gamme des fréquences étudiées. On constate qu'il faut effectivement faire appel à une analyse paramétrique complémentaire pour pouvoir déterminer avec précision les caractéristiques des deux composantes fréquentielles, mais qu'il

serait vraisemblablement prudent de rechercher plus de composantes fréquentielles avec une analyse paramétrique que ceux que l'on voit apparaître dans le spectre puisque le choix du nombre de fréquences calculées a une incidence directe sur le nombre de composantes fréquentielles émergeant effectivement dans les spectres globaux.

En Figure 6, on illustre ce que donnent les tracés du spectre et du sonagramme dans le cas d'un extrait de chant lyrique enregistré en chambre anéchoïque. Le spectre global a été calculé et est tracé avec un nombre très important de fréquences, ce qui permet de mieux dessiner le spectre en faisant mieux apparaître les lobes secondaires des différentes occurrences de sinus cardinal. Mais, on ne peut que faire le constat que ce spectre comporte beaucoup trop d'informations pour être facilement exploitable surtout qu'on n'obtient qu'une information statistique sur la présence potentielle de chaque fréquence calculée à chaque instant dans le signal temporel (numérique). Le tracé du sonagramme permet de voir apparaître un peu d'information concernant l'évolution au cours du temps du signal étudié. On peut constater la présence en couleurs chaudes, associées aux forts niveaux en décibels, de composantes fréquentielles dans les fréquences graves présentant des ondulations et des pentes variables au cours du temps. Ces petites portions ondulées correspondent à l'utilisation de vibrato par la chanteuse pour des notes vraisemblablement tenues ►

**2**

Sur l'harmonica diatonique, comme il n'y a pas de dispositif de découplage entre les deux anches libres présentes dans le canal d'harmonica, on peut réussir à produire des notes avec des fréquences comprises entre les deux fréquences des notes, soufflée et aspirée normale, en adaptant la forme de la cavité buc cale. Ces notes sont obtenues avec une aspiration sur les canaux 1 à 6, avec une expiration sur les canaux 7 à 10. Les harmonicistes nomment ces notes (altérations) bends car ils ont l'impression de plier la note normale, plus aigüe, pour obtenir une note d'environ un demi-ton en dessous. Sur tous les canaux, sauf le canal 3, les notes normales sont séparées d'un ton donc les bends sont produits à la note associée au demi-ton intermédiaire ; sur le canal 3, on a deux tons qui séparent les deux notes normales, donc on peut produire trois bends pour chacun des 3 demi-tons intermédiaires en modifiant la forme de la cavité buc cale.



**Figure 7 : Influence du choix de l'échelle pour les fréquences sur le tracé du spectre d'amplitude global de Fourier pour un bend aspiré réalisé sur le troisième canal d'un harmonica diatonique (tonalité G), mesure à l'intérieur du canal : tracés avec une échelle linéaire (à gauche) et logarithmique (à droite).**

puisque ces portions ondulées semblent réparties de manière régulière verticalement, et, parce que l'on peut constater qu'on a utilisé une échelle linéaire pour représenter les fréquences. On peut aussi noter la présence d'énergie jusque dans les très hautes fréquences, avec des traces d'ondulations temporelles et une répartition verticale de ces ondulations assez régulière, donc vraisemblablement associées à des harmoniques supérieurs des notes chantées. Mais, pour obtenir une description fine du contenu effectif de ce signal, il faudrait mettre en œuvre une analyse paramétrique basée sur un modèle mathématique de signaux modulés au cours du temps.

La figure 7 correspond au tracé d'une note correspondant à un *bend* aspiré<sup>2</sup> sur le quatrième canal d'un harmonica diatonique en tonalité G (tonalité en sol), signal enregistré avec un microphone de mesure de diamètre 2,54 mm inséré au fond du canal. Cette figure permet de constater la présence d'un nombre conséquent de raies fréquentielles espacées régulièrement en fréquence. Ces raies régulières correspondent aux harmoniques générés par l'écoulement au travers d'une anche libre en amont de celle-ci, haché par les oscillations de l'anche. Cependant, il ne faut pas se fier à la position des raies pour déterminer les fréquences avec précision et il vaudrait mieux mettre en œuvre, une fois de plus, une analyse paramétrique utilisant des sinusoïdes amorties

comme modèle mathématique de signal élémentaire. On peut noter que la composante de niveau le plus important se situe au voisinage de 0 Hz dans le tracé du spectre global avec une échelle linéaire pour les fréquences. En revanche, dans le tracé avec une échelle logarithmique des fréquences, cette raie n'apparaît pas parce que la fréquence nulle n'est pas représentée (puisque'elle pose un problème de logarithme de valeur infinie...). Cela étant, ce tracé en échelle des fréquences logarithmique permet de constater qu'il existe potentiellement de l'énergie sous les 100 Hz et qu'il ne faut pas se limiter à observer seulement ce qui se passe pour les composantes fréquentielles associées aux harmoniques de la fréquence fondamentale. Il s'avère qu'il faut donc « jongler » entre les deux tracés de la Figure 7 pour mieux comprendre les phénomènes à l'œuvre lors de la production de ce *bend*.

Sur la base de ces illustrations, on comprend que l'analyse de Fourier, globale ou à court terme, ne constitue qu'une analyse préliminaire (ou de contrôle ?) qui présente beaucoup d'inconvénients. C'est pour essayer de trouver une réponse à ces problèmes que l'on présente une méthode d'analyse-resynthèse alternative dans la suite.

## Principe de l'analyse-resynthèse IDS

Au vu de l'étude des analyses de Fourier, on peut définir un cahier des charges minimal pour l'analyse recherchée : permettre de conserver une trace des composantes fréquentielles variant sensiblement au cours du temps ; permettre l'écoute des éléments d'analyse seuls sans introduction de distorsion (pas de repliement spectral, par exemple) ; resynthèses partielles à totale sans distorsion.

En cherchant une analyse permettant de respecter ce cahier des charges minimal, nous avons fait le lien avec l'analyse analogique par Intégration de Densité Spectrale (analyse IDS) proposée par Emile Leipp (Leipp, 1977) dans les années 1970. Comme nous allons le voir, l'analyse-resynthèse IDS proposée peut être envisagée, *a posteriori*, comme un croisement entre une version numérique de l'analyse IDS analogique de Leipp et l'analyse par ondelettes.

## Principe de l'analyse IDS

L'analyse proposée consiste à découper le signal étudié en sous-bandes fréquentielles grâce à un banc de filtres complémentaires : *cela signifie que la partie du signal qui ne figure pas complètement dans une sous-bande, parce qu'elle se situe dans la bande de transition entre deux sous-bandes consécutives, figure automatiquement dans la sous-bande suivante et que si l'on additionne toutes les sous-bandes, on resynthétise le signal initial sans aucune erreur en théorie, et aucune erreur perceptive en pratique. Ceci est garanti par le soin apporté au calcul des réponses impulsionnelles des filtres constituant le découpage fréquentiel considéré. Et, s'agissant de ces filtres, on peut choisir librement la*

largeur de la bande de transition, ainsi que le niveau maximum des ondulations dans la bande atténuée. Les filtres constituant un découpage fréquentiel sont des filtres à phase nulle, donc n'introduisant aucune distorsion de phase, passe-bas pour la sous-bande la plus basse, passe-bande pour toutes les autres sous-bandes, sauf la sous-bande la plus haute où on trouve un filtre passe-haut.

En adoptant ce principe (illustré en Figure 8), on peut choisir de mettre en œuvre n'importe quel découpage fréquentiel pour réaliser l'analyse d'un signal ou d'une scène sonore puisqu'il suffit de calculer les signaux en sous-bandes associés à l'utilisation de chacun des filtres de sous-bande pour chacune des voies constituant la scène sonore. Par défaut dans les outils IDS proposés, on trouve trois découpages fréquentiels différents :

découpage « Leipp » (*étendu*) comportant 10 sous-bandes (ajout de la première et la dernière sous-bandes) : 0-50 Hz, 50-200 Hz, 200-400 Hz, 400-800 Hz, 800-1200 Hz, 1200-1800 Hz, 1,8-3 kHz, 3-6 kHz, 6-15 kHz, 15 kHz-fréquence de Nyquist ;

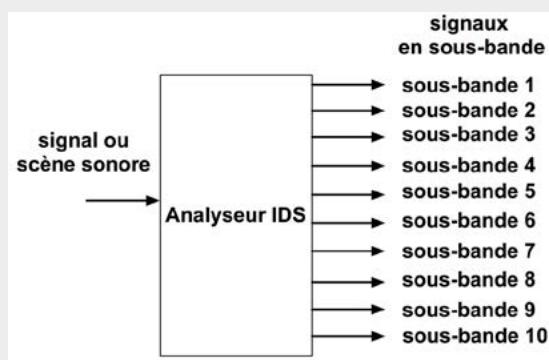


Figure 8 : module d'analyse IDS dans le cas d'un découpage comportant N sous-bandes fréquentielles (ici N est égal à 10). Chaque piste constituant la scène sonore ou le signal temporel est filtré par chacun des N filtres associés au découpage fréquentiel choisi ce qui conduit à la génération de N signaux en sous-bandes.

découpage « Octaves » comportant 11 sous-bandes : 0-22 Hz, 22-44 Hz, 44-88 Hz, 88-176 Hz, 176-353 Hz, 353-707 Hz, 707-1414 Hz, 1414-2828 Hz, 2828-5656 Hz, 5656-11313, 12313-fréquence de Nyquist ;

découpage « Audio » proposé par Pascal Spitz, enseignant à l'ENS Louis-Lumière, comportant 13 sous-bandes : 0-30 Hz, 30-60 Hz, 60-80 Hz, 80-100 Hz, 100-120 Hz, 120-250 Hz, 250-500 Hz, 500-1000 Hz, 1-2 kHz, 2-4 kHz, 4-8 kHz, 8-16 kHz, 16 kHz- fréquence de Nyquist.

À partir des signaux en sous-bandes, que l'on pourra ou non conserver suivant que l'on désire ou non procéder à des opérations de resynthèse, on peut calculer puis tracer ce que l'on nomme un portrait IDS. Ce portrait est constitué du tracé des poids relatifs de chacune des sous-bandes correspondant à la valeur exprimée en décibels relatifs du rapport entre l'énergie cumulée (somme des carrés des échantillons du signal) pour le signal de sous-bande et l'énergie cumulée pour le signal étudié (ou la voie pour une scène sonore). On peut alors tracer la courbe des poids relatifs en fonction de la sous-bande fréquentielle. Le portrait IDS est complété par les indications du niveau moyen du signal (ou de la voie) données en dB FS, ainsi que par le tableau des poids relatifs (au besoin pour chaque voie).

En Figure 9, on illustre le tracé du portrait IDS pour les trois découpages fréquentiels proposés par défaut dans le cas du signal de *bend* aspiré sur le canal 4 d'un harmonica diatonique (tonalité : G) considéré pour le tracé des spectres de la Figure 7.

Le choix du découpage fréquentiel modifie bien évidemment un peu l'allure des portraits IDS, mais cette description globale se révèle très synthétique et elle semble de nature à mieux restituer ce qui se passe du point de vue de la répartition de l'énergie pour le signal étudié. En effet, on retrouve pour chacun des trois portraits IDS, une séparation en trois zones. La première zone correspond à la première sous-bande où la majeure partie de l'énergie se trouve concentrée, ce qui explique que le tracé des évolutions temporelles ne soit pas centré autour de la pression nulle, mais adopte une forme unipolaire (cf. Figure 10) : les phénomènes physiques à l'œuvre sont des écoulements sonores, pas des ondes acoustiques (Millot et Baumann, 2007). La deuxième zone s'étend jusqu'à la sous-bande où on trouve la fréquence fondamentale (environ à 415 Hz, car la note émise est très proche d'un la bémol), et il n'y a pas ou peu d'énergie dans cette zone fréquentielle. La troisième zone correspond au contenu supérieur du spectre où se déploient les fréquences audibles

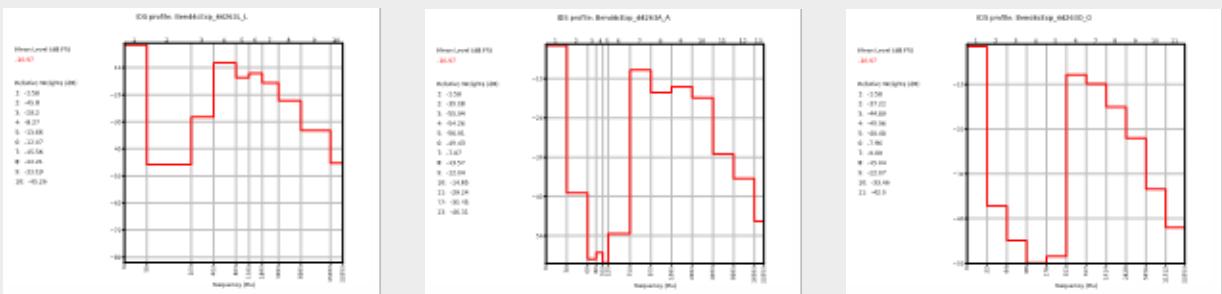
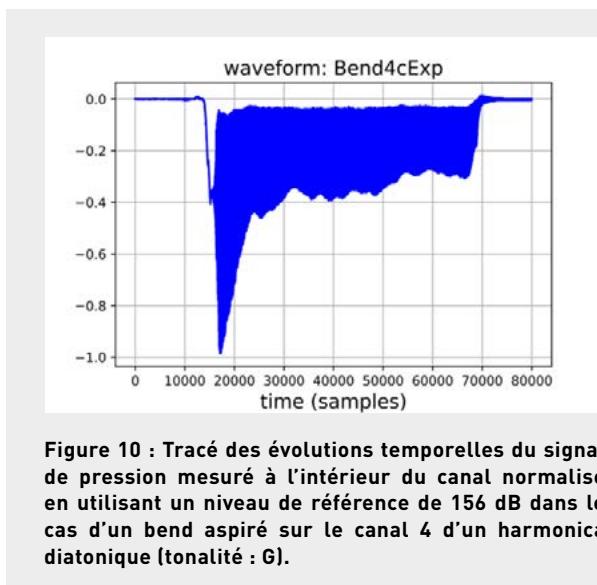


Figure 9 : Illustration de portraits IDS pour les trois découpages fréquentiels proposés par défaut dans le cas d'un bend aspiré sur le canal 4 d'un harmonica diatonique (tonalité : G) : Leipp étendu (à gauche), Audio (au milieu) et Octaves (à droite).

associées à ce signal : la fréquence fondamentale, ainsi que tous les harmoniques jusque quasiment 20 kHz. On peut éventuellement distinguer deux sous-zones dans cette troisième zone : une première sous-zone avec les gammes de fréquences présentant un niveau plus conséquent d'énergie jusqu'à environ 4 ou 6 kHz ; une deuxième sous-zone au-dessus de 4 à 6 kHz, où les harmoniques concentrent sensiblement moins d'énergie.

Ainsi, on peut déjà proposer que l'observation des portraits IDS, plus synthétiques, permet de distinguer des informations qui sont difficilement visibles si on étudie seulement le tracé des spectres globaux. Mais, l'analyse de Fourier globale est complémentaire du tracé des portraits IDS en ce sens qu'elle permet de constater la présence de raies spectrales régulièrement espacées donc *a priori* en rapport harmonique. Cela étant, pour s'en assurer, il faudrait aussi mettre en œuvre une analyse paramétrique, mais, cette fois-ci, elle pourrait être menée sous-bande fréquentielle par sous-bande fréquentielle, ce qui pourrait s'avérer plus simple. En fait, on voit apparaître l'idée classique consistant à « *diviser pour mieux régner* », donc découper le problème d'étude en sous-problèmes, pouvant être constitués ici d'une sous-bande ou d'un regroupement de sous-bandes au besoin, donc de sous-problèmes *a priori* plus faciles à résoudre.

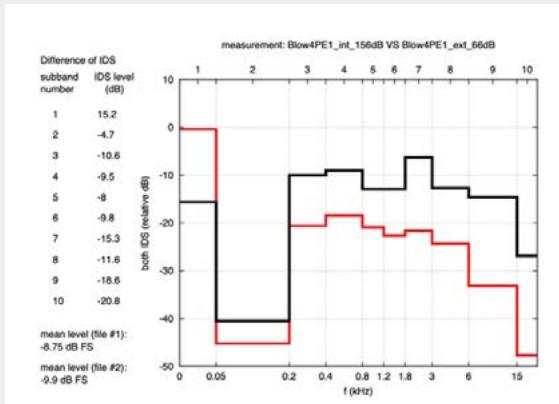
En Figure 11, on a tracé sur la même figure les portraits IDS des signaux de pression mesurés à l'occasion d'une note soufflée (*blow*) sur le canal 4, respectivement à l'intérieur (courbe en rouge) et à l'extérieur (courbe en noir) du canal d'harmonica diatonique, à 10 cm en aval de l'instrument. Ces signaux ont été respectivement normalisés pour pouvoir être exportés au format WAV avec des niveaux de 156 dB SPL pour la mesure



**Figure 10 : Tracé des évolutions temporelles du signal de pression mesuré à l'intérieur du canal normalisé en utilisant un niveau de référence de 156 dB dans le cas d'un bend aspiré sur le canal 4 d'un harmonica diatonique (tonalité : G).**

à l'intérieur du canal et 66 dB SPL pour la mesure à l'extérieur du canal. Sur le portrait IDS, les indications numériques données sur la gauche correspondent à la différence entre les poids relatifs des deux mesures pour chaque sous-bande ainsi qu'aux niveaux moyens en dBFS pour ces deux sons normalisés.

On peut ainsi constater que les niveaux moyens des signaux normalisés sont assez proches mais, si on prend en compte la différence entre les deux niveaux utilisés pour la normalisation, égale à 90 dB SPL, on constate qu'il existe une atténuation de niveau de l'ordre de 90 dB lorsqu'on passe de l'intérieur à l'extérieur du canal de l'harmonica. Au travers de l'exemple de ces deux signaux, on trouve un exemple de la justification de l'hypothèse commune, lorsqu'on travaille sur la modélisation par modèle physique des instruments de musique à vent, qui consiste à considérer que la pression en sortie de l'instrument (ou du pavillon terminal, par exemple) est égale à la pression atmosphérique : pour étudier les conditions ►



**Figure 11 : Comparaisons de portraits IDS pour une note soufflée sur le 4ème canal d'un harmonica diatonique (tonalité G) : signal de pression mesuré à l'intérieur du canal (en rouge) normalisé en utilisant un niveau de référence de 156 dB SPL ; signal de pression mesuré à l'extérieur de l'instrument (en noir) à une dizaine de centimètres en aval et dans l'axe en utilisant un niveau de référence de 66 dB SPL.**

correspond toujours à une zone où il y a peu d'énergie dans le signal soit entre 50 Hz et une fréquence inférieure à la fréquence fondamentale de la note produite. On observe que le poids relatif de l'énergie dans la première zone, entre 0 et 50 Hz, diminue sensiblement de 15 dB environ. Et, on peut noter une augmentation du poids relatif des sous-bandes fréquentielles où se trouvent la fréquence fondamentale et les harmoniques, ce qui signifie qu'il se produit un rééquilibrage de l'énergie (donc de la balance spectrale), certes accompagné d'une perte conséquente de niveau, au profit des fréquences correspondant à la partie audible du signal de pression rayonné. On observe aussi qu'il n'y a plus, pour la pression rayonnée, de sous-zone fréquentielle au-dessus de 6 kHz bénéficiant de moins d'énergie. En effet, dans le signal de pression externe, on ne trouve plus que la sous-bande des fréquences supérieures à 15 kHz qui se trouve sensiblement en retrait par rapport au reste de cette troisième zone. Cette remontée du contenu fréquentiel dans la sous-bande 6-15 kHz semble correspondre au fait que, à l'écoulement au travers de l'anche libre qui se présente sous la forme d'un jet libre intermittent et donc d'une concentration spatiale comme fréquentielle en basse fréquence de l'énergie, succède un écoulement bruité associé à la dissipation par turbulences du jet libre d'air dans l'air environnant.

de fonctionnement de l'instrument, et donc de production des notes de musique, on suppose que l'instrument ne produit aucun son, qu'il n'y a pas de rayonnement de pression acoustique. En fait, pour l'étude des instruments à vent, on sépare l'étude en deux problèmes : l'étude des conditions de fonctionnement à l'intérieur de l'instrument avec l'hypothèse de l'absence de rayonnement ; l'étude du rayonnement acoustique à l'extérieur de l'instrument, construite en exploitant comme source acoustique les signaux (ou expressions analytiques quand c'est possible) de flux volumique s'échappant de l'instrument.

Ces quelques illustrations de portraits IDS permettent de montrer l'intérêt de considérer une analyse telle que celle proposée, puisqu'elle permet une analyse et une compréhension plus fine des phénomènes physiques à l'œuvre, complétant les autres données expérimentales telles que des visualisations de l'écoulement au travers des anches libres (par la méthode dite de « *Schlieren visualization* » par exemple). L'analyse par tiers d'octave, cas particulier de transformée de Fourier

En étudiant les modifications de balance spectrale pour la pression entre l'intérieur (courbe rouge) et l'extérieur (courbe noire), on constate que la séparation observée en trois zones pour la pression interne subsiste pour la pression externe. On constate que la deuxième zone

discrète mise en œuvre dans les analyseurs de spectres numériques, ne permet qu'une description approchée, si on respecte les préconisations sur la conception des filtres d'analyse (norme ISO 266 :1997), car ils ne sont pas complémentaires et introduisent des distorsions de phase puisqu'il s'agit de filtres récurrents<sup>3</sup>. En utilisant des filtres complémentaires à phase nulle, on pourrait donc envisager d'utiliser un découpage fréquentiel en tiers d'octave, mais ce découpage comporterait plus de sous-bandes fréquentielles que le découpage en octaves proposé par *défaut*, on aurait donc un découpage probablement trop lourd du point de vue des calculs puisqu'il faut procéder à un filtrage coûteux par sous-bande fréquentielle.

L'analyse globale de Fourier permet de compléter l'étude dans les cas présentés, puisqu'elle permet de mettre en évidence une structuration potentiellement harmonique du spectre pour les notes émises. Mais l'étude et le tracé des *évolutions temporelles* des signaux de sous-bandes permettraient probablement d'envisager cette structuration, puisque les signaux en sous-bandes ont une largeur fréquentielle réduite et qu'ils comportent donc un nombre limité de fréquences harmoniques pour toutes les sous-bandes de fréquences supérieures à 200 Hz. Et, pour obtenir une caractérisation fine des composantes fréquentielles jouant un rôle dans le signal étudié, il semble encore nécessaire d'envisager une analyse complémentaire paramétrique reposant sur un modèle mathématique de sinusoides amorties, vraisemblablement modulées en amplitude, voire en fréquence, pour tenir compte des vibratos des notes produites et des variations de fréquence fondamentale potentielles. Mais, même s'il faut envisager plusieurs analyses paramétriques, jusqu'à une par sous-bande intéressante, ces analyses paramétriques sont probablement plus simples à mener,

puisque elles concernent des signaux en sous-bandes comportant un nombre très réduit de fréquences. On peut donc considérer l'analyse des portraits IDS comme une analyse préliminaire (voire de contrôle), mais qui semble, pour autant, donner plus d'informations que l'analyse d'un spectre global. Et, on peut au besoin opter pour un autre découpage fréquentiel comportant plus (ou moins) de sous-bandes pour l'adapter au signal étudié.

### Principe de la resynthèse IDS

À partir des signaux en sous-bandes, qui ne font pas l'objet d'un sous-échantillonnage contrairement à ce qui est pratiqué dans le cas des analyses ou transformées en ondelettes (notamment rapides), on peut *étudier les évolutions temporelles* de ces signaux en les traçant, mais on peut aussi procéder à des resynthèses totales ou partielles suivant les sous-bandes que l'on choisit de considérer pour la resynthèse. Puisque la phase d'analyse a été réalisée avec des filtres complémentaires à phase nulle, la resynthèse correspond à la sommation simple des signaux des sous-bandes retenues. Et, si on a calculé et stocké au préalable ces signaux de sous-bandes, on peut procéder à des resynthèses et modifier les sous-bandes retenues en activant ou désactivant simplement les sous-bandes. Ainsi, l'étape de resynthèse correspond simplement à la sommation, donc au mixage, des signaux de sous-bande.

On peut donc choisir d'écouter les éléments d'analyse, un par un, comme choisir de supprimer ou d'ajouter à la resynthèse une sous-bande donnée. Ce faisant, partant de la prise en compte de toutes les sous-bandes dans la resynthèse, on peut tester les sous-bandes que l'on peut supprimer sans que cela n'introduise de modification ►

conséquence par rapport au signal original, qui peut être écouté par simple bascule à la volée avec le signal resynthétisé. On peut, sinon, procéder à l'inverse en commençant par prendre en compte la ou les sous-bandes les plus importantes perceptivement (ou au contraire objectivement) dans le signal original afin de constituer une resynthèse pouvant être considérée comme une approximation satisfaisante du son. En fait, de nombreuses approches sont possibles.

Puisque la resynthèse s'opère par simple mixage ou sommation des signaux de sous-bande, on peut, partant d'une resynthèse totale, envisager d'augmenter ou de réduire le poids d'une sous-bande afin d'étudier la sensibilité perceptive par rapport à un changement du poids de cette sous-bande. On peut aussi envisager une resynthèse partielle en modifiant le poids relatif des sous-bandes retenues pour modifier ou sculpter différemment le signal, principe qui a été utilisé dans le cadre de deux études préliminaires des problèmes d'audition pour les personnes équipées d'aides auditives, mais aussi pour modifier le son restitué au casque dans le cas de la pratique de la trompette avec une sourdine (Armant, 2018).

## Exemples d'utilisations déjà mises en œuvre

Depuis que nous avons commencé à travailler sur la mise en œuvre de cette analyse alternative (Millot, 2004 ; Millot, 2007), nous avons eu l'occasion de vérifier son intérêt au travers de son utilisation dans des mémoires de Master de l'ENS Louis-Lumière, comme de travaux de recherche personnels.

Comme évoqué dans la section précédente, l'analyse-

resynthèse IDS a été au cœur de travaux de mémoire pour améliorer le son diffusé au casque dans le cas du jeu sur une trompette équipée d'une sourdine (Armant, 2018). L'objectif était de permettre une pratique quotidienne de l'instrument, *même quand on ne dispose pas d'un accès à une pièce ou un studio de répétition et qu'on doit travailler son instrument dans un environnement urbain*, dans des pièces ne bénéficiant pas d'une isolation acoustique particulière par rapport au voisinage.

On a aussi utilisé l'analyse-resynthèse comme principe central pour deux mémoires qui ont porté sur l'étude de l'intérêt de proposer, à terme, aux personnes équipées d'aides auditives de choisir le profil spectral du son diffusé dans leurs aides auditives et d'adapter ce profil spectral en fonction de l'ambiance sonore dans laquelle elles se trouvent (Ralle, 2017 ; Ricciuti, 2019). Il s'agissait de permettre de changer les poids relatifs des sous-bandes retenues pour améliorer le confort d'écoute, augmenter l'intelligibilité, diminuer la gêne auditive comme la fatigue auditive, en permettant à chaque personne équipée de définir des pré-configurations adaptées aux ambiances sonores entre lesquelles elle a l'habitude de « naviguer », voire d'adapter en temps réel une pré-configuration à une nouvelle ambiance sonore. Ce travail, pour être poursuivi, requiert des partenariats avec des acteurs industriels (fabricants d'aides auditives) comme des acteurs médicaux (chercheurs travaillant sur les problèmes d'audition), sans oublier un nombre suffisant de personnes concernées par les problèmes d'audition et donc bénéficiaires potentiels de ces travaux.

Toujours dans le cadre d'un mémoire de fin d'études, l'analyse-resynthèse IDS a été utilisée pour l'étude de signaux cardiaques et pulmonaires enregistrés avec

un stéthoscope numérique (Dellaby, 2011). Malgré la petite taille du corpus de signaux disponible à l'époque, on a pu constater que l'analyse-resynthèse IDS permet la mise en évidence de particularités associées à des pathologies cardiaques ou pulmonaires en procédant à l'écoute de resynthèses très partielles.

Cette analyse-resynthèse constitue aussi un outil permettant d'aller chercher des signaux élémentaires dans des scènes ou ambiances sonores complexes (Carr-Brown, Colcy et Delatte, 2004 ; Millot *et al*, 2005 ; Millot et Pelé, 2007) ou d'étudier la variabilité d'ambiances sonores (Debeugny, 2021). Elle a aussi été utilisée pour réaliser des classifications objectives de guitaristes (Navarret, 2006, Jamain, 2011) quand on a constaté que l'analyse d'enregistrements, *réalisés sur trois prototypes de guitares électriques, avec des essences de bois différentes pour le corps de la guitare*, a révélé non pas des différences de jeu des guitaristes sur les guitares, mais que ceux-ci réussissaient à adapter leur jeu sur les trois guitares pour retrouver la couleur sonore qu'ils désiraient et que l'on parvenait à regrouper objectivement les guitaristes à partir des portraits IDS de leurs enregistrements (Navarret, 2006) !

L'analyse-resynthèse IDS a aussi été utile dans le cadre de travaux ou de mémoires portant sur l'étude des phénomènes de proximité pour les sources et les microphones et/ou la modélisation physique des microphones (Valette, 2005 ; Millot *et al*, 2006 ; Millot *et al*, 2007 ; Lescure, 2011 ; Llave, 2015). L'analyse-synthèse IDS est aussi l'outil central de travaux ayant porté sur le comportement des salles et leur caractérisation objective (Millot et Pelé, 2009 ; Millot, 2009 ; Millot, 2013-2016), travaux ayant donné lieu à deux *mémoires de fin d'études* (Hendrickx et Derode, 2010 ; Hémerly et Mallebay, 2013). Une étude (Lemée, 2017) sur le transport du mixage son de l'auditorium cinéma vers la salle de

cinéma a été menée et elle a montré que l'analyse-resynthèse IDS pourrait être utilisée pour caractériser et entendre un aperçu des distorsions introduites lors de ce transfert.

## Outils d'analyse IDS

Pour le moment, il est possible de télécharger des outils d'analyse-resynthèse IDS se présentant sous forme d'applications *standalone* tournant sous Mac OS 10.10, ainsi que trois resynthétiseurs tournant sous Pure Data, associés aux trois découpages fréquentiels proposés, depuis le site [www.autopsience.com](http://www.autopsience.com). De plus, des vidéos en français et en anglais pour installer et utiliser ces outils sont disponibles sur la chaîne Autopsience sur YouTube<sup>4</sup> dont on trouve un lien direct sur le site [www.autopsience.com](http://www.autopsience.com). Une nouvelle version des outils testés sous Mac OS 10.15.6 et Windows 10 sera disponible d'ici juillet 2022, accompagnée de resynthétiseurs fonctionnant sous Max 8.

À ce stade, les outils développés en Python 3 sont les suivants :

- analyseur IDS avec 3 découpages proposés par défaut (« *Leipp* », « *Audio* » et « *Octaves* ») ;
- calculateur de gradients et bi-gradients IDS par lot ;
- traceur de portraits, gradients et bi-gradients IDS par lot ;
- analyseur IDS préparant toutes les données nécessaires (signaux originaux et de sous-bandes, portraits IDS associés) pour l'utilisation des resynthétiseurs sous Max et rangeant automatiquement ces données dans les sous-dossiers utilisées par les resynthétiseurs. ►

<sup>4</sup> [www.youtube.com/channel/UC7MBULHthydNcSoUHDYTCZg](https://www.youtube.com/channel/UC7MBULHthydNcSoUHDYTCZg)

## Conclusions et travaux futurs

Dans cet article, nous avons expliqué pourquoi un autre outil d'analyse est nécessaire pour étudier les signaux temporels, de quelque nature physique qu'ils soient. L'outil proposé permet notamment la resynthèse partielle à complète des signaux temporels étudiés, et donc l'écoute de ces signaux resynthétisés, contrairement aux analyses de Fourier. Nous avons dressé un premier panorama des applications déjà réalisées de cet outil qui se révèle très large et concerne des domaines d'application assez nombreux, tant dans les métiers du son, de la musique, en physique, que dans le domaine *médical*, surtout que l'on peut aussi procéder à des resynthèses incluant des modifications du poids des différents signaux en sous-bandes fréquentielles.

En 2022, deux mémoires utiliseront l'analyse-resynthèse IDS : une étude sur l'identification et la classification des effets vocaux extrêmes dans le Métal ; une étude sur la détermination du genre des voix des adolescents pré-pubères. Et, nous devrions reprendre l'étude sur l'analyse et la classification de signaux cardiaques en vue d'observer si on peut envisager une identification, voire une aide pour poser un pré-diagnostic en identifiant des pathologies potentielles sur la base des portraits IDS.

Il semble que cet outil pourrait être employé *comme une aide à la formation à l'écoute au stéthoscope des personnels médicaux*. Et, sur la base d'études de classification, que nous avons menées depuis sur d'autres corpus de signaux, nous pensons qu'il serait possible d'étudier la pertinence de cet outil pour établir ou guider des pré-diagnostic de pathologies, dès lors qu'une analyse préalable sur un corpus de taille suffisante pourra être menée. Dans le cadre d'un travail s'appuyant vraisemblablement sur l'intelligence artificielle, il serait intéressant de vérifier si cet outil peut servir

pour aider les personnels médicaux opérant dans des conditions extrêmes (médecine de catastrophe ou de guerre) ou pour résoudre, pour partie, les problèmes des déserts médicaux. Il pourrait ainsi être possible de procéder à des pré-diagnostic sans spécialiste présent sur le terrain mais avec juste un personnel médical disposant d'une formation complémentaire adéquate, afin d'aider à déterminer les priorités ou à confirmer, par exemple, les suspicions de pathologies graves. Ce genre d'outil pourrait d'ailleurs être utilisé dans d'autres contextes médicaux où on procède à l'enregistrement de mesures temporelles tels que, par exemple, les électro-encéphalogrammes, les électrocardiogrammes, les *électroglottogrammes* (pour l'étude de la voix) ou encore pour l'analyse des tests de conduction nerveuse, tels que les électromyogrammes.

Parmi les applications envisagées mais qui n'ont pas été testées pour le moment, on peut citer, par exemple :

- le suivi et la détection sonores d'usure de pièces mécaniques ou d'outils d'usage ;
- l'aide à l'apprentissage des langues maternelles comme étrangères ;
- la rééducation à la parole pour les personnes ayant subi des opérations lourdes de l'appareil phonatoire ;
- la formation de l'oreille pour les musiciens, les acteurs ou les comédiens, conférenciers, les ingénieurs du son, voire des gens comme les « oreilles d'or » dans les sous-marins ;
- l'analyse prédictive de cours d'actions cotées en bourse.

L'extension de l'analyse-resynthèse IDS aux images fixes et animées devrait être étudiée à nouveau en 2022 avec Alain Sarlat, enseignant de l'école Louis-Lumière, Master Photographie.

Enfin, le travail sur le portage depuis Matlab vers Python 3, ainsi que la finalisation d'un outil de détermination quasi-automatique d'un découpage fréquentiel adapté à un corpus de signaux temporels donnés devraient aussi être continués en 2022, ce qui permettra de tester encore plus de signaux temporels de nature physique différente.

## Références bibliographiques

- Armant, A. (2018), *Sourdine muette pour trompette : vers un modèle plus adapté*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Bellanger, M. (2006), *Traitement numérique du signal : théorie et pratique*, Collection Sciences Sup, Dunod.
- > Carr-Brown, A., Colcy, M., Delatte, N. (2004), *Description spatiale d'enregistrement stéréophonique*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Delaby, F-X. (2011), *Écoute au stéthoscope, description et analyse d'enregistrements de « bruits » physiologiques et pathologiques, perspectives en vue d'une application pédagogique*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Debeugny, T. (2021), *Influence de l'audio dans la perception de la variabilité des environnements virtuels explorables en jeux vidéo*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Derode, L., Hendrickx, E. (2010), *La pertinence de l'analyse par intégration de densité spectrale numérique dans la caractérisation objective de l'acoustique des salles*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Hémerly, E., Mallebay, F. (2013), *Mise au point du cahier des charges du logiciel SABRE et des protocoles de mesures pour la caractérisation d'installations sonores par une méthode alternative basée sur l'analyse IDSE*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Jamain, C. (2011), *Classification objective de guitaristes par analyse IDS*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Leipp, É. (1977), *L'intégrateur de densité spectrale (IDS) et ses applications, bulletin N°94 du GAM (Groupe d'Acoustique Musicale)*, Université Paris Pierre et Marie Curie, 19.12.1977.
- > Lemée, G. (2017), *Étude et simulation du transport entre l'auditorium de mixage et la salle de cinéma*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Lescure, J. (2011), *Simulation temps réel de prise de son multicanale*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Llave, A. (2015), *Simulation et modélisation de l'éloignement d'une source*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Mallat, S. (2000), *Une exploration des signaux en ondelettes*, Les éditions de l'École Polytechnique, Palaiseau.
- > Max, J., Lacoume, J-L. (1996), *Méthodes et techniques de traitement numérique du signal et applications aux mesures physiques – 1. Principes généraux et méthodes classiques*, Collection Enseignement de l'Électronique, Masson, Paris.
- > Meyer, Y. (1992), *Ondelettes et algorithmes concurrents*, Hermann, Paris. ►

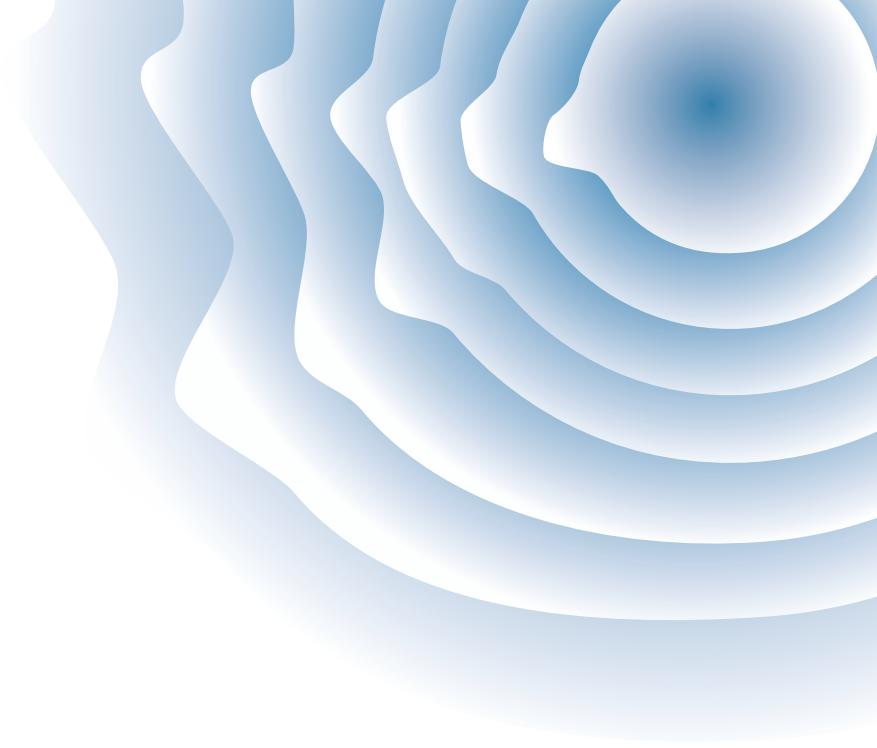
- > Millot, L. (2004), Some clues to build a sound analysis relevant to hearing, *116th Convention of the Audio Engineering Society*, Berlin, Allemagne, 8-11 mai 2004, Paper 6041.
- > Millot, L., Pelé, G., Eliqu, M. (2005) Using perceptive subbands analysis to perform audio scenes cartography, *118th Convention of the Audio Engineering Society*, Barcelone, Espagne, 28-31 mai 2005, Paper 6340.
- > Millot, L., Valette, A., Lopes, M., Pelé, G., Eliqu, M., Lambert, D. (2006), Listening broadband physical model for microphones: a first step, *120th Convention of the Audio Engineering Society*, Paris, France, 20-23 mai 2006, Paper 6638.
- > Millot, L., Baumann, C. (2007), A proposal for a minimal model of free reeds, *Acta Acustica united with Acustica*, 93(1), pp. 122-144.
- > Millot, L. (2007), *Traitement du signal audiovisuel - Applications avec Pure Data*, Collection Audio-Photo-Vidéo, Dunod/ENS Louis-Lumière.
- > Millot, L., Lopes, M., Eliqu, M., Pelé, G., Lambert, D. (2007), Revisiting proximity effect using broadband signals, *122th Convention of the Audio Engineering Society*, Vienne, Autriche, 8-10 mai 2007, Paper 7106.
- > Millot, L., Pelé, G. (2007), An objective and subjective alternative audio sounds and scenes analysis: the IDS, *International Symposium on Musical Acoustics (ISMA 2007)*, Barcelone, Espagne, 9-12 Septembre 2007.
- > Millot, L., Pelé, G. (2009), *Méthodes de caractérisation des dispositifs ou installations audio : critiques et propositions alternatives / Protocole de caractérisation objective d'une installation audio*, Pôle de Compétitivité « Cap Digital » projet HD-3D Initiative pour une Industrie Ouverte (HD-3D IIO) / ENS Louis-Lumière.
- > Millot, L. (2009), Why using complex stimuli for acoustical measurements may be necessary to get physically relevant analysis results, *ASA's 158th meeting (American Society of Acoustics)*, San Antonio, USA, 26-30 octobre 2009.
- > Navarret, B. (2006), *Etude de l'influence des éléments de lutherie de la guitare électrique « solid body »*, mémoire de fin d'études de la classe d'Acoustique Musicale, CNSMDP.
- > Ralle, T. (2017), *Usage de l'analyse IDS pour une meilleure compréhension de la parole dans le bruit*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Ricciuti, G. (2019), *Utilisation de l'Intégrateur de Densité Spectrale (IDS) pour améliorer la compréhension de la parole dans le bruit pour les personnes équipées d'aides auditives*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).
- > Truchetet, F. (1998), *Ondelettes pour le signal numérique*, Collection traitement du signal, Hermes, Paris.
- > Valette, A. (2005), *Modèles physiques de microphones*, Mémoire de Fin d'Études de l'ENS Louis-Lumière (section Son).

## LAURENT MILLOT

Laurent Millot, ingénieur Arts et Métiers (ENSAM, 1991-1994), est titulaire d'un Master 2 Acoustique, Traitement du signal, Informatique Appliqués à la Musique (ATIAM, Sorbonne Université Paris, 1995) et d'un Doctorat en Mécanique/Acoustique (Sorbonne Université Paris, 1999). Il enseigne le traitement du signal et l'acoustique à l'ENS Louis-Lumière depuis octobre 1996, et il y est Maître de Conférences en traitement du signal depuis septembre 2010. Il enseigne aussi le traitement du signal depuis 1998 au sein de la Formation Supérieure aux Métiers du Son (Conservatoire National Supérieur de Musique et Danse de Paris). Ses travaux actuels se situent à la croisée du traitement du signal, de l'acoustique et, notamment, de la mécanique quantique. Sur la base des travaux réalisés pendant son Master 2 ATIAM et la préparation de son Doctorat, il travaille à une reformulation de l'acoustique qui a déjà débouché sur une formulation alternative pour la mécanique quantique, formulations théoriques qui seront discutées ailleurs.

*Laurent Millot, an Arts et Métiers engineer (ENSAM, 1991-1994), holds a Master's degree in Acoustics, Signal Processing and Computer Science applied to Music (ATIAM, Sorbonne University Paris, 1995) and a PhD in Mechanics/Acoustics (Sorbonne University Paris, 1999). He has been teaching signal processing and acoustics at ENS Louis-Lumière since October 1996, and is a lecturer in signal processing since September 2010. He also teaches signal processing since 1998 at the Formation Supérieure aux Métiers du Son of the CNSMDP. His current work is at the crossroads of signal processing, acoustics and, notably, quantum mechanics. Based on the work done during his Master 2 ATIAM and the preparation of his PhD, he is working on a reformulation of acoustics which has already led to an alternative formulation for quantum mechanics, theoretical formulations which will be discussed elsewhere.*





# Notes de recherches

## Les enjeux de la transition dans la composition des paysages sonores

Hugo Corbel,  
Master Son 2021, ENS Louis-Lumière

L'aube dans un environnement naturel est un processus très complexe en matière de son. Les mécanismes en jeu lors de cette période sont encore peu connus, bien que le phénomène soit familier à beaucoup de personnes vivant en lisière de forêt. C'est un processus de transition entre deux moments, à savoir la nuit et le jour. Le concept de transition est très présent dans les enregistrements et les compositions de paysages sonores. Les enjeux de ce mémoire sont de déterminer comment faire ressentir cette transition dans une composition à base de paysages sonores et d'en rendre compte dans une installation en ambisonie pour que l'expérience soit immersive.

En premier lieu, nous nous sommes intéressé au processus de transition, à ses rôles dans les compositions à base de paysages sonores et aux différents procédés de composition mis en œuvre pour la faire ressentir. Par l'analyse de la pièce *La Selva* de Francisco López, nous avons déterminé qu'il existait trois grands groupes de transitions différentes, rassemblant sept types de transitions. Tout d'abord les transitions fréquentielles, qui impliquent des modifications ou non dans la répartition fréquentielle. Elles se décomposent en deux types de transitions : les transitions avec élément constant (un élément sonore continu va rester tout au long de la transition) et les transitions par changement de répartition fréquentielle (la coloration spectrale avant et après la transition sera différente). Ensuite, les transitions en intensité qui reposent sur une impression de changement d'intensité sonore. On y retrouve trois types de transitions : les transitions

par changement de niveau sonore global (on passe d'un niveau global faible à un niveau global fort, ou inversement), les transitions par changement d'intensité d'un élément géophonique (changement de dynamique d'un élément géophonique, soit le vent ou l'eau) et les transitions par changement d'intensité d'un élément anthropophonique (changement de dynamique d'un élément anthropophonique, soit des sons produits par l'activité humaine). Enfin, on retrouve les transitions rythmiques, qui reposent sur un changement de rythme dans la pièce. Dans ce groupe se retrouvent les transitions par déplacement d'une espèce animale (on tire parti de l'activité animale afin de faire sentir un déplacement dans l'espace) et les transitions par déplacements d'individus (qui repose sur la présence d'êtres humains dans la scène sonore, qui discutent, chantent ou se déplacent).

Dans la suite du mémoire, nous nous sommes demandé comment utiliser les principes précédents afin de créer une pièce sonore reposant sur une transition naturelle, celle de la nuit au jour, que nous avons nommée *L'Heure Bleue*. Nous avons vu quelles étaient les caractéristiques sonores principales d'une forêt à l'aube, puis comment utiliser ces spécificités pour la composition de la pièce. Ainsi, le lever du jour se décompose en trois phases. Premièrement la nuit, dont la sensation principale est le silence. Il y a très peu d'espèces animales actives, et les sons qu'elles émettent sont très éparés. Ensuite vient le chœur matinal, qui est un phénomène qui se produit au lever du jour, pendant lequel les oiseaux commencent à chanter tous en même temps. Il commence entre 30 et 90 minutes avant le lever du soleil et s'arrête brusquement quand l'intensité lumineuse augmente. C'est un facteur déterminant dans le mécanisme sonore de transition de la nuit au jour puisque c'est un des seuls facteurs sonores qui évolue dans le temps lors de l'aube. Enfin, la dernière étape est le jour. Lorsque

le soleil est levé, on assiste à une accalmie des oiseaux, qui se mettent à communiquer plus directement. La vie humaine commence vraiment, il y a donc davantage de sons de circulations ou d'activités extérieures. Ainsi, afin de retransmettre l'expérience sonore de la transition de la nuit au jour, il convient de tenir compte de tous ces paramètres et de les incorporer dans la composition, le chorus matinal jouant un rôle déterminant dans cette transition. *L'Heure Bleue* a donc été découpée en trois parties, rendant compte des caractéristiques décrites.

*L'Heure Bleue* est une tentative de transmettre de façon immersive l'expérience du lever du jour dans une forêt. L'espace sonore étant un espace en trois dimensions, la question des procédés à mettre en œuvre pour recréer cet enveloppement se pose. Ainsi, nous nous sommes demandé par quels procédés se créait l'immersion et comment l'appliquer à une pièce sonore dite « immersive », puis de la pertinence d'un système multicanal pour permettre l'immersion. Nous avons vu que l'immersion se créait à partir de trois stratégies : l'enveloppement, la saturation et l'implication. L'enveloppement est un débordement de l'espace visuel et sonore du public. Cela permet de placer le spectateur et la spectatrice au cœur de l'action. La saturation est le détachement du monde réel en saturant de nouvelles informations sensorielles. Il s'agit de donner des informations provenant de toutes les directions, et d'en donner beaucoup, comme on pourrait le ressentir dans la réalité, pour détacher le spectateur et la spectatrice du monde réel. Enfin, l'implication est un engagement avec le monde virtuel, le monde du récit, par le moyen d'interfaces d'interaction réelles ou non. Dans le domaine sonore, ces trois stratégies sont applicables plus ou moins simplement. Concernant l'enveloppement, il est réalisable assez facilement avec l'utilisation de technologies surround et 3D ou par le son binaural. La saturation est tout aussi applicable, par la

composition de la scène sonore et/ou en couplant le dispositif sonore à des dispositifs sollicitant d'autres sens tels que la vue. Les dispositifs multicanaux participent à ce procédé en permettant de donner des informations provenant de plusieurs directions à la fois. Enfin, l'implication peut se faire grâce à des capteurs qui permettent de générer des événements sonores suivant les actions du public par exemple. On peut même réaliser cet effet plus simplement encore en laissant un espace de déambulation à l'auditeur et à l'auditrice. Dans le cadre de *L'Heure Bleue*, l'intérêt d'un dispositif sonore multicanal est grand, car il permet de donner des informations provenant de toutes les directions. Un système reposant sur la diffusion à travers des enceintes permet de s'approprier l'espace en tournant la tête par exemple. Afin de focaliser les auditeurs et les auditrices sur l'écoute de la pièce, il est possible et plutôt simple de les priver de la vue en plongeant la pièce dans le noir.

Il existe beaucoup de technologies permettant la reproduction du champ sonore en trois dimensions et il convenait donc d'en choisir une pour la diffusion de *L'Heure Bleue*. Une première distinction peut être faite entre les dispositifs se basant sur la diffusion par haut-parleurs et les dispositifs diffusant par un casque. Pour ce qui est de la diffusion au casque, il existe la technologie binaurale, qui a l'avantage de rendre des pièces immersives accessibles au grand public. Mais l'écoute binaurale peut également constituer un inconvénient majeur car elle ne permet pas une écoute collective dans un contexte d'installation sonore. L'écoute de *L'Heure Bleue* se voulant collective, il a fallu avoir recours à une diffusion à partir d'enceintes. Il existe beaucoup de systèmes multicanaux et nous en avons détaillé quelques-uns, en commençant par la technologie surround 5.1 et 7.1 utilisée au cinéma. Ce sont des techniques adaptées pour le cinéma, et ►

la sensation d'enveloppement n'est pas vraiment atteignable sans un support visuel. De plus, elles ne reproduisent pas les sons sur le plan de l'élévation. Pour le cinéma toujours, Dolby a développé le système Dolby Atmos, qui offre deux canaux en hauteur. Cependant, c'est un système pensé pour le cinéma et pour être diffusé dans des salles de cinéma, avec une forte prévalence pour la partie frontale. Il se prête difficilement aux installations immersives. En matière de technologies audio 3D, non orientées spécifiquement vers le cinéma, nous pouvons citer le VBAP, la WFS ou l'ambisonie. Le VBAP ou *Vector Based Amplitude Panning* est une généralisation des panoramiques stéréos sur une, puis deux dimensions supplémentaires. Le VBAP offre la possibilité de travailler en élévation. La WFS ou *Wave Field Synthesis* est un modèle de synthèse de front d'onde. Elle se base notamment sur des calculs de délais visant à reproduire la forme d'un front d'onde. Le principal inconvénient de cette technologie est qu'elle ne prend pas en compte l'élévation. Cependant, elle permet une bonne reproduction de la profondeur du champ sonore. Enfin, l'ambisonie est un système qui repose sur la décomposition de l'espace en harmoniques sphériques (en 3D) ou circulaires (en 2D). Chaque source est donc une somme précise de ces harmoniques. C'est un format qui permet de travailler sur toute une sphère, en azimut et en élévation, sans introduire de distorsion de l'espace ou de timbre lorsqu'il est convenablement décodé. C'est le système qui a été choisi pour la diffusion de *L'Heure Bleue*.

Enfin, après avoir défini tous les outils qui allaient être utilisés pour la réalisation de *L'Heure Bleue*, nous avons détaillé toutes les étapes de la composition de la pièce. Cela passe d'abord par l'enregistrement. Les sessions d'enregistrements ont eu lieu les 4, 5, 6 et 7 mars 2021, de 5h à 8h30 environ. L'aube étant un processus assez long au niveau sonore, l'enregistrement

était lancé environ une heure avant le début de l'aube et tournait pendant environ trois heures trente. Les enregistrements ont eu lieu aux alentours du village de Treffort dans le département de l'Ain, dans des vallées qui sont peu peuplées et à l'abri de beaucoup d'activités humaines, surtout le matin. Le micro choisi était le Sennheiser Ambeo VR Mic, car il est simple d'utilisation et une bonnette anti-vent était fournie par l'entreprise de location DC Audiovisuel. Ce microphone permet d'enregistrer en ambisonie de premier ordre au format A (qui correspond aux signaux enregistrés par une matrice de quatre microphones cardioïdes en forme de tétraèdre). Ces quatre jours d'enregistrement ont produit 17h30 de rushes, qui ont été dérushés pendant le mois de mars 2021. Les enregistrements en ambisonie, écoutés d'abord au casque via un rendu binaural, donnent une grande sensation de profondeur. Beaucoup d'événements se sont produits pendant les enregistrements, ce qui a permis d'avoir de nombreuses options pour le montage. Seuls les éléments enregistrés en mars 2021 ont été utilisés dans la pièce, afin de se concentrer sur la même zone géographique et de rendre compte spécifiquement de cet environnement. L'objectif de cette création étant de retranscrire l'expérience de la forêt, le montage des événements se devait d'appuyer le propos. Enfin, le mixage a été fait dans la salle multicanale de l'ENS Louis-Lumière. Les enjeux de cette étape sont multiples. Il s'agit d'affiner la localisation des sources, ainsi que leurs niveaux et leurs timbres, de bien répartir l'énergie acoustique dans l'espace, de s'assurer de l'attention des auditeurs et des auditrices et d'améliorer les transitions.

La création de la pièce *L'Heure Bleue* nous a permis de confronter nos recherches à une démarche artistique présentée à un public. Nous avons vu que, par une composition précise des différents moments clés de l'aube, il était possible de faire comprendre aux

spectateurs et aux spectatrices qu'une transition avait eu lieu. De même, par l'utilisation d'un système ambisonique et la privation de la vue, nous avons observé que dans la plupart des cas une immersion était provoquée, ou en tout cas que les auditeurs et les auditrices se concentraient davantage sur le son et l'espace sonore qui leur était présenté. L'objectif de *L'Heure Bleue* était d'amener l'expérience de l'aube à un public en mettant en lumière les mécanismes qui sont en jeu à cette période de la journée, en gardant l'aspect enveloppant et « naturel » d'un tel environnement.

*L'Heure Bleue* a été composée pour un système ambisonique. De ce fait, son écoute en dehors de tels systèmes est très compliquée. Un rendu binaural a cependant été effectué et peut être écouté via le lien suivant : <https://soundcloud.com/user-204283851/sets/lheure-bleue>

## Filmer la brume, le brouillard et la fumée : enjeux artistiques et techniques

Camille AUBRIOT,  
Master Cinéma 2020, ENS Louis-Lumière

La brume et le brouillard sont des phénomènes météorologiques naturels hautement cinématographiques : ils cachent et révèlent tour à tour paysages et personnages dans une symphonie poétique du mouvement. Tous deux porteurs de sens, autant visuellement que scénaristiquement, ils modifient les perceptions du spectateur et ses repères spatio-temporels. Nombre d'apparitions et de disparitions se font ainsi dans le brouillard... Ces deux phénomènes naturels sont très délicats à filmer du fait de leur brièveté, car il n'y a aucune assurance de continuité. Des alternatives ont rapidement été trouvées pour les reproduire artificiellement et les rendre plus maîtrisables, notamment avec des « machines à fumée ». Au-delà de recréer des brouillards artificiels en extérieur, la fumée permet de réaliser de nombreux effets visuels particulièrement appréciés des directeurs de la photographie, tels que matérialiser des rayons lumineux, créer une diffusion ou de la profondeur dans l'image, texturer un décor, donner une ambiance particulière (surréaliste, inquiétante...), marquer un temps au passé, etc. La fumée permet aussi de marquer visiblement l'effet d'un feu de bois, de renforcer la sensation de froid dans une séquence de neige, ou encore de faire fumer un café, un barbecue ou un moteur de voiture durablement.

Si la fumée n'est pas toujours visible en tant que telle, notamment dans le cas de diffusions légères, elle est employée dans de très nombreux films tout au long

de l'histoire du cinéma. Elle se retrouve souvent dans les films de science-fiction, de *fantasy* ou fantastiques, d'horreur, de films d'époque, mais pas seulement. La fumée est notamment beaucoup utilisée dans la cinématographie anglo-saxonne pour son esthétique visuelle, très ancrée dans leur culture.

Notre recherche, sous la direction de Sylvie Carcedo et Jean-Marie Dreujou, nous a permis de détailler principalement l'approche des directeurs de la photographie concernant ces effets de brume, de brouillard et de fumée, mais également leur mise en œuvre globale sur un tournage. Comment les filmer, que ces phénomènes soient naturels ou reproduits artificiellement ? Quelles sont les contraintes de tournage inhérentes à ces effets, produits généralement par les accessoiristes ou les superviseurs SFX (effets spéciaux physiques, réalisés directement sur le plateau), voire par les superviseurs VFX (effets spéciaux numériques) ? Afin de répondre à ces questionnements, le travail de recherche s'est basé sur une documentation variée, sur une pratique personnelle grâce à la Partie Pratique de Mémoire, ainsi que sur plusieurs rencontres avec des professionnels, présentées sous la forme d'entretiens en annexes. Des directeurs de la photographie ont partagé leurs expériences avec la fumée : Nicolas Bolduc (CSC), Benoît Chamillard (AFC), Rémy Chevrin (AFC), Jean-Marie Dreujou (AFC), Éric Guichard (AFC) et Philip Lozano (AFC). Il y a également parmi ces entretiens un accessoiriste, Benoît Herlin ; deux superviseurs d'effets spéciaux physiques SFX, Philippe Hubin et Jean-Christophe Magnaud (BigBang SFX) ; et un superviseur d'effets spéciaux numériques VFX, Bruno Maillard. De plus, les rencontres avec les loueurs d'effets spéciaux, Flam and Co et Pyrofolies, ont été très enrichissantes et déterminantes dans la réalisation de la Partie Pratique.

Ces différentes discussions étaient capitales pour

approfondir le sujet, car il existe peu de ressources concernant l'aspect technique de l'emploi de la fumée au cinéma. Le sujet semble être inédit dans le domaine de la recherche – au-delà d'analyses esthétiques –, du moins en France. Cela est étonnant, tant la fumée est largement employée, que ce soit dans des films récents comme *Blade Runner 2049* (Denis Villeneuve, 2017), *Dans la brume* (Daniel Roby, 2018), *La Belle époque* (Nicolas Bedos, 2019), ou dans des films plus anciens.

Nous avons analysé en premier lieu les phénomènes de la brume, du brouillard et de la fumée d'un point de vue scientifique, afin de mieux les comprendre. Les brumes sont moins denses que les brouillards, mais tous deux sont constitués de microgouttelettes d'eau en suspension dans l'air, tandis que la fumée est un mélange de particules solides et de gaz. Brumes et brouillards résultent de la condensation de la vapeur d'eau contenue dans l'air, suite à différents processus liés à des changements de température, de pression ou d'humidité. La fumée résulte généralement, quant à elle, d'une combustion. Leurs principes de formation respectifs sont détaillés, ainsi que les différents types de brouillard existants. Le *smog* (mélange de fumée et de brouillard) et les brumes sèches de type tempêtes de sable sont également cités, de même que des anecdotes de tournages en brouillard naturel.

Afin de pallier les problèmes rencontrés pour filmer ces phénomènes naturels, une multitude de « machines à fumée » sont utilisées aussi bien en extérieur qu'en intérieur, souvent couplées avec des ventilateurs. Fonctionnant à l'électricité, au gaz ou à l'essence, les modèles varient en taille et en puissance, des toutes petites Tiny Pocket ou Tiny S sur batterie (de marque LOOK) aux énormes machines permettant de remplir une vallée entière. Ces « machines à fumée » produisent des microgouttelettes en suspension dans

l'air, donc des brouillards, plutôt qu'une véritable fumée. Cependant, d'autres techniques produisent de vraies fumées (au sens propre du terme), telles que la combustion d'encens, de papier d'Arménie, de tablettes à fumée, ou l'emploi de *beesmokers*, ces appareils servant à enfumer les ruches. Des techniques plus archaïques étaient auparavant employées, comme le fait de brûler des pneus du temps de *Quai des brumes* (Marcel Carné, 1938), ce qui produisait une fumée toxique interdite de nos jours. Il existe également une sorte de fumée très particulière appelée fumée lourde, qui reste au sol de manière presque magique, produite à base de carboglace ou d'un liquide spécial. Le terme de « fumée » est ainsi employé de manière globale, bien que souvent de manière inappropriée scientifiquement parlant.

Les interactions de la fumée, brume ou brouillard avec la lumière sont ensuite expliquées, afin de mieux les maîtriser. La nature de la lumière (une onde électromagnétique), ainsi que les processus de diffusions principalement à l'œuvre sont ainsi détaillés, de même que certaines curiosités optiques liés au brouillard, tels que l'arc-en-ciel blanc.

Le travail des directeurs de la photographie avec la fumée constitue le point central du mémoire, que ce soit du point de vue de la lumière ou de la caméra. L'éclairage représente une grande part du travail, et certaines notions sont capitales à connaître en fonction de l'effet souhaité. Par exemple, pour matérialiser des rayons lumineux, il convient d'employer une source de lumière très directionnelle et puissante à contre-jour (ou *a minima* en latéral). La fumée nécessite également d'être vigilant concernant l'exposition, que ce soit en pellicule ou en numérique. Ces questions sont abordées, ainsi que la coloration de la fumée et l'apport possible de l'étalonnage et des effets spéciaux numériques (VFX). ►



Figures 1 à 4 : vidéostills du film *Into the fog* © Camille Aubriot, 2020

La fumée influence également le cadrage, car elle adoucit et décontraste l'image, tout en ajoutant un élément visuel qui, parfois, vient même obstruer le sujet filmé. La composition de l'image doit ainsi tenir compte de la fumée, qui, souvent, permet de créer une certaine profondeur. L'usage de filtres de type *Fog* ou *Low Fog* peut s'avérer utile en complément de la fumée, notamment dans le cas d'un tournage à plusieurs caméras ; car la densité de fumée ressentie à la caméra ne sera pas la même en fonction de la focale employée ou de l'axe caméra par rapport à la lumière. Cela nécessite donc d'adapter constamment le niveau de fumée sur le tournage, ce qui peut être vite délicat. De plus, la mise en œuvre de tels effets est souvent synonyme de gros moyens financiers et logistiques, car cela nécessite généralement du matériel imposant, des techniciens dédiés et un certain temps de mise en place ; sans oublier certaines contraintes liées à la fumée (désactiver les alarmes incendie, le vent en extérieur qui peut chasser toute la fumée, le bruit des machines pour les prises de son, etc.). La fumée n'est donc pas si simple que cela à employer.

Les données théoriques ont été mises à l'épreuve d'expérimentations personnelles dans le double objectif d'effectuer des essais comparatifs filmés, afin de mieux appréhender la gestion et les rendus visuels de la fumée, et d'appliquer les effets de fumée spécifiques, tels que de la fumée lourde et des cocktails bouillonnants avec de la carboglace grâce à la réalisation d'un court métrage expérimental.

## Reconstitution du procédé de photographie en relief de Louis Lumière : « la Photo-stéréo-synthèse »

| Clémence Lavigne,  
Master Photographie 2021, ENS Louis-Lumière

### Comment retrouver un dispositif disparu ?

La représentation du relief est une problématique très ancienne qui ne cesse de se renouveler. Dès ses premières représentations figuratives datant du paléolithique, l'être humain se questionne et expérimente la profondeur. En étudiant l'état des surfaces rocheuses, il joue avec leurs diverses aspérités pour texturer son dessin, mais aussi suggérer la notion d'espace. Premiers bas-reliefs de l'Histoire, les fresques préhistoriques ne sont que les prémices d'une succession d'études et d'expérimentations, tant artistiques que scientifiques. Les progrès et les évolutions théoriques dans le domaine de l'optique font naître de nouvelles formes de représentations qui visent à reproduire la vision de la profondeur. Des fresques romaines à la perspective italienne, les techniques picturales d'illusion de la troisième dimension évoluent. Dès la révolution industrielle du XIXe siècle marquée par l'invention de la photographie, la course à l'innovation s'accélère. La conception et la production de dispositifs photographiques cherchant à restituer une sensation de relief ne cessent d'accroître. Du stéréoscope à miroirs jusqu'aux installations immersives plus contemporaines, la photographie entretient l'espoir d'être, un beau jour, capable d'observer une image en profondeur.

En 1920, Louis Lumière s'inscrit dans cette quête du relief et dépose un nouveau brevet d'invention<sup>1</sup>. Il y décrit un procédé photographique expérimental atypique : la photo-stéréo-synthèse. Cette technique repose sur la réalisation d'une succession d'images négatives de ce solide, à échelle et point de vue fixes, en décalant la mise au point sur différents plans parallèles, équidistants ou non. Si nous prenons l'exemple d'un portrait, le bout du nez serait net sur la première image, les lèvres sur la seconde, les yeux sur la troisième et ainsi de suite. Les images négatives photographiées sont tirées en positif sur plaque de verre, pour ensuite être superposées les unes derrière les autres. Leurs espacements sont proportionnels aux distances séparant les plans de mise au point et au rapport de grandissement. Les plaques sont montées dans une boîte en bois fermée, constituée d'une plaque de protection en verre à l'avant et d'un verre dépoli à l'arrière qui permet d'observer la superposition par transparence avec un système de rétroéclairage. Si l'observateur se place dans l'axe, à la distance orthoscopique du montage, les photographies s'alignent et il peut admirer le portrait en relief.

Afin que cette sensation soit véritable, les images doivent cependant présenter une profondeur de champ extrêmement réduite. En 1920, les objectifs disponibles ne permettent pas d'obtenir la quantité de flou nécessaire au procédé de photo-stéréo-synthèse. Ils sont sujets aux aberrations optiques qui impactent considérablement la qualité du flou et leur diamètre d'ouverture reste trop faible. Louis Lumière imagine alors un dispositif mécanique qui réduit artificiellement la zone de netteté. En d'autres termes, il agrandit fictivement l'ouverture de l'objectif grâce à un dispositif conçu sur le principe des transformations homographiques des plans de front. Afin d'expliquer et mieux comprendre ce concept optique, nous nous ►

1

Lumière, L. (21 janvier 1920), Procédé de stéréo-synthèse-photographique par stratification, brevet d'invention n°523 962.

appuyons sur le schéma suivant, proposé dans le brevet d'invention.

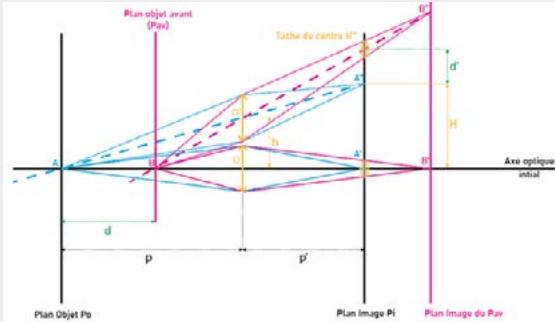


Figure 1: Reproduction de la figure originale dessinée par Louis Lumière pour la démonstration de son système optique.

Nous prenons un objectif classique, ici représenté par une lentille convergente de centre O. La mise au point est ajustée sur notre sujet, le point A. Son image A' apparaît nette sur notre surface sensible P<sub>i</sub>. En déplaçant la lentille de centre O d'une distance h en O', nous obtenons le point image A'', conjugué du point objet A par O', toujours sur le plan image P<sub>i</sub>. Mais si en parallèle, le plan image P<sub>i</sub> effectue lui aussi un déplacement proportionnel H, dans son plan et sans rotation, les points-images A' et A'' se confondent. Le mouvement de l'objectif est annulé par le mouvement corrélé de la plaque photographique dont l'amplitude est déterminée par le théorème de Thalès.

À présent, considérons un point objet B, situé sur l'axe optique à une distance d du plan de mise au point P<sub>O</sub> et de A. Son image B' se forme derrière le plan. En gardant l'objectif fixe, nous voyons que les points A, B, O, A' et B' restent alignés sur l'axe. Mais si la lentille translate vers sa seconde position, les points A, B, O', A'' et B'' ne

sont plus du tout alignés. Deux axes bien distincts se dessinent (en bleu et rose sur le schéma). Sur le plan P<sub>i</sub>, l'axe (BB'') s'éloigne de l'axe (AA'') d'une distance d'. Cette dernière représente l'augmentation artificielle de la tache de flou. La tache formée par le point B'' décrit une couronne autour du point A', une couronne d'un diamètre largement supérieur à la tache de flou initiale.

Grâce à ce principe optique, un objectif ouvert à 4.5 permet d'obtenir une quantité de flou équivalente à celle d'un objectif fixe ouvert à 1.6.

Malgré l'élégance du concept, très peu d'images réalisées par photo-stéréo-synthèse ont été conservées. Pour des raisons techniques comme économiques et commerciales, le dispositif de prise de vues est resté au stade de prototype et demeure aujourd'hui disparu.

Afin de mieux le comprendre et tenter de retrouver une esthétique éteinte, nous avons entrepris sa reconstruction. À notre connaissance, il n'y a jamais eu de tentative de ce type depuis 1920. Ce travail a donc nécessité une étude approfondie du principe optique et mécanique de la chambre à partir du brevet d'invention,

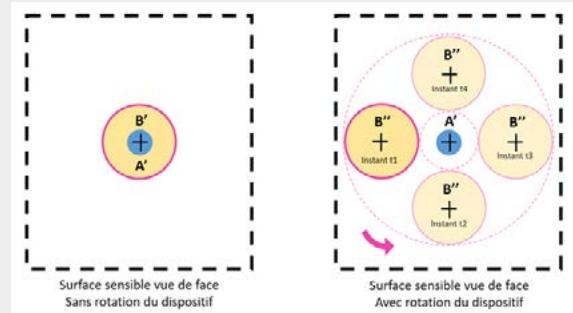
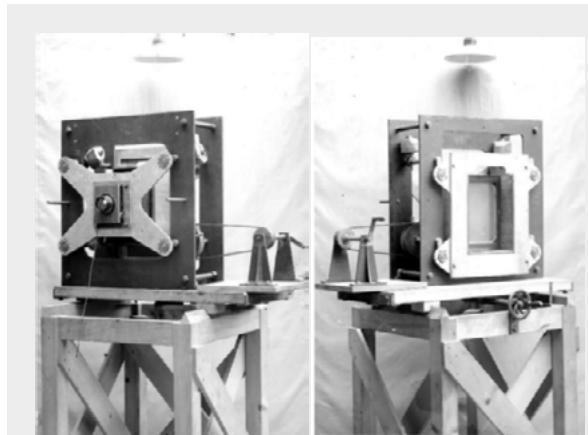


Figure 2 : Surface sensible vue de face avec ou sans rotation du dispositif.

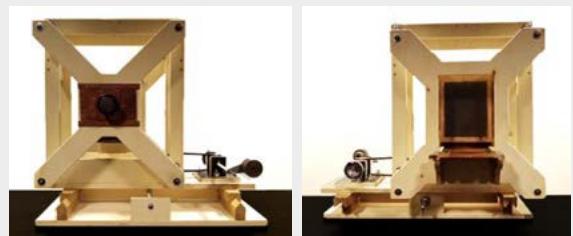
pour ensuite modéliser et concevoir notre prototype. Avec l'aide de mon premier frère Benoit, ébéniste, nous avons ensuite usiné et monté toutes les pièces en bois. Avec mon second frère Bixente, mécanicien, nous avons conçu et assemblé tout le système mécanique de mise en rotation. Des tests de fonctionnement ont suivi le montage final et des améliorations techniques ont été apportées pour compenser certaines erreurs. Nous obtenons le prototype suivant :

Malheureusement, ce dispositif présente des défauts et des décalages assez importants inhérents à la conception du système de rotation. Les images capturées avec ce dispositif sortent en fin de compte entièrement floues et nous n'avons pas pu atteindre l'étape de la superposition finale des images. Mais ce projet de recherche me tient trop à cœur pour se terminer ainsi. Nous venons d'entreprendre le démontage et la reconception de notre prototype. Nous pensons revoir le système de rotation, plus particulièrement la conception des bielles qui n'était pas adaptée au dispositif. Nous envisageons même d'incruster un système électronique afin d'automatiser la rotation pour une meilleure fluidité de mouvement et une vitesse constante.

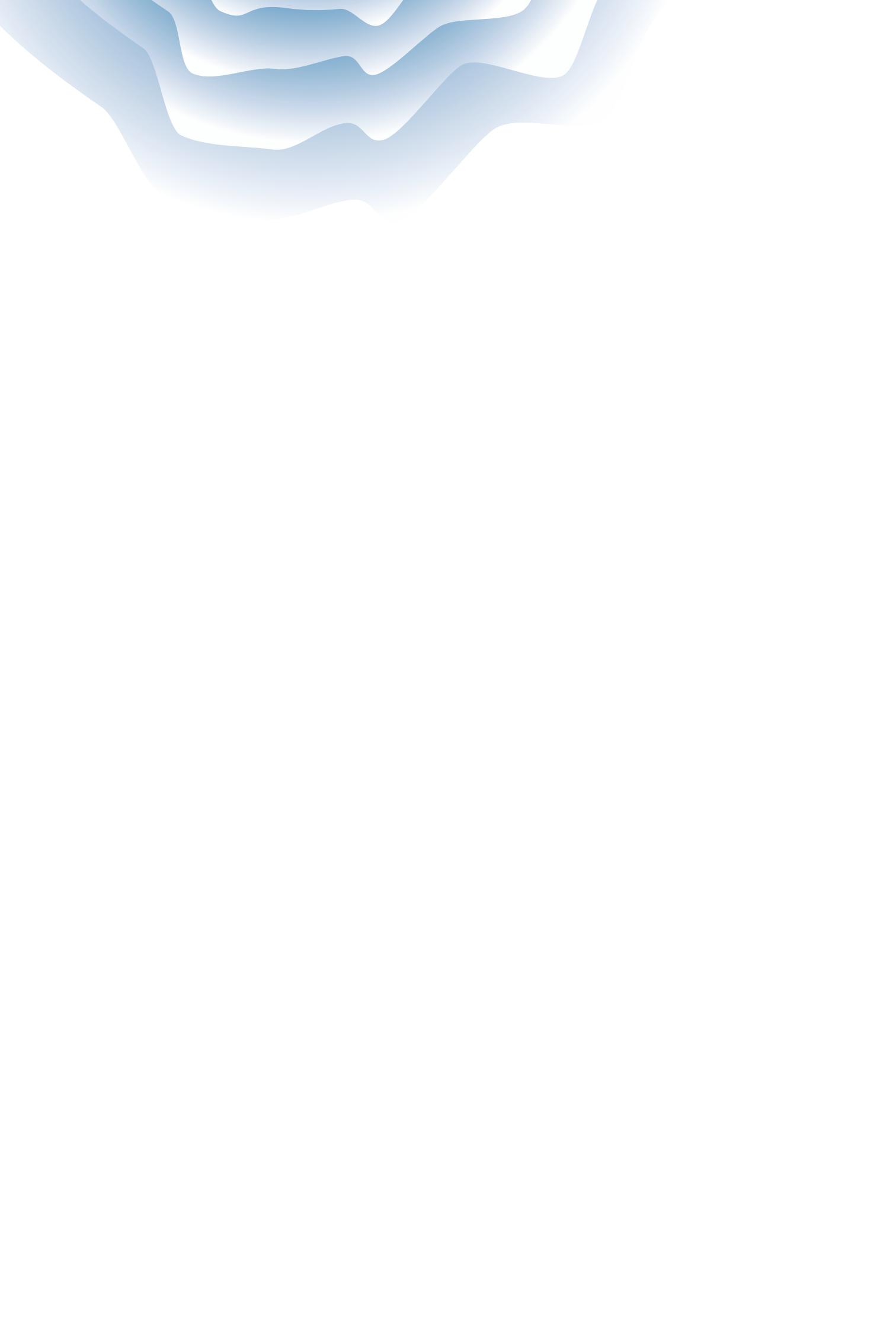
Cette reconstitution représente un défi ambitieux, mais ravive l'espoir d'une photographie dite intégrale, une photographie présentant au spectateur un espace en trois dimensions sans nécessiter un dispositif de visualisation. Et force est de constater combien ce procédé était particulièrement novateur, car est-il besoin de rappeler que les scanners utilisés aujourd'hui dans le domaine médical fonctionnent sur un principe similaire de stratification des volumes.



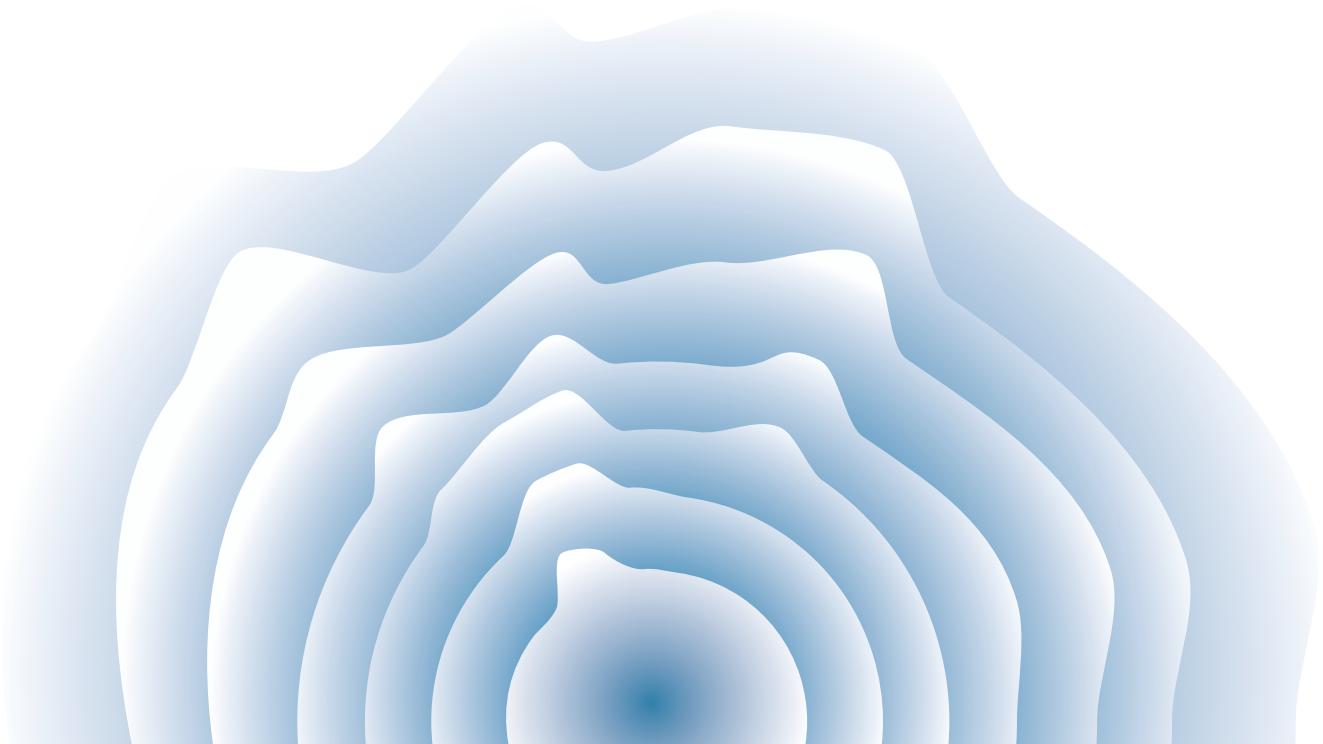
**Figure 3 : Photographies du dispositif historique tirées du brevet d'invention de 1920.**



**Figure 4 : Photographie de notre prototype du dispositif de photo-stéréo-synthèse.**



# Notes de lectures



## ***Sensory Experiences: Exploring meaning and the senses,***

**Danièle Dubois, Caroline Cance, Matt Coler, Arthur Paté, Catherine Guastavino, John Benjamins Publishing, 2021**

| Corsin Vogel

Comment étudier la connaissance du monde que nous construisons à travers nos sens ? Comment mettre en œuvre des procédures expérimentales rigoureuses qui respectent nos pratiques du quotidien et le sens que nous attribuons aux objets de notre environnement ? Répondre à ces questions, tel est l'ambitieux objectif que se sont fixés les auteurs de *Sensory Experiences: Exploring meaning and the senses* (que l'on traduira par : « Expériences sensorielles : explorer le sens et les sens »), paru en 2021.

La démarche de Danièle Dubois et ses collègues se situe au sein des sciences cognitives en intégrant, dans une approche sémantique et psycholinguistique, les aspects culturels et sociaux bien souvent absents de l'approche cognitiviste classique qui fonde les processus psychologiques sur le modèle du traitement de l'information. C'est à partir de l'analyse de descriptions verbales produites par les participants lors d'études perceptives que les auteurs développent à la fois une théorie sémantique du sens accordé à l'expérience sensible et des méthodes empiriques pour rendre objective cette expérience subjective. Les auteurs proposent ainsi un glissement théorique et méthodologique vers une approche située de la cognition, qui permet aujourd'hui de tenir compte de la diversité à la fois culturelle et historique des expériences humaines.

Le livre se compose de deux grandes parties. La première se présente comme une réflexion menée à l'aune de nombreuses années d'expérimentations dans les différentes modalités sensorielles. Après un premier chapitre retraçant les approches psychophysiques, cognitivistes et comportementalistes, ancrées dans la description du monde physique, et les apports des sciences de l'information, les auteurs proposent une autre perspective cognitive dans laquelle les différentes modalités perceptives sont *situées*. En effet, en sollicitant les différents sens, les recherches empiriques menées pendant plus de trois décennies ont conduit Danièle Dubois et ses collègues à identifier ce qui *fait sens* dans la production et l'usage des objets du quotidien. Les chapitres 2 à 8 décrivent ainsi un ensemble de pratiques et usages dans quatre modalités perceptives, en commençant par l'expérience visuelle et la perception des couleurs. Puis, la modalité auditive est traitée dans trois chapitres, à travers les recherches sur les *soundscape*s, le traitement de la parole et celui de la musique. Nous y retrouvons notamment la référence au concept de milieu sonore, ou plus largement d'*Umwelt*, autre formulation de la notion d'environnements immersifs proposée par Uexküll, avec lesquels l'humain, et plus largement l'animal, interagissent. Suivent enfin les recherches sur l'olfaction et le goût (le sens haptique du toucher ne sera pas abordé dans cet ouvrage). Le chapitre 9 conclut la première partie de l'ouvrage en théorisant les résultats issus de ces études empiriques et en mettant en exergue leurs conséquences sur la façon de questionner les expériences sensorielles et d'élaborer des protocoles expérimentaux. Une phrase résume le positionnement des chercheurs : « Le monde entre les sujets est caractérisé par des pratiques situées faisant sens et basées sur une construction collective d'un savoir partagé ou d'une mémoire expérientielle<sup>1</sup> ». Ce faisant, les questionnements ancrés en psychologie et en linguistique bénéficieront des sciences sociales

1

Notre traduction de : « *the world between subjects is characterized by sense-making practices in situ based on the co-construction of common shared knowledge or experiential memory* », p. 354

telles que l'histoire, l'anthropologie, mais également l'histoire des arts et des technologies à travers les productions artistiques et la *culture matérielle*.

La seconde partie de l'ouvrage (chapitres 10 à 15) présente des conséquences méthodologiques globales pour mener à bien l'exploration du sensible, en dépassant les oppositions entre évaluations subjectives et objectives. Ainsi, il s'agit de trouver comment questionner les individus et la manière dont il faut les considérer : sont-ils des *sujets*, génériques et indifférenciés, qui subissent les expériences, ou plutôt des *participants*, qui, en fonction de leur expérience, leurs savoirs et savoir-faire, sont actifs dans leurs rapports aux objets perçus et questionnent eux-mêmes les *a priori* de l'expérimentateur ? En plaçant les acteurs humains, tels les musiciens, ingénieurs du son, designers, parfumeurs, œnologues ou « simples » usagers, dans leurs pratiques ordinaires ou expertes, au centre des questionnements, les auteurs développent une démarche expérimentale qui s'appuie sur les concepts de validité écologique des recherches en laboratoire, et portent ainsi une attention particulière sur le choix des stimuli, des échantillons à tester, à la formulation des questions dans les questionnaires, aux consignes dans les tâches de catégorisation libre, avant de développer des analyses sémantiques précises des verbalisations produites par les participants.

De par la richesse des sujets traités et par la productivité des concepts et des méthodes introduits pour mener à bien une étude sur la perception, ce livre constitue un ouvrage de référence pour toute personne souhaitant faire une recherche rigoureuse, et permet d'éviter les nombreux biais et *a priori* qui apparaissent inévitablement lors de toute exploration systématique sur la perception et le sens. Outre son intérêt dans le domaine applicatif, l'ouvrage contribue

au développement de cadres théoriques renouvelés en psychologie et linguistique cognitives. Un ouvrage riche, très bien structuré et pédagogique, précieux guide méthodologique et indispensable outil pour les chercheurs et les étudiants à partir du niveau de Master, mais aussi pour les professionnels qui contribuent aux recherches appliquées.

## ***Puissances de la parole. À l'écoute des films,***

**Mathias Lavin, Éditions Mimésis, 2021**

| Corsin Vogel

L'enjeu de l'ouvrage de Mathias Lavin sur la parole au cinéma est ambitieux : trouver un espace pour penser la question de la parole qui n'ait pas à choisir entre les dimensions sonores et visuelles du cinéma, un espace qui permette de tenir « le refus d'un partage entre images et sons ». Posant d'emblée que sa réflexion peut trouver un point d'inspiration dans un corpus d'écrits qui, depuis les années 1980, traitent du son au cinéma et, partant du constat que les études cinématographiques restent encore en grande partie marquées par un primat de l'image sur le sonore, il précise cependant que son objet se situe ailleurs que dans la lignée des *sound studies*, tout en reconnaissant leur pouvoir de stimulation. Ainsi, de manière complémentaire aux approches historiques et technologiques (même s'il les croise et s'en nourrit), il propose une démarche analytique passionnante d'œuvres filmiques singulières, mises en avant tout au long de l'ouvrage, pour revendiquer, selon la belle expression choisie en sous-titre, une « écoute des films », une écoute qui « cependant ne rende pas aveugle à leur matière imageante ».

L'ouvrage, organisé en chapitres thématiques, dessine un territoire, celui de la parole, qui n'est pas celui de la voix – bien que leurs deux espaces se recouvrent et fassent preuve de porosité – mais plutôt celui, politique, du « droit à la parole », qui fonde notre humanité. L'auteur appuie d'emblée la dimension politique fondamentale de son propos : « L'homme est un animal politique et parlant, là où l'animal (l'appartenance à

l'animalité) ne détient qu'une capacité de phonation, de cri, éventuellement de communication, mais pas de parole en tant que telle permettant de discriminer entre le juste et l'injuste, entre le vrai et le faux, et de trancher ainsi les débats sur les raisons et les valeurs. » La parole est envisagée en termes de « puissances », c'est-à-dire en termes d'effets, d'éléments qui font sens, et qui sont fondamentalement dynamiques : l'auteur ne cherche pas à théoriser et organiser une typologie exhaustive de catégories de la parole, mais plutôt à observer et analyser, à travers cinq grands chapitres thématiques, des régimes divers et mouvants de ces « puissances de la parole ».

Le premier chapitre, consacré à la surdité, met l'accent sur la nécessité de penser la parole en rapport avec la notion d'écoute, en insistant sur la différence entre voix et parole, mais également en dépassant le strict champ du sonore. Mathias Lavin engage une réflexion sur un point limite de la parole et de l'écoute, qui rend sensible à la fois « le rôle structurant du silence » et la place essentielle du geste dans toute situation de parole. La notion de « parole gestuelle », qui évoque en premier lieu la langue des signes, lui permet de penser ce point limite, à partir de séquences de films longuement et clairement analysées. Considérant la surdité comme « lieu de convergence possible entre geste et parole », il tente alors de répondre à la question de la figuration de la parole au cinéma : « comment peut-on montrer [...] l'impossibilité d'entendre ou de parler ? ».

Le chapitre suivant, consacré à la parole au cinéma muet, s'étend à toute figuration gestuelle de la parole et même, potentiellement, à tout l'espace filmique, l'auteur montrant bien comment l'ensemble des éléments visuels, plastiques, rythmiques des films convoqués concourt à rendre sensible le fonctionnement, les effets de la parole. C'est là une puissance de contamination, de

« dissémination », que l'auteur résume sous le terme de « phonogénie » - emprunté à Epstein et calqué sur celui de photogénie - et qui est le point de départ du troisième chapitre, consacré à la question de l'outil, du média qui diffuse la parole. Cette phonogénie, serait une sorte de « supplément d'âme », dans le domaine sonore, qui permettrait d'approcher tout le non-dit, tout ce qui, dans la parole, semble échapper à celle-ci, mais se diffuse ailleurs, dans les corps, les espaces, les motifs plastiques des films. Le format radiophonique devient dans ce troisième chapitre opérant pour penser les effets de dissémination globale de la parole, leurs puissances (y compris celles de la voix), et revenir, ce faisant, à des questionnements sur le rôle politique de la parole.

Le quatrième chapitre, qui se penche sur la parole littéraire, semble faire un pas de côté dans la trame du livre, mais revient en fait, fondamentalement à cette dimension politique nodale, en analysant la capacité de trouble que peut produire, induire une parole littéraire, que l'auteur appelle aussi « parole d'écriture » (empruntant à Blanchot). Cette parole d'écriture, qui suppose qu'on énonce oralement comme on lirait, comme on écrirait, qui fait entrer la parole en littérature, disposerait d'une capacité de déplacement qui est transgressive : elle donne à voir/entendre une polyphonie, une profusion ou une suspension de la signification ordinaire de la parole. Cette transgression s'apparente à une possibilité de faire entendre une voix/une parole « autre », jamais entendue comme telle, une parole politique que poursuit l'auteur et qui viendra clore le livre, en son dernier chapitre.

Avant cela, le cinquième chapitre fait un retour sur le frottement entre voix et parole, sur le potentiel érotique de la voix. L'analyse de quelques séquences de films permet à Mathias Lavin d'en souligner plutôt

l'ambivalence, montrant que, si la voix peut érotiser et déborder l'image, elle peut aussi contrarier ce mouvement, contredisant ou déplaçant ce que le visible figure d'une manière toute singulière. Là aussi, c'est la capacité de résistance à un sens univoque de la voix (au cinéma) qui intéresse l'auteur, et qui l'amène en toute logique à ce chapitre conclusif, consacré, lui, à un seul film : *La Blessure* de Nicolas Klotz. À partir de ce film, il revient à son interrogation première, celle des paroles politiques au cinéma, cherchant à circonscrire la moins évidente, celle qu'il appelle la parole « autre » ou inouïe, et qu'il touche ici dans une de ses versions les plus violentes : celle de la parole entravée, empêchée, violentée de migrants africains sur le territoire français.

Comment donner à entendre au cinéma cette parole « autre », comment l'écouter, et comment répondre à cette écoute ? C'est cet enjeu éminemment politique, appuyé par un questionnement inédit, que l'auteur mène à travers un tissage précis et d'une grande érudition d'exemples filmiques et de références littéraires et théoriques. Ce faisant c'est aussi une petite histoire du cinéma, en partie subjective et évidemment lacunaire, que dessine l'ouvrage. On ressort de cette lecture avec l'envie de voir et revoir tous les films analysés, prêt à être pleinement à leur écoute.

# Cahier Louis-Lumière

Revue annuelle

*Corsin Vogel et Sylvain Lambinet tiennent à remercier très chaleureusement :*

- l'ensemble des autrices et auteurs qui ont contribué à l'élaboration de ce numéro,
- Giusy Pisano, pour ses propositions de notes de lectures et ses relectures attentives,
- Danièle Dubois, pour sa disponibilité, ses nombreuses relectures et ses remarques constructives,
- Barnchade Pourvali, pour ses relectures consciencieuses et le suivi iconographique,
- Jérémy Piot, pour ses nombreuses maquettes préliminaires,
- Méhdi Aït-Kacimi, pour avoir mis ce beau projet sur les rails,
- Vincent Lowy, pour l'invitation à co-diriger ce numéro et la confiance témoignée tout au long de cette aventure.

## Comité scientifique

- Raphaëlle Bertho, Maître de Conférences, Université de Tours - Intru.
- Nicole Brenez, Professeure des Universités, Université Paris 3 - IRCAV.
- Kira Kitsopanidou, Professeure des Universités, Université Paris 3 - IRCAV.
- Martin Lefebvre, Professeur des Universités, Concordia University, Canada - Figura NT2.
- Mathieu Saladin, Maître de Conférences, Université Paris 8 - Teamed.
- Makis Solomos, Professeur des Universités, Université Paris 8 - Musidance.
- Myriam Tsikounas, Professeure des Universités, Université Paris 1 - ISOR/CRH.
- Benoît Turquety, Professeur des Universités, Université de Lausanne, Suisse - UNIL.

## Directeur de publication

Vincent Lowy,  
Directeur de l'ENS Louis-Lumière

## Direction scientifique (n°15 – 2022)

Corsin Vogel et Sylvain Lambinet

## Ont participé à la rédaction de ce numéro

Camille Aubriot  
Valentin Bauer  
Alan Blum  
Léa Chevrier  
Hugo Corbel  
Danièle Dubois  
François-Xavier Féron  
Cyril Holtz  
Jean-Pascal Jullien  
Sylvain Lambinet  
Clémence Lavigne  
Alexis Ling  
Vincent Lowy  
Jean-Marc Lyzwa  
Jean-Christophe Messonnier  
Laurent Millot  
Benoît Turquety  
Corsin Vogel

## Pilotage de projet

Jérémy Piot  
(j.piot@ens-louis-lumiere.fr)

## Mise en page

Jérémy Piot

## Création graphique

Olivier Sallon

## Enseignants-chercheurs de l'ENS Louis-Lumière

David Faroult,  
Maître de conférences  
Véronique Figini,  
Maître de conférences  
Pascal Martin,  
Professeur des Universités  
Laurent Millot,  
Maître de conférences  
Giusy Pisano,  
Professeure des Universités  
et au titre de la Direction : Vincent Lowy,  
Professeur des Universités

## Diffusion

ENS Louis-Lumière

## Copyright

Chaque auteur pour sa contribution.  
L'ENS Louis-Lumière pour l'ensemble.

Imprimé en France  
par l'imprimerie STIPA  
Achevé d'imprimerie en juillet 2022  
Dépôt Légal : juillet 2022  
ISBN : 978-2-9576795-1-5  
ISSN : 2-9520267-9-3

Toute reproduction, même partielle  
de cet ouvrage est interdite.  
Sa copie ou sa reproduction par quelque procédé que  
ce soit, photocopie, microfilm, bande magnétique,  
disque ou autre, constitue une contrefaçon passible  
des peines prévues par la loi du 11 mars 1957 sur la  
protection des droits d'auteur.

# L'ÉCOLE NATIONALE SUPÉRIEURE LOUIS-LUMIÈRE

L'École nationale supérieure Louis-Lumière propose une formation initiale professionnalisante, théorique et pratique, technique et artistique. Placée sous la tutelle du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation, l'École est un établissement public qui recrute à Bac+2 par voie de concours. Elle dispense un enseignement dans le cadre de trois masters – cinéma, son, photographie – couronné par un diplôme de niveau Bac+5

La scolarité est gratuite pour les 150 étudiants.

L'ENS Louis-Lumière participe à des projets de recherche et dispense des cours en formation professionnelle continue.

Créée en 1926 sous l'impulsion de personnalités comme Louis Lumière ou Léon Gaumont, pionnière des écoles de cinéma et de photographie, l'École a emménagé à la Cité du Cinéma en juillet 2012. En 2013, elle a ouvert une classe "égalité des chances" en soutien à la préparation de son concours. Depuis 2014, elle est membre associé de l'Université Paris Lumières (UPL).

## La Recherche

L'ENS Louis-Lumière développe une recherche appliquée et théorique, associée à la création, qui prend forme à travers des projets spécifiques, originaux, à la croisée de l'art, de la science et de la technologie.

L'ENS Louis-Lumière est impliquée au sein de l'EUR ArTeC, une Graduate School à la française intégrée à l'Université Paris-Lumières. Lancée en septembre 2018, ArTeC nous permet d'ouvrir un échelon doctoral pour nos étudiants, à travers une réflexion sur l'outil et la technique, les dispositifs de médiation, l'éthique du futur et l'âge du post-numérique.

Les projets de recherches menés par l'ENS Louis-Lumière et ses partenaires ont comme horizon commun des problématiques qui sont également au cœur de la formation et s'articulent autour de quatre axes transversaux et transdisciplinaires :

- › Innovations technologiques et pratiques associées
- › Recherche et création artistique
- › Histoire des techniques, des technologies et des pratiques
- › Expertise et caractérisation scientifique

**Le cahier *Louis-Lumière*, a pour ambition de mettre en valeur les relations entre la recherche et la création. Il s'adresse aussi bien aux chercheurs qu'à toutes celles et tous ceux intéressés par l'évolution de la réflexion dans les domaines de l'image et du son : professionnels, industriels, responsables d'institutions culturelles et de lieux de diffusion, étudiants...**

Précédentes éditions :

**2003 :** Questions de cinéma

**2004 :** Espaces pluriels, images et son

**2005 :** Territoires audiovisuels, errances, itinérances et frontières

**2006 :** Les dispositifs (textes issus d'une colloque international organisé par l'ENS Louis-Lumière et l'Université de Marne-la-Vallée)

**2007 :** Coupe, découpe, découpage

**2008 :** Vide, vacuité, désœuvrement

**2009 :** Nouvelles perspectives pour les photographes professionnels (actes d'un colloque organisé au Sénat par l'ENS Louis-Lumière)

**2010 :** Un cinéma du subjectif (Actes du colloque organisé par l'ENS Louis-Lumière et l'École supérieure d'Audiovisuel de l'Université Toulouse II le Mirail)

**2015 :** Mémoires d'écoles

**2016 :** Archéologie de l'audiovisuel

**2018 :** Le cinéma face aux histoires du regard

**2019 :** Art filmique et expérimentations optiques contemporaines

**2020 :** Les dispositifs immersifs : vers de nouvelles expériences de l'image et du son ?

**2021 :** Aaton : le cinéma réinventé

Disponibles sur commande pour la version papier

En ligne sur les sites [www.persee.fr](http://www.persee.fr) et [www.ens-louis-lumiere.fr](http://www.ens-louis-lumiere.fr)



