

Ecole Nationale Supérieure Louis-Lumière

MÉMOIRE DE FIN D'ÉTUDES

Sonoriser et spatialiser la voix pour le spectacle vivant :
la gestion des mouvements des acteurs

Clément Vallon

Directeur Interne : Thierry Coduys
Directeur Externe : Marc Piera

Rapporteurs : Franck Jouanny
Frank Gillardeau

Soutenu le 14 juin 2019 à Saint-Denis

Table des matières

Résumé	5
Abstract	7
Introduction	9
Enjeux estétiques	10
Porter la voix : l'art de la technique vocale	12
Le microphone ouvre la voie	13
Le problème de la voix dissociée du corps	15
1 Rappels sur l'audition	19
1.1 Mécanismes de localisation	19
1.1.1 Indices monoraux et interauraux	19
1.1.2 Le rôle de l'acoustique	20
1.2 Effet Haas ou loi du premier front d'onde	21
I État de l'art	23
2 Pratiques courantes de sonorisation	25
2.1 Prendre le son de la voix	25
2.2 Systèmes de diffusion « classiques »	27
2.2.1 Stéréophonie (L/R)	27
2.2.2 Stéréophonie avec point central (L/C/R)	29
2.2.3 Stéréophonie avec point central (L/C/R) en système A/B	30
3 Pratiques courantes de spatialisation	33
3.1 Techniques « simples » de spatialisation	33
3.1.1 Panoramique L/R ou L/C/R et AUX	33
3.1.2 Spatialisation par matrice	34
3.2 Des matrices au mixage objet	35
3.2.1 Les premières solutions informatisées	35
3.2.2 Vers les processeurs et le mixage objet	35

II	Spatialisation de la voix avec les « nouveaux » outils	39
4	Acquisition de la position et du mouvement	41
4.1	Suivi de la position par la méthode cognitive	42
4.2	Suivi de la position par automation et time-code	45
4.2.1	Caractérisation des outils d'automation	45
4.2.2	Outils à disposition	46
4.2.3	Conclusion sur l'utilisation d'automations	50
4.3	Suivi automatique de la position	50
4.3.1	Caractéristiques des systèmes de localisation	51
4.3.2	Les méthodes de localisation	53
4.3.3	Les technologies de localisation	56
4.3.4	Les systèmes disponibles en 2019	60
4.4	Interfaçage (protocoles de communication)	62
4.4.1	Open Sound Control (OSC)	62
4.4.2	Real Time Tracking Protocol - Motion (RTTrPM)	63
4.4.3	MIDI Show Control (MSC)	64
4.5	Corrélation	65
5	Reproduire le mouvement	67
5.1	Algorithmes à synthèse de champs sonores	68
5.1.1	Wave Field Synthesis, W.FS	68
5.1.2	Ambisonie	72
5.2	Algorithmes multicanaux	78
5.2.1	Distance Based Amplitude Panning (DBAP)	78
5.2.2	Vector Base Amplitude & Intensity Panning (VBAP, VBIP)	79
5.2.3	Layer Based Amplitude Panning (LBAP)	79
5.2.4	Panning d'amplitude et de délais basé sur la distance	80
6	Mixer en conséquence	81
6.1	Précision de la localisation par traitement du signal	81
6.1.1	Traitement de la source : délai, atténuation et filtrage	81
6.1.2	Traitement auxiliaire : réverbération	84
6.2	Vers le mixage 100% objet	86
6.2.1	Habitudes de mixage à conserver	86
6.2.2	Possibilité offertes par le mixage objet	90
7	Description de la partie pratique du mémoire	93
	Conclusion	97
	Bibliographie	103

Résumé

La sonorisation de la voix dans le spectacle vivant (théâtre, opéra, comédie musicale) est présente dans de très nombreuses productions. En comédie musicale elle est même devenue pratiquement omniprésente. Les pratiques courantes de sonorisation placent la voix au centre du dispositif de sonorisation, pratique qui produit une dé-localisation de la voix, et peut s'avérer gênante à certaines oreilles, voire venir diminuer la capacité d'attention des spectateurs lorsque de nombreux acteurs interviennent sur scène.

De nombreux procédés pour répondre à ce problème ont été utilisés par le passé, mais il se dessine aujourd'hui de nouvelles solutions. Certains précurseurs ont proposé dès la fin des années 1990 des processeurs de spatialisation sonore pensés pour la voix. C'est cependant au cours des six dernières années que les processeurs de spatialisation sonore ont commencé à fleurir, offrant un large choix aux sonorisateurs. Dans le même temps, les technologies de localisation des acteurs sur scène se sont améliorées. Cette conjoncture fait qu'il semble aujourd'hui plus simple que jamais de mettre en œuvre de tels systèmes, les anciennes techniques restant toujours valables pour des productions n'ayant pas les moyens de mettre en œuvre de telles solutions.

De par leur nature, ces solutions bousculent les habitudes que peuvent avoir les sonorisateurs en termes d'écoute, de mise en place et d'outils à utiliser. Nous proposons donc dans ce mémoire de passer en revue les différents concepts qui nous semblent nécessaires de connaître lorsque l'on s'intéresse à ce type de sonorisation, ainsi que les différentes problématiques auxquelles peuvent être confrontés les sonorisateurs.

Une partie pratique de mémoire consistant en une série d'essais au Théâtre National de Chaillot aura pour but de relever les différences entre quelques-uns des algorithmes de spatialisation à disposition, et d'analyser comment peut être optimisée l'intégration d'un système de localisation pour spatialiser les sources à l'emplacement exact de l'interprète.

Mots clefs

Spatialisation sonore, sonorisation de la voix, mixage objet, spectacle vivant, localisation.

Abstract

The sound reinforcement of voices is used in many shows (theater, opera and musicals). In musicals it even became quite impossible to see a show without it. The usual techniques for voice reinforcement place the voices at the central position of the sound system, which causes the voice to be emitted far away from the performer's position, and can seem odd to some spectators, and even diminish the focus of the audience when many performers are on stage.

Several methods that tried to resolve this issue were used in the past, but nowadays new tools have emerged. Some forward-thinking processors were developed at the end of the 1990s specifically for voice reinforcement. But during the past six years, a large number of immersive sound processors emerged, offering sound designer wide possibilities of choice. At the same time, performers' stage tracking techniques improved. This context makes us think that it is easier than ever to use such systems, previous techniques still being relevant for smaller-scale productions.

By nature, these new techniques disrupt sound designers' habits and customs, in terms of listening, design and tools use. Therefore we offer in this master's thesis to review the different concepts that seem important to us, for the designer to be aware of when getting to use such voice reinforcement systems, and the different problematics that a designer can face.

A practical side of the thesis will consist of tests at the Théâtre National de Chaillot. It will aim at observing differences between some available immersive sound algorithms, and analyzing how to optimize the interaction between the performers localization system and the immersive sound processor.

Keywords

Immersive sound, voice sound reinforcement, object-oriented mixing, source-oriented reinforcement, live performance, shows, localization.

Introduction

La voix a toujours été une problématique centrale dans le spectacle vivant —qu’il s’agisse de l’opéra ou du théâtre—, à travers la technique vocale, l’acoustique puis la sonorisation. Les siècles ont façonné la voix des artistes et leur manière de la projeter : des techniques vocales spécifiques se sont développées dans l’opéra, en fonction des genres, ainsi que dans le théâtre.

L’arrivée du microphone a changé le rapport à la voix : son amplification a eu pour conséquence d’en modifier les caractéristiques. Ceci a valu au microphone d’être adulé par certains, mais honni par d’autres. Des genres comme la comédie musicale l’ont rapidement acquis, tandis que son usage au théâtre et à l’opéra continue de faire régulièrement débat. L’un des griefs que l’on trouve le plus à l’encontre de la sonorisation de la voix est qu’elle rompt le lien de localisation entre le corps de l’acteur, et sa voix, émise à un autre emplacement : celui des haut-parleurs.

Les sonorisateurs ont toujours trouvé des solutions pour donner l’impression que le système de sonorisation n’existait pas : autrement dit, ils ont trouvé des méthodes permettant de donner l’illusion que le son amplifié provenait du corps de l’acteur, ou tout du moins suffisamment proche du corps de l’acteur pour que le cerveau s’y méprenne. La mise en œuvre de ce type de solution semble clairement minoritaire à l’heure actuelle, mais elles connaissent depuis quelques années un tournant qui pourrait bien s’avérer décisif. En effet, ce type de systèmes de sonorisation impliquait auparavant une certaine complexité matérielle et logistique, mais avec l’avènement des réseaux audionumériques et de l’informatique dans le spectacle vivant, ce type d’installation semble avoir perdu en complexité de mise en œuvre matérielle.

Un survol des enjeux esthétiques liés à la sonorisation de la voix nous permettra en introduction d’envisager la dimension artistique du travail de sonorisateur, et l’impact qu’il pourra avoir sur une production à travers ses choix de sonorisation et de mixage.

La première partie du mémoire est consacrée à un état de l’art des techniques de sonorisation, ou nous aborderons d’abord les systèmes où la voix est statique, puis nous présenterons les premières techniques de spatialisa-

tion de la voix, avant de dresser une chronologie des principaux processeurs de spatialisation sonore aujourd'hui à disposition des sonorisateurs.

La deuxième partie du mémoire sera consacrée à l'analyse des « nouveaux » outils qui permettent de spatialiser la voix. Ceux-ci sont apparus et se sont perfectionnés au cours des deux dernières décennies. Cette partie s'intéresse spécialement aux processeurs de spatialisation sonore, et aux systèmes de localisation des acteurs (ceux-ci permettent d'asservir la position des sources virtuelles sur la position réelle des interprètes sur scène). Le premier chapitre sera donc dédié à l'acquisition de la position des interprètes, sous quelque forme que ce soit, puis le second chapitre présentera les algorithmes de spatialisation sonore proposés par les processeurs. Enfin nous concluons sur un chapitre portant sur les techniques de mixage avec ces nouveaux outils.

Nous présenterons ainsi dans ce mémoire les concepts et les connaissances permettant d'aborder dans la pratique la spatialisation sonore de la voix dans le spectacle vivant. Les explications de concepts scientifiques n'entrent pas dans les détails tels que les équations, car elles n'aident pas le concepteur sonore à réaliser son travail, et qu'elles sont étudiées de manière extensive dans de nombreux ouvrages. Nous préférons aborder les conséquences pratiques de tous ces concepts scientifiques, et l'implication qu'elles auront sur le travail du concepteur sonore.

Enjeux esthétiques

En touchant à la voix des acteurs, le microphone s'attaque à leur outil le plus précieux. Allant du simple renfort sonore aux créations les plus folles, l'arrivée du microphone a mené les arts de la scène vers un Nouveau Monde débordant de possibilité. Avec les évolutions technologiques et l'inventivité sans faille des concepteurs sonores, il semble aujourd'hui désuet d'adopter vis-à-vis du microphone simplement une position pour ou contre. S'il est détesté de certains, c'est souvent par méconnaissance des possibilités qu'il offre, et peut être aussi car il arrive encore aujourd'hui qu'il ne soit pas intégré à un spectacle avec suffisamment de précautions. Car sonoriser des voix est porteur de nombreux enjeux esthétiques qui doivent être pris en compte par le concepteur sonore, évidemment, mais dont doivent avoir également conscience tous ses partenaires, en particulier le metteur en scène, les interprètes, et l'équipe de production. En effet, si les microphones sonnent parfois mal c'est peut-être à cause du manque de connaissance ou de sensibilité de l'ingénieur du son, mais cela peut être aussi car on ne lui a pas donné les moyens (humains, temporels, et matériels) de travailler comme il aurait fallu. La production d'un spectacle est un travail de groupe, et le son est l'affaire de

tous. Il est donc important que tous les partenaires comprennent les enjeux esthétiques qui se cachent derrière la sonorisation.

Question.s de genre.s

Commençons tout d'abord par définir à quels types de spectacles nous nous intéressons dans ce travail. Le spectacle vivant est un domaine vaste où la voix n'est pas nécessairement présente dans la forme qui nous intéresse ici. L'objet de notre étude est : tout type de spectacle comportant un argument et des dialogues, qu'ils soient parlés ou chantés. Cela inclut donc évidemment le théâtre, quel qu'il soit (antique, classique ou contemporain) ainsi que les spectacles musicaux comme l'opéra et la comédie musicale. Nous n'étudierons pas le cas du concert, du spectacle apparenté au cabaret, ou encore de la danse.

D'autre part, nous ne considérerons ici que les spectacles se jouant dans des théâtres ou des lieux de représentations à taille humaine, ne nécessitant pas des écrans pour voir les interprètes. Les raisons de cela sont simples : nous allons nous intéresser à la localisation des sons, et notre oreille n'a pas la finesse de distinguer la localisation de sources très éloignées de nous. Dès lors que la scène ne représente plus que quelques centimètres dans notre champ de vision, il ne sert à rien de spatialiser les sons en cohérence avec ce qui est visible sur scène, car d'autres problématiques spécifiques entrent alors en jeu. Ces productions n'étant pas majoritaires dans l'univers du théâtre, de l'opéra et de la comédie musicale, nous nous contenterons de considérer que les représentations ont lieu dans des théâtres.

Quand et pourquoi sonorise-t-on la voix ?

Quel que soit le type de spectacle, le cas de sonorisation le plus fréquent est le renfort vocal, souvent utilisé lorsque la représentation a lieu dans un endroit pas adapté au spectacle vivant. Prenons l'exemple d'un spectacle joué en plein air devant une très large audience. Étant donné l'absence d'une acoustique favorable, il est probable que les voix ne portent pas assez pour que tout le monde les entende. Sonoriser les voix dans ce cas de figure permet de les rendre audibles de tout le public. Nous pourrions également imaginer une troupe de théâtre souhaitant interpréter un texte dans un espace désaffecté s'avérant très réverbérant. Dans ce cas, les voix pourront porter suffisamment pour que le public les entende, mais la réverbération du lieu risque de brouiller l'intelligibilité de la parole, et à moins d'immenses efforts de débit vocal et de diction, le public ne comprendrait rien à ce qui se dirait. Cette fois-ci, sonoriser les voix permettrait de récupérer l'intelligibilité du texte.

Ces deux cas de figure présentent clairement des situations où l'objectif de la sonorisation est d'apporter une solution discrète, presque purement technique, où les enjeux esthétiques liés au microphone sont moindres, car le microphone ne contribue pas à l'esthétique des voix —les interprètes ne modifient pas leur voix et leur jeu par rapport au microphone. En somme, l'on chercherait dans ce cas de figure plutôt une sonorisation « naturelle » ou dite « transparente ».

Mais il existe également des cas de figure inverses, où la transparence n'est plus le critère principal : lorsque le microphone est utilisé pour produire un effet. Un exemple célèbre est la voix du Fantôme de l'Opéra dans la comédie musicale éponyme de Andrew Lloyd Webber, qui peut résonner dans le théâtre avec des niveaux sonores surhumains, ou bien dans les opéras contemporains où de nombreuses transformations peuvent être appliquées sur la voix.

Le troisième cas de figure, plus subtile, concerne l'utilisation de la sonorisation de la voix comme support pour une autre technique vocale, qui serait impossible sans le microphone. Cela se retrouve dans le théâtre contemporain pour faire murmurer des acteurs, ou dans la comédie musicale, où le microphone est devenu indispensable au chant.

Finalement, tout comme les interprètes (comédiens et chanteurs) vont travailler leur technique vocale en fonction du genre du spectacle, du lieu de représentation et des volontés de mise en scène, le sonorisateur travaille en prenant en compte les mêmes critères.

Porter la voix : l'art de la technique vocale

L'une des premières leçons que reçoit tout étudiant en art dramatique de la part de son professeur est : « ne parle pas dos au public, et projette ta voix jusqu'au fond de la salle ». De tout temps, le théâtre, tout comme l'opéra, a été purement acoustique. Pour se faire entendre, les comédiens travaillent donc leur voix en conséquence et apprennent à avoir une voix « qui porte », même dans les nuances du jeu : un chuchotement de théâtre n'aura pas la même sonorité d'un chuchotement au cinéma. Cela se joue également dans la mise en scène : un chuchotement au théâtre est dit face au public, en général plutôt à l'avant-scène, dans l'intimité des feux de la rampe.

Avec la taille des salles, la technique vocale change également : il suffit d'aller à l'opéra pour remarquer que dans les scènes parlées (dans les *singspiel* ou opéra-comiques et autres genres mêlant jeu et chant), les chanteurs prennent une « grosse voie » dont la nature est bien différente de celle que l'on entend au théâtre, puisqu'elle est une hybridation entre la technique de chant et le débit parlé. Les techniques de chant ont justement été façonnées

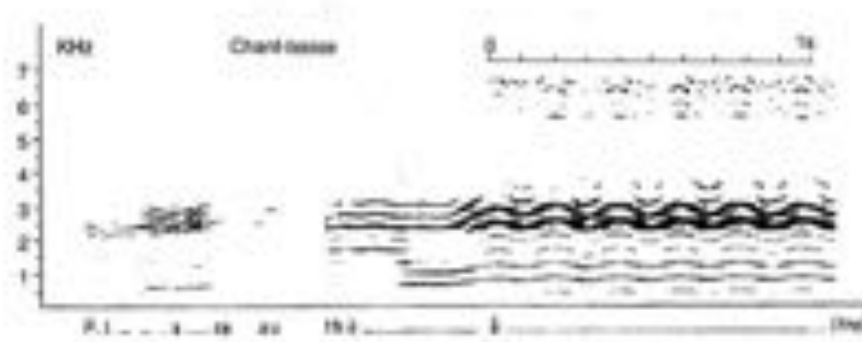


FIGURE 1 – Sonagramme d'un chanteur lyrique basse. Le formant présentant le plus d'énergie se trouve entre 2 kHz et 3 kHz, surpassant même la fréquence fondamentale. Cette zone coïncide avec la zone fréquentielle la plus sensible de notre audition. (Source : CASTELLENGO, Michèle in *Le livre des techniques du son*, 4e édition, MERCIER, Denis (ed.), Dunod, 2010, p. 79).

par la nécessité d'être entendu jusqu'au fond d'une maison d'opéra, même par-dessus un orchestre symphonique, et ceci se retrouve dans le timbre de leur voix. De manière analogue, les chanteurs de rue par exemple ont une voix nasillarde qui leur permet d'être entendus même dans la foule, les chanteurs de rock ont eux aussi développé des techniques vocales propres qui leur permettent de ressortir parmi les guitares hurlantes.

Ces caractéristiques des techniques vocales se retrouvent en analysant ces voix qui portent. En fonction de leur registre, ces voix présentent¹ un formant situé entre 2 kHz et 4 kHz qui s'avère correspondre à la zone la plus sensible de notre audition (fig. 1).

Le microphone ouvre la voie

Pour chacun des trois genres étudiés, l'ajout du microphone a ouvert la voie à de nouvelles possibilités. Chaque genre a vu s'opérer des transformations spécifiques.

Transformations de la voix au théâtre

En ajoutant un microphone aux comédiens, l'on retire la nécessité de porter la voix, le système électroacoustique s'en chargeant pour eux. La troupe peut néanmoins continuer de jouer de la même manière, mais cela ouvre pourtant le champ à de nombreuses possibilités : il est désormais possible de se retourner et pourtant de rester audible, de chuchoter naturellement et rester audible, de murmurer, de réduire l'articulation, de se cacher du public

1. Mercier, Denis et Castellengo, Michèle, *Le livre des techniques du son, tome 1 : Notions fondamentales*, 4^e édition. 2010.



FIGURE 2 – La chambre vitrée de *Angels in America*, mise en scène de Aurélie Van Den Daele en 2015

en restant audible. La technique vocale peut donc se reconfigurer autour de cet outil.

Puisque la voix reste audible où que soit l'acteur, le microphone permet d'explorer de nouvelles scénographies et de nouvelles occupations de l'espace scénique. Le metteur en scène peut maintenant par exemple placer les comédiens tout à fait au fond de la scène, comme dans *Richard III*, Acte III, Scène 5, dans la mise en scène de Thomas Jolly (2015), et pourtant les entendre clairement grâce à un microphone caché dans le décor. Dans la mise en scène d'*Angels in America* par Aurélie Van Den Daele (2015), la scénographie comporte une chambre transparente, de laquelle aucun son ne peut sortir, puisqu'elle est censée pouvoir se remplir de fumée. Les microphones des comédiens permettent là par exemple d'étendre l'espace de jeu à cette partie du décor (fig. 2).

Transformations de la voix à l'opéra

Du côté de l'opéra, le microphone n'a pas changé la technique vocale dans l'interprétation des œuvres de répertoire, mais a néanmoins ouvert le champ à la mise en scène comme au théâtre, permettant d'entendre les chanteurs même s'ils sont placés là où ils ne devraient pas être. C'est plutôt du côté de la musique contemporaine (opéra contemporain, théâtre musical) que le microphone a permis une évolution des techniques vocales, jusqu'à l'extrême dans l'opéra *Kein Licht* (2017) composé par Philippe Manoury, où le compositeur, à l'aide de traitements conçus par l'Ircam, utilise les microphones pour transformer la voix *parlée* des artistes en voix *chantée*.

Transformations de la voix dans la comédie musicale

Même si aujourd'hui il s'agit de l'unique genre scénique où le microphone est quasi systématiquement présent, la comédie musicale a commencé comme le théâtre et l'opéra, par être un genre uniquement acoustique. Ce genre puise ses sources musicales dans les musiques populaires (au sens anglo-américain du terme, *popular music*² tel que défini³ par Philip Tagg). À partir du tournant des années 1920, les musiques populaires sont bouleversées par l'arrivée du microphone. De chanteurs avec du coffre, l'on passe aux crooneurs, qui développent une technique de chant proche parfois du fredonnement, impossible en concert sans utiliser de microphones. Son usage se généralise ensuite avec les musiques enregistrées, et s'impose naturellement en concert, permettant à l'interprète de jouer avec son timbre, entre douceur et puissance vocale. Ceci nécessite l'utilisation d'outil de contrôle du niveau (compresseur, mixage de niveau) afin de mixer et d'ajuster la voix au reste des instruments, qui doivent eux aussi être travaillés à la console de mixage. Le microphone est donc devenu indispensable dans les musiques populaires. Sa première utilisation remonterai⁴ à 1964 sur Broadway dans *Funny Girl*. Son adoption dans la comédie musicale est peu à peu devenue inévitable à partir des années 1980, *Cats* d'Andrew Lloyd Webber ouvrant le bal cette année-là.

Le problème de la voix dissociée du corps

La sonorisation de la voix a connu et connaît toujours de nombreux détracteurs, qui lui adressent des reproches variés, souvent valables, mais parfois parfaitement discutables. Parmi les arguments les plus discutables se retrouvent souvent ceux qui tiennent de la nostalgie ou du conservatisme. Certains auteurs tels que Mark N. Grant⁵ défendent que les nouvelles techniques vocales reposant sur le microphone et développées dans les musiques actuelles ont moins de valeur que les techniques vocales anciennes. Doit-on considérer que parce qu'un chanteur de musiques actuelles utilise un microphone pour exercer son art, il a moins de valeur qu'un chanteur lyrique ? A nos yeux, il ne peut pas être fait de comparaison de valeur entre des techniques simplement différentes, toutes deux héritières d'un passé et d'un contexte les ayant façonnées différemment.

2. Tagg, Philip, "Analysing popular music : theory, method and practice", *Popular Music* 1982, Nr. 2, p. 37-65.

3. Dans cet ouvrage, Philip Tagg distingue la musique savante (*art music*), la musique traditionnelle (*folk music*) et la musique populaire (*popular music*) à l'aide de critères spécifiques qui permettent de classer la musique des comédies musicales dans les musiques populaires.

4. Burston, Jonathan, "La technologie audio, le corps chantant reconfiguré et les megamusicals", *Théâtre/Public* 2011, Nr. 199, p. 95-99.

5. Grant, Mark N., *The Rise and Fall of the Broadway Musical*, Boston: Northeastern University Press, 2004.

Il en va de même pour le théâtre, lorsqu'il n'est pas utilisé pour du renfort sonore, parler avec un microphone n'est pas une preuve de faiblesse vocale, mais une volonté de développer une technique vocale différente. Il ne peut pas être fait de hiérarchisation de valeur entre ces deux pratiques, elles sont simplement différentes. Le gout pourra être dirigé en faveur de l'un ou de l'autre néanmoins.

Des chercheurs comme Jean-Marc Larrue⁶ analysent cette hiérarchisation fictive comme provenant d'un dernier sursaut pour défendre le théâtre face au cinéma et la radio du temps où ceux-ci semblaient une menace pour le spectacle vivant. Pour simplifier, les gens ont considéré par le passé que ce qui permettait de distinguer le théâtre des médias (cinéma, télévision, radio) était la co-présence : cette faculté à percevoir que l'on est dans le même espace que l'interprète. La co-présence se serait cristallisée autour de la voix, ayant au théâtre la caractéristique d'être non médiatisée (c'est-à-dire non sonorisée). L'auteur analyse que l'arrivée de la voix médiatisée était inévitable au théâtre, suivant un mouvement global dans l'ensemble de la société. En se basant sur cette différenciation, il était évident son utilisation serait rejetée. L'utilisation du microphone au théâtre devient alors presque hérétique si l'on se place dans cette logique, puisqu'elle pulvériserait ce qui fait le propre du théâtre : la co-présence. Ce type de réactions épidermiques est documenté par Gareth Fry, un célèbre concepteur sonore britannique, dans une anecdote juteuse prenant place dans les journaux britanniques⁷ :

In 1999 the National Theater made the front page of several national newspapers because it was using radio microphones on actors in several plays, including, to the horror of many, a production of Shakespeare's *Troilus and Cressida*. There then followed much outcry and disdain about microphones resulting in lazy projection from actors, about their offering 'television performance', and the state of actor training. What wasn't noted was that radio microphones had been used for about six months by the time it was reported, and several shows had opened, with the major theatre reviewers having attended all of these, and no one had noticed that radio microphones were being used! The only reason the newspapers became aware of it was because a staff member at the NT [National Theater, ndr] leaked the information to them.

The microphones were being used to feed the difficult areas of the auditorium where intelligibility was poor, by gently lifting the volume of the performers' voices in those areas. Whilst no one had noticed that radio microphones were being used, complaints from the audience sitting in the problematic areas about not being able to hear were significantly reduced.⁸

6. Larrue, Jean-Marc, "Son, présence et ontologie", *Le son du théâtre : XIXe - XXIe siècle* 2016.

7. Fry, Gareth, *Sound Design for the Stage*, The Crowood Press, avril 2019, p. 96.

8. En 1999 le National Theater fit la une de plusieurs journaux nationaux pour avoir utilisé des microphones sans fil sur les comédiens dans plusieurs pièces, dont, à l'effroi de beaucoup, une production de *Troilus et Cressida* de Shakespeare. Il en suivit un tollé et du mépris sur les microphones, cause de la projection paresseuse des acteurs, sur leur jeu tel qu'à la télévision,

Enfin, d'autres⁹ questionnent la légitimité de produire des spectacles dans des lieux non adaptés, au prix de devoir sonoriser la voix. L'on pourra alors opposer cette affirmation : si on veut faire une représentation dans un lieu atypique, sans que la sonorisation soit faite détrimement de la qualité de l'interprétation, et que le lieu apporte un plus à la création du spectacle, pourquoi ne pas le proposer ? D'autant que le théâtre, l'opéra et la comédie musicale sont certes des arts, mais dépendent également de leur public. Si le public est au rendez-vous et le spectacle de qualité, pourquoi ne pas le produire ?

Il n'en demeure pas moins que de nombreux arguments opposés au microphone restent valables. Daniel Deshays dans « L'impossible façade »¹⁰ cite ce qu'il nomme avec emphase « les facteurs du drame de la sonorisation » : « (1.) La dénaturation de la voix des acteurs ; (2.) La grossièreté des outils, la négligence de leur bon emploi et du savoir-faire qui lui est lié ; (3.) La délocalisation. ». Cet article est extrêmement précieux pour nous sonorisateurs, puisqu'il cite les causes techniques de toutes les gênes pouvant être exprimées par les profanes de la sonorisation. Sans entrer dans un argumentaire détaillé, il nous semble aujourd'hui d'une part que les facteurs évoqués sont surmontables, et d'autre part que les outils et techniques de prise de son, de mixage et de sonorisation ont atteint un niveau suffisant de qualité sonore pour ne plus être mis en cause.

Cependant il est intéressant de relever que lorsque l'auteur développe chacune de ces trois catégories, revient à chaque fois la question de la localisation : ou plutôt de la délocalisation. Dans le premier, la dénaturation de la voix est en partie imputée à « [la] perte de possibilité de localiser la source », due à l'utilisation d'un système de sonorisation type stéréo. Dans le deuxième point, c'est l'absence des informations de localisation de l'acoustique du lieu qui est reprochée, ou plutôt le fait que ces échos et réverbérations soient écrasés par le système de sonorisation. Enfin dans le dernier point, particulièrement dédié à cette problématique, révèle la source du problème : il faudrait pouvoir localiser le son à l'emplacement de l'acteur pour que l'auditeur n'ait pas de resynchronisation à opérer.

et sur la condition de la formation d'acteurs. Ce qui n'avait pas été relevé c'est que les microphones sans fils étaient utilisés depuis six mois lorsque l'on eut vent de l'affaire, et plusieurs spectacles avaient alors commencé, les plus grands critiques ayant assisté à leurs premières, sans remarquer l'usage de microphones sans fils ! La seule raison pour laquelle les journaux surent cela, c'est qu'un membre du National Theater leur avait fourni l'information.

Les microphones étaient utilisés pour sonoriser les zones difficiles de l'auditorium où l'intelligibilité était mauvaise, en augmentant légèrement le niveau des voix dans ces zones. Bien que nul n'ait remarqué que des microphones étaient utilisés, les plaintes du public assis dans les zones problématiques à propos de l'inintelligibilité des comédiens diminuèrent considérablement.

9. Deshays, Daniel, "L'impossible façade : l'usage du microphone et son incidence sur la scénographie des espaces", *Le son du théâtre : XIXe - XXIe siècle* 2016.

10. *Ibid.*

Dans cet écrit et dans de nombreux autres, la voix sonorisée séparée du corps de l'interprète semble le principal grief reproché, non plus alors au microphone, mais à la sonorisation. Mais depuis quelques années, une conjoncture technique semble plus que favorable à la conception de systèmes de sonorisation permettant justement de localiser les sons à l'emplacement de l'acteur. Nous explorerons au cours de ce mémoire les différents aspects techniques de la mise en œuvre d'une sonorisation de spatialisation de la voix.

Chapitre 1

Rappels sur l'audition

Avant de parler des pratiques courantes de sonorisation, quelques rappels concernant notre audition sont nécessaires. Dans ce chapitre, nous évoquons les principaux mécanismes de l'audition ayant trait à la localisation des sons, en retenant principalement les concepts intéressants pour notre étude. Ceux-ci sont principalement basés sur les caractéristiques morphologiques de notre tête, mais nous verrons que d'autres critères entrent en jeu, comme l'acoustique du lieu par exemple.

1.1 Mécanismes de localisation

1.1.1 Indices monoraux et interauraux

Notre appareil auditif utilise plusieurs mécanismes pour localiser les sons. Les deux mécanismes principaux sont les indices monoraux (relatifs à chaque oreille, individuellement) et les indices interauraux (relatifs aux deux oreilles). Les indices monoraux nous permettent d'évaluer principalement la distance ainsi que l'élévation d'une source, et de distinguer si elle se situe devant ou derrière nous. Cette évaluation se fait grâce à l'analyse de l'évolution du contenu spectral des sons¹. Les indices interauraux nous permettent quant à eux d'évaluer la position en azimut (c'est-à-dire la latéralisation) des sons. Ils se divisent eux-mêmes en deux familles, les différences interaurales temporelles et de niveau. Enfin, la localisation des sons est affinée grâce aux mouvements, même légers, de la tête.

Différences interaurales temporelles (ITD, *Interaural Time Differences*)

La largeur de notre tête induit une différence de temps d'arrivée du son entre nos deux oreilles, variant avec la position de la source. Lorsque la source

1. Blauert, Jens, *Spatial hearing : the psychophysics of human sound localization*, Revised édition. Cambridge: MIT Press, 1997.

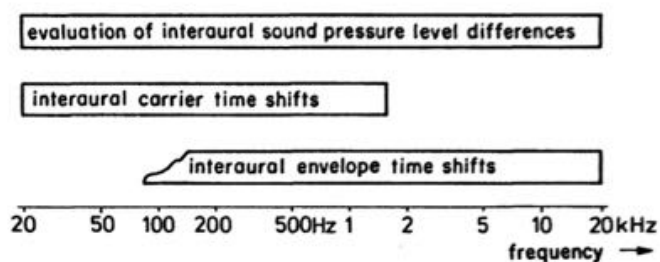


FIGURE 1.1 – « Plages de fréquence d'évaluation des différents attributs interauraux ». En haut sont les ILD, au milieu les ITD et en bas les ITD évaluées à l'aide de l'enveloppe du signal. (Source : Blauert, p. 164)

se trouve au centre, face à l'auditeur ou derrière lui, le décalage temporel est nul ; lorsque la source se trouve sur le côté il peut atteindre $0,65 \text{ ms}^2$. Notre audition est capable de localiser les sons à l'aide des différences de temps pour les fréquences jusqu'à $1,6 \text{ kHz}$, mais cette faculté diminue par la suite. Néanmoins, l'enveloppe sonore entre également en jeu dans la localisation, particulièrement au niveau de l'attaque et de l'extinction. Cette localisation via l'enveloppe d'un son se fait quelle que soit la fréquence de la porteuse, c'est à dire, quelle que soit la hauteur du son (à partir de $100\text{-}200 \text{ Hz}$)³.

Différences interaurales de niveau (ILD, *Interaural Level Differences*)

Le son s'atténue au cours de sa propagation, ce qui engendre des différences de niveau entre chaque oreille, en fonction de la position de la source. C'est ce phénomène qui est principalement exploité dans la stéréo, grâce à la technique de la source fantôme. Mais au-delà de l'atténuation dans l'air, d'autres facteurs introduisent des différences de niveau : ce sont les obstacles que constituent notre tête, nos épaules, ainsi que la morphologie de notre oreille. Toutes les réflexions, diffractions et autres phénomènes qu'ils engendrent sur le son vont modifier le contenu spectral des sons à notre oreille. Ces caractéristiques sont uniques pour chaque individu, en les reproduisant, il est possible de simuler, via l'écoute au casque, des scènes sonores 3D.

1.1.2 Le rôle de l'acoustique

Si les indices monauraux entrent en compte dans l'évaluation de la distance d'une source, la contribution de l'acoustique d'une salle permet également à notre oreille de percevoir cette information. Lorsqu'une source émet

2. Rumsey, Francis, *Spatial Audio*, Focal Press, 2001.

3. Blauert, Jens, *Spatial hearing : the psychophysics of human sound localization*, Revised édition. Cambridge: MIT Press, 1997.

un son dans une salle, celui-ci se propage dans de multiples directions.

Nous percevons d'abord le son direct, c'est-à-dire le son de la source s'étant propagé en notre direction. L'acoustique de la salle n'a aucune incidence sur ce son.

Puisque le son est rayonné dans plusieurs directions par la source, il se réfléchit contre les parois de la salle, puis arrive transformé par la salle à nos oreilles, et depuis différentes directions. Elles sont appelées *premières réflexions* ou *réflexions précoces* (*early reflections* en anglais), et nous arrivent dans les 100 ms après le son direct⁴. Elles sont porteuses d'information de localisation, puisqu'elles proviennent de directions précises, dépendentes du positionnement de la source dans le lieu.

Dans un troisième temps intervient le champ diffus qui est quand à lui est chaotique (ne peut pas être localisé) et homogène dans l'espace (ne dépend pas de la position de la source). Il porte lui des informations concernant la taille du lieu et sa sonorité.

Si l'on souhaite simuler acoustiquement la position d'une source sonore, le travail sur la réverbération aura une importance non négligeable, pour donner les bonnes impressions à l'auditoire. Cela peut relever d'un travail naturaliste (tenter de « coller » à l'acoustique du lieu, ou bien de reproduire une acoustique naturelle), ou bien d'un travail de création (tenter de représenter un lieu complètement différent, comme par exemple donner une acoustique de cathédrale à une petite salle insonorisée, ou carrément de créer des acoustiques imaginaires ou impossibles).

1.2 Effet Haas ou loi du premier front d'onde

L'effet Haas est connu sous plusieurs noms : loi du premier front d'onde, effet de préséance (ou « précédence », par anglicisme) et autres variantes. Ce phénomène physchoacoustique fait que si nous percevons deux sources distinctes, espacées, mais émettant le même son, alors, notre cerveau ne percevra qu'une seule source, dont la localisation sera celle de la source ayant émis le front d'onde étant arrivé le premier à nos oreilles, et ce, même si le premier signal est plus faible que le second.

Ce phénomène est très utilisé en sonorisation, car cela permet, par exemple, de renforcer le son provenant d'une enceinte un peu trop lointaine en ajoutant une nouvelle enceinte plus proche de l'auditeur. En retardant le son sur cette nouvelle enceinte (de sorte que le front d'onde atteigne l'auditeur après celui de la première enceinte), il sera possible de renforcer le niveau sonore, tout en conservant la sensation que le son vient de la première enceinte.

4. Fischetti, Antonio, *Initiation à l'acoustique*, Paris: Belin, 2003.

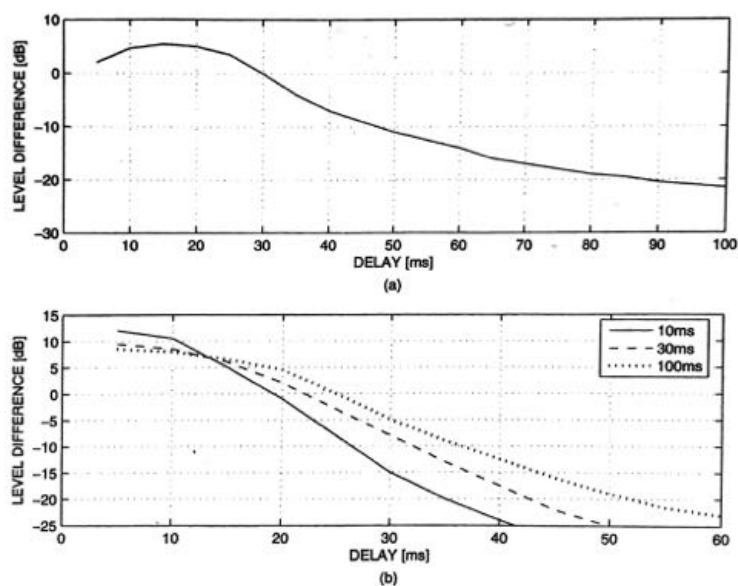


FIGURE 1.2 – Seuils d'écho pour (a) de la voix; (b) des pulsation de bruits blanc de différentes durées. (Source : Breebaart and Faller, *Spatial Audio Processing : MPEG Surround and Other Application*, John Wiley & Sons, 2007, p. 38)

Néanmoins ce phénomène possède une limite au-delà de laquelle notre oreille perçoit un écho. Cette limite varie avec la différence de niveau entre les deux sources ainsi que leur décalage temporel (fig. 1.2). L'ordre de grandeur de cet effet est de quelques dizaines de millisecondes. Il fonctionne mieux sur des sons continus que sur des sons percussifs, il n'est donc pas possible d'appliquer un délai aussi long sur une caisse claire (son percussif très bref) que sur une voix par exemple (son plus continu).

Première partie

État de l'art

Chapitre 2

Pratiques courantes de sonorisation

2.1 Prendre le son de la voix

Avant l'arrivée des microphones sans fil, les premières techniques de renfort sonore se faisaient avec des microphones suspendus à la cage de scène, cachés dans le décor, ou bien placés à l'avant-scène, qui captaient le son d'ensemble du plateau. Les bruits de pas, de vêtements ou de manipulation d'objets étaient donc captés dans les mêmes proportions que la voix. Pour quelques spectacles cependant, des microphones à main étaient utilisés, nécessitant soit d'être tenus par le chanteur, soit posés sur un pied. À l'évidence, cet usage n'est pas discret, il n'est pas étonnant que ce type de microphones aient été principalement utilisés du côté des *rock-musicals* ou des opéras rocks tels que *Jesus Christ Superstar* et *Starmania*, dont le style musical et l'esthétique justifiaient facilement d'avoir recours à de tels microphones.

C'est finalement la miniaturisation des microphones et des émetteurs (sans fil) qui a permis l'essor des voix sonorisées dans le spectacle vivant.

Aujourd'hui, les interprètes¹ portent un microphone miniature au niveau de leur visage. Il peut être soit fixé sur le visage, généralement au niveau du front, de la tempe ou bien de la joue, soit amené à peu près au niveau de la bouche, par une fine tige métallique. Les premiers sont dits « micros goutte », et les seconds sont appelés « microphone serre-tête », de par leur mode de fixation puisque la tige métallique est maintenue par un casque discret qui repose sur le tour d'oreille de l'interprète.

1. L'on utilisera ce terme pour désigner indifféremment des acteurs, chanteurs, danseurs,... ou bien les trois à la fois. En anglais l'on aurait parlé de *performer*.

Choix du type de microphone

Deux directivités de microphones peuvent être utilisées : les microphones omnidirectionnels et les microphones cardioïdes.

Les microphones omnidirectionnels captent le son dans toutes les directions, mais ils présentent néanmoins l'avantage d'avoir naturellement une réponse en fréquence plus plate que les cardioïdes. En raison de leur nature technologique², ils sont moins sensibles au vent et à l'effet de proximité que les microphones cardioïdes.

Les microphones cardioïdes quant à eux ne captent pas le son dans toutes les directions. Ceci les rend avantageux dans des environnements très chargés en niveaux sonores, ce qui permet d'isoler plus aisément la voix de l'artiste des autres sons présents sur le plateau.

Émission et réception du signal

La capsule du microphone est relié à un fil qui court le long de la tête et du dos de l'artiste, au bout duquel il est connecté à un émetteur qui transmet le signal audio du microphone à des récepteurs situés en coulisses. Des antennes situées dans la cage de scène permettent d'assurer une bonne réception du signal ainsi qu'une couverture de l'ensemble du plateau, voire de la salle si les artistes sont amenés à s'y déplacer. Le signal des microphones est ensuite transmis à la console de mixage.

Qualité sonore

À ses débuts³, la technologie sans fil présentait une plage de dynamique étroite ainsi qu'un bruit de fond non négligeable, rendant quasiment son usage impropre pour les arts de la scène. Les évolutions technologiques l'ont rendue tout à fait performante pour une telle utilisation, et bien qu'acceptable en termes de qualité sonore et de discrétion depuis des années, le matériel continue de se miniaturiser et de s'améliorer en qualité audio.

Les microphones actuels sont si fins qu'ils peuvent être aisément dissimulés au ras de la chevelure des artistes ou bien même intégrés dans des prothèses de peaux, et l'un des constructeurs majeurs de microphones, DPA, vient de lancer début 2019 un modèle encore plus miniature.

Les émetteurs quant à eux sont assez petits pour être cachés sous le costume, même pour les scènes où les tenues sont des plus légères. Leur qualité audio ne cesse d'augmenter, les dernières évolutions de ce côté étant la

2. Rappelons qu'il s'agit de microphones de pression, contrairement aux microphones cardioïdes, qui eux sont en conséquence plus sensibles au vent et au souffle.

3. Leonard, John A., *Theater Sound*, A C Black, 2001.

conversion du signal analogique vers numérique effectuée avant la transmission sans fil, permettant ainsi d'avoir une chaîne audio complètement numérique débutant au plus près du microphone, et ainsi d'éviter la dégradation du signal audio lors de la transmission sans fil.

2.2 Systèmes de diffusion « classiques »

Une fois le son mixé dans la console de mixage, le signal audio doit être retransmis sur des haut-parleurs pour le public. Sans entrer dans les détails de la conception des systèmes de diffusion, nous allons ici nous intéresser aux systèmes que l'on rencontre couramment pour sonoriser des voix.

2.2.1 Stéréophonie (L/R)

La stéréophonie est la base de nos systèmes de diffusion aujourd'hui. Le principe est simple, grâce à deux enceintes, il est possible de reconstruire une scène sonore sur la largeur de ces deux haut-parleurs. La musique, les chaînes hi-fi, les salles de spectacle... tous nos systèmes de retransmission du son sont basés sur la stéréophonie. Dans les salles de spectacle, cela se concrétise par deux points de diffusion de part et d'autre de la scène.

Nous ne nous intéressons ici qu'au système principal de diffusion. Les haut-parleurs complémentaires (ou *fills*) ne seront pas évoqués dans ce mémoire.

Le principe de la source fantôme

Si l'on place un auditeur en face de deux enceintes qui reproduisent exactement le même signal, il aura la sensation que le son provient du centre des deux enceintes. Ce phénomène est appelé source fantôme, et il permet de localiser les sons entre deux haut-parleurs. En augmentant le volume de l'enceinte droite, l'image fantôme va se décaler vers la droite, et faire de même vers la gauche si celle-ci est plus forte que l'enceinte droite. Ce procédé fonctionne avec plus de deux enceintes, en ajustant les niveaux entre chaque haut-parleur, il se crée des images fantômes entre chacun d'eux.

Le problème du *sweet spot*

L'inconvénient de la stéréophonie est le « *sweet spot* », cette position idéale unique, qui rend la perception de l'image stéréo peu performante pour une large audience.

Pour bien percevoir l'image stéréophonique sans décalage, il est nécessaire que l'auditeur soit placé au centre du dispositif, idéalement en formant

un triangle équilatéral avec les deux enceintes. Lorsque l'auditeur se décale sur le côté, il ne peut plus percevoir correctement la stéréophonie ; plus il se décale, plus il aura progressivement l'impression que le son ne provient que d'une enceinte, celle qui est la plus proche de lui.

Ce phénomène s'explique par les mécanismes à l'œuvre dans notre perception de la localisation des sons. En effet, comme nous l'avons vu au chapitre 1.1 page 19, notre système auditif localise les sons à l'aide de deux phénomènes, appelés indices de localisation :

- les différences de niveau sonore ;
- les différences de temps d'arrivée.

Si l'auditeur fait face à une source sonore, le son de cette source arrivera au même moment et au même niveau sonore à ses deux oreilles. S'il se décale, le son de la source atteindra une oreille avant l'autre, et à un plus fort niveau. Le cerveau utilisera alors ces différentiels de niveau et de temps pour déduire la localisation de la source.

Cela aura alors deux conséquences : le niveau sonore de l'enceinte opposée sera atténué, et un effet de préséance se produira en faveur de l'enceinte la plus proche.

Ainsi, pour le spectateur qui n'est pas au centre (et a plus fort raison pour celui qui est carrément face à une des deux enceintes), il localisera les sources sonores comme provenant de l'enceinte la plus proche de lui.

Mixer la voix au centre

Pour de nombreuses raisons, il est courant que la voix soit positionnée au centre dans cette technique de sonorisation. D'une part, il est quasiment impossible pour l'ingénieur du son de faire des suivis pour chaque interprète sur la console de mixage, d'autre part, les systèmes stéréo ne permettent pas une localisation très précise et cohérente avec ce que voit le spectateur. Le standard est de conserver la voix au centre.

***Sweet spot* défavorable à la voix au centre**

Ce phénomène de localisation sur l'enceinte la plus proche est très marqué lorsque les deux enceintes reproduisent exactement le même signal audio (donc lorsqu'un son est placé au centre de la stéréo). Les deux enceintes sont alors deux sources cohérentes, et il s'agit du cas de figure que nous venons d'exposer dans le paragraphe précédent, causant la localisation du son sur l'enceinte la plus proche de nous. Pour des sons non cohérents (c'est-à-dire que les enceintes ne reproduisent pas exactement le même signal), ce qui est le cas de la musique par exemple (le son des instruments étant généralement réparti de gauche à droite), le phénomène est moins marqué, bien que présent

(puisque les sons ne sont jamais tous complètement à gauche ou à droite). L'image stéréo se décale, mais la sensation de localisation sur les enceintes gauches ou droites est moins radicale.

2.2.2 Stéréophonie avec point central (L/C/R)

Afin de remédier à ce problème de localisation de la voix lorsqu'elle est positionnée au centre sur un système L/R, la solution est d'utiliser un point de diffusion central, destiné uniquement à la voix, en complément de la stéréophonie. Dans ce cas de figure, l'on retrouve ainsi le point de diffusion au centre du cadre de scène, et la stéréo de part et d'autre, cette dernière servant alors pour les réverbérations, ainsi que la musique. Ce type de systèmes de sonorisation est le plus fréquent aujourd'hui, quand les productions et les budgets le permettent.

Les salles de spectacles équipées d'un point central de sonorisation sont rares, la majorité étant équipée simplement en stéréo. En France toutes les productions ne font pas l'effort d'équiper leur spectacle de ce point de diffusion supplémentaire, là où la majorité des théâtres londoniens le font.

Filtrage en peigne

Il subsiste néanmoins un problème avec ces systèmes, qui existe également dans les systèmes simplement stéréo : il s'agit du filtrage en peigne causé par les microphones omnidirectionnels lorsque les acteurs se tiennent près l'un de l'autre.

Dans ce cas de figure, la voix de l'interprète sera captée par son propre microphone, mais également par le microphone de celui qui lui fait face. Un très court décalage temporel est induit par la distance entre les deux interprètes. Et lorsque ces deux signaux sont sommés (au centre de la stéréo ou dans le canal central), il se produit alors des phénomènes d'interférences destructives et constructives qui résultent en l'élimination de certaines fréquences et l'augmentation d'autres. Cette modification du signal, appelé filtrage en peigne, altère de manière très audible le signal notamment puisque la coloration de ce filtrage change avec la distance entre les comédiens, ce qui le rend d'autant plus audible lorsqu'ils sont en mouvement.

Le filtrage en peigne se présente comme une série harmonique : un premier pic intervient à une fréquence f_{c1} (interférence constructive), il est ensuite suivi par une annulation (interférence destructive) à $f_{d1} = f_{c1}/2$, suivi d'une série de pics et de creux.

Considérons d_m la distance entre les microphones (en mètres), f_{c1} la fréquence de la première interférence constructive, $f_{d1} = f_{c1}/2$ la première interférence destructive (en Hertz), et $c = 340 \text{ m.s}^{-1}$ la vitesse du son ; l'équation

$f = c/d$ nous permet de déterminer⁴ f_{d1} :

$$f_{d1} = \frac{1}{2} \times \frac{c}{d_m}$$

Pour une distance de un mètre séparant les deux comédiens, la première annulation a lieu à 170 Hz, à 50 cm elle a lieu à 340 Hz et à 10 cm elle intervient à 1700 Hz. Dès lors que les comédiens sont près les uns des autres, le filtrage en peigne affecte le timbre de leur voix de manière audible et gênante.

Solution intuitive : *line-by-line mixing*

La solution la plus évidente à ce problème est de mixer les voix des comédiens, en les suivant *réplique par réplique* (soit *line-by-line*, nom de cette technique de mixage en anglais, aussi appelée *line crunching*⁵). En baissant le niveau du comédien qui ne parle pas, l'on évite ainsi l'effet de filtrage en peigne. Lorsque deux comédiens se font face l'on peut même ne conserver d'un seul des deux microphones, généralement celui du plus grand des deux, puisqu'il surplombe les deux interprètes. Cette pratique requiert de l'ingénieur du son qu'il ait une bonne connaissance du texte et de la dextérité, lors les échanges rapides de répliques.

2.2.3 Stéréophonie avec point central (L/C/R) en système A/B

Une autre solution pour réduire le problème de filtrage en peigne est l'utilisation d'un système de sonorisation A/B⁶. Rien ne change par rapport au système stéréo avec point central si ce n'est que celui-ci est doublé, avec un point central A et un point central B. Dès lors que deux comédiens sont proches l'un de l'autre, chacune des voix est envoyée dans un circuit différent, en programmant la console avec deux bus séparés pour les deux points centraux. Les deux signaux sont donc toujours séparés, il n'y a plus de sommation, et donc plus de filtrage en peigne.

Cette solution est néanmoins lourde et couteuse à mettre en œuvre, puisqu'elle nécessite —en plus de doubler le point central— de doubler également les enceintes de rappel permettant de couvrir la salle de spectacle de manière plus homogène. Les productions faisant ce choix sont donc assez rares.

4. Voir Fischetti, Antonio, *Initiation à l'acoustique*, Belin, Paris, 2003 et Josephs, Andrew et Penny, Ryan, « Back to basics : multiple sound sources », in *The Echo*, n°12 septembre 2016, revue de the ASD (Association of Sound Designers), Royaume-Uni.

5. Fry, Gareth, *Sound Design for the Stage*, The Crowood Press, avril 2019, p. 191.

6. Josephs, Andrew et Penny, Ryan, "Back to basics : multiple sound sources", *The Echo, the ASD magazine (Association of Sound Designers)* septembre 2016, Nr. 12.

Conclusion

Ces pratiques courantes ne procèdent pas à une localisation de la voix, préférant une diffusion homogène apportée par le point central lorsque les conditions le permettent. Il est possible de spatialiser les voix sur ces systèmes, mais de manière très limitée, ce qui a poussé très tôt les concepteurs sonores à trouver des systèmes pour spatialiser le son des voix, parfois en tirant parti de dispositifs de sonorisation initialement prévus pour localiser des effets sonores dans le décor. Le chapitre suivant est dédié à l'étude de ces techniques.

Chapitre 3

Pratiques courantes de spatialisation

Une autre approche de la sonorisation des voix, est ce que les Anglais appellent le *Source Oriented Reinforcement* ou *S.O.R.*, qui vise à concevoir le système de sonorisation de telle sorte qu'il conserve la localisation des voix des acteurs, ou bien qu'il localise les sons au plus près des acteurs. En français l'on parle souvent de sonorisation multipoints ou bien de multidiffusion.

Nous proposons dans ce chapitre de dresser une liste des différents techniques de spatialisation applicables à la voix qui peuvent être utilisées dans le théâtre. Cet état des lieux présente des techniques historiques, qui continuent tout de même d'être utilisées aujourd'hui, car elles peuvent souvent être implémentées avec du matériel moins coûteux, et peuvent suffire à produire un effet suffisant pour des productions n'ayant pas accès à du matériel plus complexe.

Nous ferons la distinction entre les méthodes à panoramique (et leurs dérivés), et les méthodes à matrice.

3.1 Techniques « simples » de spatialisation

3.1.1 Panoramique L/R ou L/C/R et AUX

Il s'agit de l'approche la plus basique de la spatialisation des voix. En utilisant le panoramique intégré dans la console, il est possible de déporter le son sur la gauche ou sur la droite. Ce paramètre est généralement utilisé pour limiter le filtrage en peigne et le masquage entre les différentes voix des chœurs, en répartissant les voix dans l'espace stéréophonique. Il peut être complété par l'envoi du son de la voix vers des enceintes placées dans le décor, via un envoi d'auxiliaire. Plusieurs enceintes dans le décor reçoivent alors une sortie de console sur laquelle est connectée un aux, comme l'on

ferait pour les retours musiciens par exemple. Cette approche sera généralement préférée pour des positionnements statiques.

Pour la fonction de renfort sonore réduite à son plus simple appareil, et jouant sur l'effet Haas, il suffit d'appliquer un délai fixe à l'ensemble du système de sonorisation. Il n'y a pas besoin d'outils complexes pour cela, mais le phénomène d'écho peut rapidement se faire entendre, dès lors que les comédiens ou les spectateurs auront des positions critiques.

Pour monter un peu en précision, il est possible, plutôt que de retarder l'ensemble du système, de retarder la voix des comédiens un à un, en prenant des valeurs de délai différentes en fonction des scènes et de la position du comédien sur le plateau dans ces scènes.

3.1.2 Spatialisation par matrice

Dans un cas de figure plus complexe, si l'on souhaite exploiter plusieurs enceintes, il faut pouvoir associer à chacun des comédiens, un délai différent pour chaque enceinte en fonction de sa position sur scène. Le recours à une matrice de délais et/ou de niveau est alors l'unique solution, il est évident qu'à partir d'un grand nombre de comédiens et d'enceintes, il n'est plus humainement possible d'effectuer tous ces calculs ; une solution informatisée sera alors préférable.

Cette approche fonctionne en définissant plusieurs zones sur le plateau. Pour chaque zone seront définis des réglages de niveau et/ou de délai pour chaque haut parleur. L'on paramétrera dans la console autant d'envoi d'aux (ou « mix ») que de zones sur scène. Les aux nourriront les entrées d'une matrice (interne à la console de mixage ou externe en fonction des besoins et du matériel à disposition). Les sorties de cette matrice sont les canaux audio destinés à chaque haut-parleur. En fonction des réglages définis plus tôt en quadrillant l'espace, l'on règlera les niveau et/ou délai de chaque point de croisement de la matrice.

Ainsi, pour chaque tranche de la console, il suffit d'envoyer le signal dans l'aux correspondant à la position de l'interprète sur scène. Les positions peuvent être préenregistrées dans la console via des *scènes*. Certaines consoles permettent d'appliquer un fondu aux paramètres choisis, d'une scène à l'autre. Cette option permettra d'avoir des changements de position plus doux. Enfin, si la console est pilotable par un protocole quelconque, il est possible d'asservir les niveaux d'envoi d'aux sur le positionnement de l'interprète, quelle que soit la méthode utilisée pour en acquérir la position (voir le chapitre 4).

3.2 Des matrices au mixage objet

3.2.1 Les premières solutions informatisées

Au début des années 2000, John A. Leonard documente dans *Theater Sound*¹ l'existence de matrices contrôlées par ordinateur, telles que la « Richmond Sound Design AudioBox » ou encore la « Level Control Systems SuperNova », permettant de rappeler différents réglages de niveaux, de délai ou de points de connexion dans la matrice, et de faire des fondus entre ces valeurs. Ces matrices, contrôlées par des ordinateurs sous BeOS, permettaient également la lecture de fichiers audio, et le système SuperNova permettait de déplacer des sons facilement dans l'espace, sans avoir à régler soi-même les changements de niveaux pour chaque enceinte, dans la matrice, lors du déplacement.

La solution de Outboard UK, « TiMax », que mentionne également l'auteur, est la première solution proposant d'automatiser le fonctionnement par matrice décrit en 3.1.2. Les calculs de valeurs de délai et de niveau sont automatisés, mais peuvent être ajustés par l'opérateur. Plutôt qu'opter pour une mise à jour en continu des valeurs de niveau et de délai, TiMax reste fidèle à la méthode matricielle propose de calculer ces valeurs pour des positions clefs (appelées *image definitions*), choisies par l'opérateur, puis effectue des fondus entre ces valeurs.

Dans sa seconde itération, la matrice s'est vu adjoindre un système de suivi de position des acteurs (*tracking system*) permettant d'automatiser les changements de position des acteurs.

Ce produit pose un changement de paradigme. Auparavant, l'on devait spatialiser les voix en modifiant chacun des paramètres de spatialisation sonore soi-même (soit le niveau et le délai). Avec ce produit, l'on dispose maintenant d'une interface graphique où il suffit de positionner la source sonore au même emplacement que l'acteur sur la carte affichée à l'écran et le son est spatialisé. Cet outil ouvre la voie à une nouvelle technique de mixage : le mixage objet.

3.2.2 Vers les processeurs et le mixage objet

TiMax le précurseur

Bien que fréquemment utilisé dans le monde anglo-saxon (avec notamment des spectacles comme *Tanz der Vampire* ou *Aladdin*), *TiMax* semble être méconnu en France. Nos recherches ne nous ont permis de trouver la trace que de deux utilisations en France : pour la version française de *Tanz der Vam-*

1. Leonard, John A., *Theater Sound*, A C Black, 2001.

pire au Théâtre Mogador en 2015² (il s'agit d'un concepteur sonore anglo-saxon, venant pour la re-création du spectacle en France), et pour *Peer Gynt* au Grand Palais, mis en scène par Eric Ruf et sonorisé par Jean-Luc Ristord en 2012³. Pas de surprise à n'avoir trouvé que si peu de traces d'utilisations ici, car d'après ce dernier article, Audiolead aurait été le premier et le seul distributeur français à proposer *TiMax* dans son catalogue, et ce, seulement à partir de 2012. Depuis le produit initial de la fin des années 1990, *TiMax* a produit des itérations plus modernes de son produit, qui s'intègre désormais dans les réseaux audionumériques.

Le tournant des années 2000 a vu une évolution majeure des outils permettant de mettre en œuvre des sonorisations avec spatialisation sonore : les solutions industrielles se sont multipliées, de nombreux constructeurs proposant désormais leur propre processeur de spatialisation sonore. Ceux-ci ont néanmoins été conçus dans une logique relativement différente que des systèmes comme *TiMax*, puisqu'ils se basent presque tous sur des algorithmes et non sur des matrices.

Barco et Sonic Emotion : synthèse de front d'onde

En 2004, la société Isono est la première à proposer un processeur de synthèse de front d'onde (WFS). Depuis 2014, Isono est la propriété de Barco⁴. Peu de temps après, c'est la société Sonic Emotion qui commercialise à son tour une industrielle mettant en œuvre la WFS avec le processeur Wave I, dans les années 2010⁵, puis une seconde itération nommée Wave II. Développé en coopération avec l'Ircam par des anciens de chez Studer⁶, ce processeur équipe plusieurs salles en France, dont le Théâtre National de Chaillot, la Maison de la Radio, l'Institut du Monde Arabe, la Marbrerie, *etc.* et à l'international comme à l'Aura Club de Zurich, au Stuttgart Staatstheater Schauspielhaus et en Chine au National Indoor Stadium, *etc.*; des prestataires s'équipent également comme On/Off (Matt Pokora, *Dirty Dancing* le musical), Urbs Audio (Jazz à la Villette); et enfin des institutions universitaires comme l'UBO, Centrale Supélec, l'Université de Lille, VirginiaTech aux USA ou le CIRMMT au Canada... Néanmoins au cours des dernières années,

2. Clark, Gisèle. « d&b, TiMax et plus encore pour le Bal des Vampires ». Réalisation, mai 2015.

3. Clark, Gisèle. « Peer Gynt au Grand Palais : TiMax le magicien du son ». Réalisation, juillet 2012.

4. barco, Communiqué de presse, *Barco gains the team and expertise from IOSONO GmbH*, (URL: <https://www.barco.com/en/News/Press-releases/Barco-gains-the-team-and-expertise-from-IOSONO-GmbH.aspx>) – visité le 2019-05-16.

5. Wanègue, Jean-José. « Sonic Emotion et la WFS : De la simulation en laboratoire au road-show en son 3D ». Réalisation, octobre 2010.

6. Studer faisait partie du projet Carusso, un projet européen de recherche autour de la WFS.

l'entreprise Sonic Emotion semble avoir arrêté le développement du produit, seule la distribution en étant maintenue (assurée par Euphonia).

Les mastodontes à l'assaut de la spatialisation sonore

Au cours des dix dernières années (rappelons que nous écrivons ces lignes en 2019), les constructeurs d'enceintes ont lancé leurs propres processeurs. C'est notamment au cours de ces trois dernières années que l'on a vu le nombre de processeurs de spatialisation sonore augmenter de manière significative. De plus, ces constructeurs sont des acteurs majeurs de l'univers de la sonorisation, l'on compte rien de moins que Meyer Sound, L-Acoustics, d&b audiotechnik et enfin Amadeus.

Meyer Sound lance en 2009 un processeur appelé « D-Mitri », une plateforme audio multi fonction, qui permet entre autres de spatialiser les sons. Son approche est assez large, de l'optimisation acoustique, au théâtre, en passant par les parcs à thème. Néanmoins, à l'inverse de produits comme ceux proposés par L-Acoustics ou d&b auditechnik, le constructeur ne met pas en avant l'utilisation du produit pour sonoriser des concerts de musiques actuelles.

Le premier constructeur d'enceintes à se lancer dans la conception d'un processeur de spatialisation sonore particulièrement conçu pour le concert de musiques actuelles, est la société française L-Acoustics, qui lance au cours de l'année 2016 « L-ISA ». Restant dans son cœur de cible principal, ce produit est actuellement surtout exploité sur ces concerts, bien qu'il soit présent également sur d'autres secteurs. Depuis son lancement, le système a été utilisé sur plusieurs concerts importants, permettant ainsi de faire découvrir ce type de sonorisation au grand public. On compte entre autres, depuis la sortie du produit, BBC Proms au Royal Albert Hall (ce festival mythique de musique classique), des concerts de alt-J, Angus et Julia Stone, Lorde, Ennio Morricone, Renaud ou bien dans l'Antarctic Dome a Coachella.

Peu de temps après, un nouveau processeur arrive également sur le marché. Lui n'est pas proposé par un constructeur d'enceintes, et semble moins se destiner à la sonorisation de concerts mais plus au spectacle vivant, puisqu'il met en avant des usages proches de ceux de TiMax. C'est Astro Spatial Audio, marque de la société néerlandaise Vanmunster BV fondée en 2012, qui a dévoilé début 2017⁷ son processeur « SARA II Premium Rendering Engine ». Ce processeur de spatialisation sonore se base sur la technologie « SpatialSound Wave », un programme développé par le Fraunhofer Institute for Digital Media Technology, organisme de recherche allemand fonctionnant majoritaire-

7. Astro-Spatial-Audio, *Sara II object rendering engine*, site web (URL: <http://www.ravepubs.com/astro-spatial-audio-debut-sara-ii-object-rendering-engine-ise-2017/>) – visité le 2019-11-26.

ment à partir de fonds privés et de partenariats industriels. Ce « SpatialSound Wave » tirerait son fonctionnement des principes de la WFS, bien que cela ne soit jamais clairement évoqué sur leur site web⁸ (seul un lien vers un document PDF s'intitulant « more information about Wave Field Synthesis » laisse supposer cela). Utilisé sur un petit nombre spectacles et installations, le produit s'est surtout illustré en 2017 lorsque l'ingénieur du son Kai Harada reçoit un Tony Awards pour la meilleure conception sonore, avec la comédie musicale de Broadway « The Band's Visit », qui exploitait leur processeur pour sonoriser l'orchestre et les voix des artistes.

Le troisième constructeur d'enceintes à tenter l'aventure de la spatialisation sonore après Meyer Sound et L-Acoustics est la marque allemande d&b audiotechnik. Ils ont lancé début 2018 le système « Soundscape », basé sur le processeur DS100, une matrice polyvalente à laquelle peuvent être ajoutées les options *EnScene* (spatialisation sonore) et *EnSpace* (réverbération multicanal à base de réponses impulsionnelles). En ce sens, ce système se rapproche de la logique de Meyer Sound : le processeur peut servir à d'autres usages que la spatialisation sonore. Les premières utilisations du produit marquent bien les différentes utilisations qui ont amené la marque à créer ce produit : une tournée avec Kraftwerk, des concerts de musique classique en plein air (pratique très fréquente en Allemagne, ce qui a inspiré la marque à proposer des réverbères à convolution issues de quelques salles allemandes), le son d'une pièce de théâtre de Broadway : *Lobby Hero* de Kenneth Lonergan, d'une comédie musicale londonienne, *Starlight Express*, de Andrew Lloyd Webber, etc.

Peu de temps après, c'est le constructeur français Amadeus qui sort le processeur « Holophonix ». Celui-ci se distingue des autres processeurs puisqu'il est le seul à proposer un vaste nombre d'algorithmes, et le seul à nommer noir sur blanc les algorithmes qu'il utilise. On compte entre autres l'Ambisonie, la WFS et le VBAP et de nombreux autres algorithmes tels que VBIP, LBAP, KNN. Ce processeur a été conçu en proche collaboration avec le laboratoire STMS (Sciences et Technologies de la Musique et du Son) de l'Ircam et permet ainsi de proposer une solution commerciale répondant aux impératifs du spectacle vivant intégrant ces algorithmes de l'Ircam. Les premiers clients de Holophonix sont le théâtre de la Comédie Française (Vieux-Colombier), les Champs Libres à Rennes, le théâtre National de Chaillot, ainsi que le Arts & Creative Technology Center en Corée du Sud, d'autres étant à venir.

8. IDMT, *Spatialsound wave*, site web (URL: <https://www.idmt.fraunhofer.de/en/institute/projects-products/spatialsound-wave.html#tabpanel-3>) – visité le 2019-11-26.

Deuxième partie

Spatialisation de la voix avec les « nouveaux » outils

Chapitre 4

Acquisition de la position et du mouvement

L'un des paramètres majeurs dont nous avons besoin pour spatialiser le son à l'emplacement d'un interprète est sa position dans l'espace. Lorsque l'on travaille avec des sources statiques (par exemple des musiciens à des positions fixes sur scène), il est aisé de positionner le son à l'aide d'une interface, aussi basique soit-elle : l'interaction entre ce que nous voyons et ce que nous entendons nous suffit à régler correctement la localisation du son en cohérence avec la position de l'interprète. Les processeurs de spatialisation sonore actuellement disponibles sur le marché présentent tous une interface permettant de positionner les sons, pour des applications relativement simples comme les concerts, cela semble largement suffisant.

En ce qui concerne le type de spectacle que nous étudions dans ce mémoire, la tâche est beaucoup plus complexe, puisqu'il faut gérer simultanément les déplacements de tous les interprètes, souvent nombreux, sur toute la durée du spectacle.

Une des solutions utilisées sur certains spectacles est de préenregistrer les positions ou les mouvements des comédiens via des automatisations. Mais comme nous allons voir, cette solution, bien que viable dans certaines applications, présente l'inconvénient de ne pas correspondre à la nature même du spectacle vivant : la variabilité de la reproduction du spectacle d'une représentation à l'autre. Il suffit que la représentation dévie de ce qui a été préenregistré pour que l'illusion s'effondre.

Pour s'assurer que les sons soient localisés au même endroit que les interprètes, il existe une troisième option : utiliser un système de localisation permettant de connaître la position exacte de chaque interprète à tout instant. L'avantage d'un tel système est que quoi qu'il arrive sur scène, la localisation du son sera synchrone aux mouvements de l'interprète, permettant même d'explorer des mises en scène plus libres avec de l'improvisation.

4.1 Suivi de la position par la méthode cognitive

Le potentiomètre

Le potentiomètre de panoramique est l'outil de localisation d'un son présent de base sur toutes les consoles. Il permet de diriger le positionnement d'un son sur l'axe gauche-droite dans un système de sonorisation L/R ou L/C/R.

Aujourd'hui sur les consoles numériques, qui sont très majoritairement utilisées en sonorisation, l'on retrouve généralement des jeux d'encodeurs rotatifs assignables (c'est-à-dire que le paramètre qu'ils contrôlent change en fonction de l'option sélectionnée). Avec certaines consoles, il est même possible de contrôler avec ces encodeurs des paramètres externes à la console, via des protocoles comme le MIDI ou l'OSC (voir 4.4). Il est donc envisageable d'utiliser ces potentiomètres pour venir piloter la position des objets dans un système de spatialisation sonore, chaque encodeur rotatif pilotant une coordonnée de l'objet (x, y, z ou a, e, d). Certains processeurs proposent même une intégration dans des consoles, accompagnant les encodeurs d'une petite interface permettant d'avoir quelques repères, comme dans la figure 4.1 page suivante.

Cette approche pour le positionnement d'objets sonores présente des avantages dans le mixage et le positionnement de sources fixes, car tous les contrôles se trouvent au même emplacement (sur la console de mixage), ce qui évite à l'opérateur d'accumuler les surfaces de contrôle pour élaborer son mixage. En ce qui concerne le suivi de sources en mouvement, cette méthode se révèle limitée, puisqu'elle exige la préhension de chaque encodeur entre le pouce et l'index, ce qui monopolise toute la main de l'opérateur. Ainsi dans l'absolu, un opérateur ne pourra effectuer le suivi que de deux interprètes simultanément (et encore, à condition de ne piloter que leur position en azimuth et pas en profondeur).

Une alternative au potentiomètre est le *joystick*, un encodeur fonctionnant en deux dimensions, permettant ainsi de déplacer une source à une main, libérant l'autre pour manipuler éventuellement une seconde source.

Ceci reste donc une solution qui n'est pas viable pour le suivi des déplacements d'une troupe complète, mais qui pourrait néanmoins trouver sa place dans des applications où elle permettrait un suivi de quelques sources, couplée avec par exemple un préenregistrement des autres positions.



FIGURE 4.1 – Capture de l'écran d'une console DiGiCo montrant l'intégration d'un système permettant le contrôle du processeur L-ISA (L-Acoustics). Sur la console, les potentiomètres visibles à l'écran sont alignés avec des encodeurs rotatifs de la surface de contrôle. (Source : Service presse de DiGiCo, « L-ISA Source Control Available On DiGiCo SD Consoles », 11 avril 2018. Consulté le 8 avril 2019. https://www.digico.biz/docs/latest_news/EVkJFuVuApEDLWmHmHn.shtml).

Interface graphique : dispositifs de pointage et écrans tactiles

Les logiciels de spatialisation sonore ainsi que les processeurs présentent tous une interface graphique représentant la zone de positionnement des sources. Cette visualisation peut être plus ou moins élaborée, allant d'une simple représentation schématique de l'espace de représentation à des rendus tridimensionnels plus complets.

La manipulation des sources se fait alors nécessairement par le dispositif de pointage de l'ordinateur qu'il s'agisse d'un souris, d'un pavé tactile (*trackpad*) ou d'un boule de commande (*trackball*). Les systèmes d'exploitation n'autorisant l'affichage que d'un seul pointeur à l'écran, il n'est possible de déplacer qu'un seul objet sonore à la fois.

Cette solution est avantageuse dans la mesure où elle consiste en une interaction directe avec le plan de la salle, permettant ainsi d'avoir accès à des repères visuels relativement précis.

L'utilisation d'écrans tactiles présente les mêmes avantages que les inter-

faces graphiques, sans toutefois être limité comme le sont les dispositifs de pointage des ordinateurs. Les tablettes multitactiles¹ donnent la possibilité d'interagir avec la surface d'affichage via de multiples points de contact, et ainsi de manipuler la position de plusieurs objets sonores simultanément. Cet outil a par exemple été utilisé² en 2010 par Markus Noisternig de l'Ircam au festival d'Avignon pour la sonorisation en WFS de la pièce « La Tragédie du Roi Richard II » mis en scène par Jean-Baptiste Sastre. Il semble évident là encore qu'il n'est pas possible pour un seul opérateur de suivre les déplacements d'une troupe entière, rendant ainsi nécessaire de combiner cette technique avec des mises en mémoire de positions. C'est ainsi³ que Markus Noisternig a opéré en 2010, en enregistrant les positions des comédiens lors de leurs entrées en scène.

Les tablettes tactiles présentent en outre l'avantage d'être aujourd'hui très répandues dans l'univers de la sonorisation, notamment pour contrôle à distance des consoles de mixage. Ainsi, intégrer un tel outil au matériel nécessaire pour une production peut s'avérer tout à fait anodin, voire même standard.

Le seul frein que nous pouvons voir à l'utilisation de cet outil est qu'il semblerait qu'aucun constructeur de processeurs ne propose d'application de contrôle pour tablette tactile, là où les constructeurs de consoles le proposent quasiment tous par exemple. Il reste possible d'utiliser un écran tactile connecté à l'ordinateur, ou bien de faire une recopie d'écrans de l'ordinateur sur une tablette, néanmoins il n'est pas dit que la fonction multitactile de l'écran soit conservée, ce qui revient à avoir la même limitation qu'une souris. Pour avoir une telle interface, il faut pour le moment que le concepteur son la prototype lui-même à l'aide d'applications ou logiciels tels que *Lemur* de Liine, *max* de Cycling'74, *Hollyhock Usine* de Senso Music.

Les limites de l'acquisition cognitive

Comme nous venons de le voir, chacune de ces méthodes d'acquisition de la position des interprètes monopolise les mains d'un opérateur et ne permet pas de suivre un grand nombre de personnes sur un plateau. Ceci pose un problème vis-à-vis des habitudes du métier : s'il n'est pas rare que les concepteurs son travaillent avec une équipe lors d'une création, la régie sonore d'un spectacle se fait généralement par un seul opérateur. Or cet opérateur est déjà occupé par des tâches telles que le mixage des voix, notamment en comédie musicale, ou bien par la régie de la création sonore.

1. Multitactile est l'appellation française pour la technologie *multi-touch*.

2. Wanègue, Jean-José, "Révolution de palais dans la Cour d'Honneur : La WFS s'invite à Avignon", *Réalisation* octobre 2010, Nr. 62.

3. *Ibid.*

Ces solutions peuvent être combinées avec l'enregistrement d'automations : c'est-à-dire des mouvements préenregistrés puis joués lors du spectacle. Elles permettent alors d'effectuer des mouvements fins sans avoir à réaliser soi-même l'intégralité des déplacements.

4.2 Suivi de la position par automation et time-code

Exception faite de l'improvisation, les mises en scène figent généralement un certain nombre de positions dans l'espace à des instants précis du spectacle, ainsi que des déplacements. Partant de ce constat, il suffirait d'enregistrer ces positionnements et mouvements, et de les déclencher au moment opportun. Il existe plusieurs méthodes pour arriver à ce résultat, qui peuvent s'avérer plus ou moins avantageuses selon le type de production considéré.

Ces différentes méthodes et techniques peuvent se distinguer selon deux critères à considérer lors du choix de l'outil : la linéarité, et la séquentialité, qui décrivent le rapport au temps de l'outil utilisé.

4.2.1 Caractérisation des outils d'automation

Linéarité

Nous caractérisons de *linéaire* un outil ou une méthode faisant appel à une unique ligne temporelle. Les différents événements sont enregistrés sur cette ligne temporelle dont le déroulement ne peut être altéré au cours du spectacle. Ces outils possèdent en outre un unique point de départ auquel se réfèrent tous les événements.

Un outil *non linéaire* est un outil dans lequel les événements ne se réfèrent pas à une ligne temporelle globale, pouvant par conséquent être rappelés à n'importe quel instant. Dans l'absolu, un outil non linéaire ne peut pas être porteur, par nature, d'une temporalité. Il s'agirait par exemple de déclencher un événement ponctuel, par exemple d'actualiser la position d'un objet.

Les outils *multilinéaires* supportent quant à eux des événements qui évoluent avec le temps de la même manière que dans les systèmes linéaires, mais en ayant la possibilité d'avoir autant de lignes temporelles que d'événements souhaités. Chaque élément nécessitant une évolution temporelle (comme un mouvement, l'évolution de n'importe quel paramètre ou encore la lecture audio), possède sa propre ligne temporelle, ce qui permet de le déclencher à l'instant souhaité, indépendamment des autres.

Séquentialité

La séquentialité se réfère à l'ordonnement des éléments dans le temps. Un outil au fonctionnement *séquentiel* contraindra le déclenchement des événements à un ordre préalablement établi. Par nature, les outils linéaires sont séquentiels. Cette distinction est donc intéressante pour différencier des outils multilinéaires ou non linéaires.

La logique *non séquentielle* ne nécessitant pas d'ordonnement des événements, chaque événement peut être déclenché indépendamment.

Néanmoins dans les équipements et outils à disposition dans le monde de l'audio et du *show control* aucun n'est réellement complètement séquentiel. Il est généralement possible de naviguer librement entre les séquences, ce qui est nécessaire pour plusieurs cas de figure : lors des répétitions, si l'on ne travaille pas l'œuvre dans l'ordre ; lors d'une interruption de représentation, en cas d'avarie ou de problème, où il peut falloir reprendre quelques instants avant le point d'arrêt ; ou encore en cas de modification de la conduite, lorsqu'un élément du spectacle est supprimé, son déclenchement sera alors ignoré, passant à la séquence suivante. Ce qui nous amène alors à qualifier ces outils de « séquentiels » c'est leur ergonomie, qui est pensée principalement pour déclencher des événements dans l'ordre prévu. C'est le cas par exemple des mémoires de console de mixage pour rappeler des scènes⁴ enregistrées : la surface de contrôle de la console présente généralement des boutons « suivante » ou « précédente » pour naviguer d'une scène à l'autre lors du spectacle, ce qui contraint à un défilement séquentiel. Il reste toujours possible de rappeler une *autre* scène en mémoire que la précédente ou la suivante, généralement via un menu spécifique (logique non séquentielle).

4.2.2 Outils à disposition

Automation

En informatique audio, l'automation consiste à programmer la variation d'un paramètre dans le temps. Il est possible de la modifier à l'aide d'une courbe représentant la variation du paramètre de 0 à 1 dans le temps. Une automation peut être définie dans un système linéaire (elle sera généralement définie sur l'ensemble de la ligne temporelle), ou bien dans un système multilinéaire, auquel cas l'automation s'étendra uniquement sur la durée de l'évènement.

La figure 4.2 page ci-contre est un exemple d'automation dans QLab, définie sur un évènement. L'automation prendra le contrôle du paramètre commandé pendant la durée de l'évènement, et libèrera la commande du para-

4. Une scène permet d'enregistrer un ensemble de réglages et de paramètres de la console.

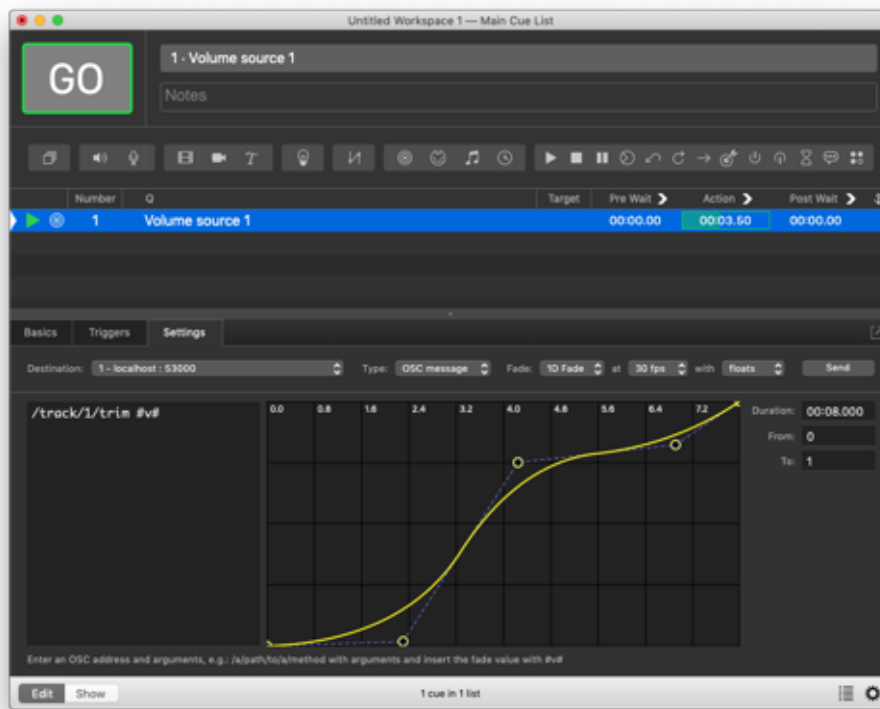


FIGURE 4.2 – Automation dans QLab

mètre une fois l'évènement terminé.

Le time-code

C'est un outil de synchronisation conçu à l'origine pour la synchronisation du son et de l'image dans la production audiovisuelle, mais qui a trouvé sa place dans le spectacle vivant. Il s'agit d'un signal permettant de maintenir la synchronisation de différents équipements sur une même ligne temporelle. Bien que ce terme désigne normalement un mécanisme de synchronisation entre une machine maître et des machines asservies, la dénomination de « time-code » se trouve généralement employée dans le métier même lorsqu'il n'y a pas de synchronisation externe, et que tout est produit dans le même logiciel.

Utiliser une synchronisation au time-code pour enregistrer les déplacements des interprètes sur scène suppose que les évènements soient reproduits à la seconde près à chaque représentation, ce qui ne correspond qu'à certains spectacles très spécifiques. D'autre part, l'on peut considérer qu'il est très fastidieux de repérer et enregistrer les déplacements de chacun des interprètes sur le plateau. Cela demande d'une part que les déplacements et positions soient fixés, ce qui peut intervenir tard dans une création, et d'autre

part d'avoir les interprètes à disposition sur scène ou bien une captation pour enregistrer les déplacements. Cette solution peut néanmoins s'avérer intéressante pour des spectacles dont on sait qu'il n'y aura pas de déviation soir après soir.

De manière générale, l'on préférera utiliser dans le spectacle vivant des solutions permettant de déclencher les événements au fur et à mesure qu'ils arrivent, sans le figer sur une ligne temporelle.

Il est également possible d'utiliser le time code pour d'autres applications que de commander l'ensemble du spectacle. Ainsi, un logiciel de commande non linéaire peut envoyer des commandes de time-code à un autre appareil, pour démarrer la lecture, ou se synchroniser à divers emplacements de la ligne temporelle de l'appareil esclave. On peut ainsi enregistrer des séquences sur time-code, tout en conservant la flexibilité qu'offre un logiciel de commande non linéaire.

Logiciels de contrôle non linéaires

Certains logiciels comme *QLab* et *Ableton Live* (tous deux étant les plus fréquemment utilisés dans le spectacle vivant) permettent de contrôler des données. *QLab* est de loin celui qui intègre le plus de protocoles de commande, en outre il intègre l'OSC. *Ableton Live* n'intègre que le MIDI comme protocole, qui comme nous allons le voir ne nous intéresse pas. Il est possible d'envoyer des données en passant par des plug-ins ou bien via *max4live*, l'intégration du logiciel *Max* de Cycling'74 directement dans *Ableton Live*.

Ces deux logiciels permettent d'avoir une approche séquentielle ou non.

Logiciels de commande de trajectoire

Pour enregistrer une trajectoire, il est évidemment possible d'enregistrer une automation de chaque paramètre (x, y, z) en déplaçant la source « à la main ». Néanmoins cette méthode présente un énorme inconvénient : les automatisations pour les trois paramètres seront enregistrées individuellement, rendant difficiles les retouches.

On préférera donc utiliser un programme permettant de dessiner des trajectoires directement sur un espace 2D (*QLab*) voir 3D (*Iannix*) [fig. 4.3 et 4.4].

Scènes, *snapshots* ou *cues*

Admettons que l'on ait positionné les source dans l'espace, et que l'on souhaite sauvegarder ces positions. Plusieurs solutions s'offrent à nous, mais elles dépendent des fonctionnalités du processeur. Certains permettent par

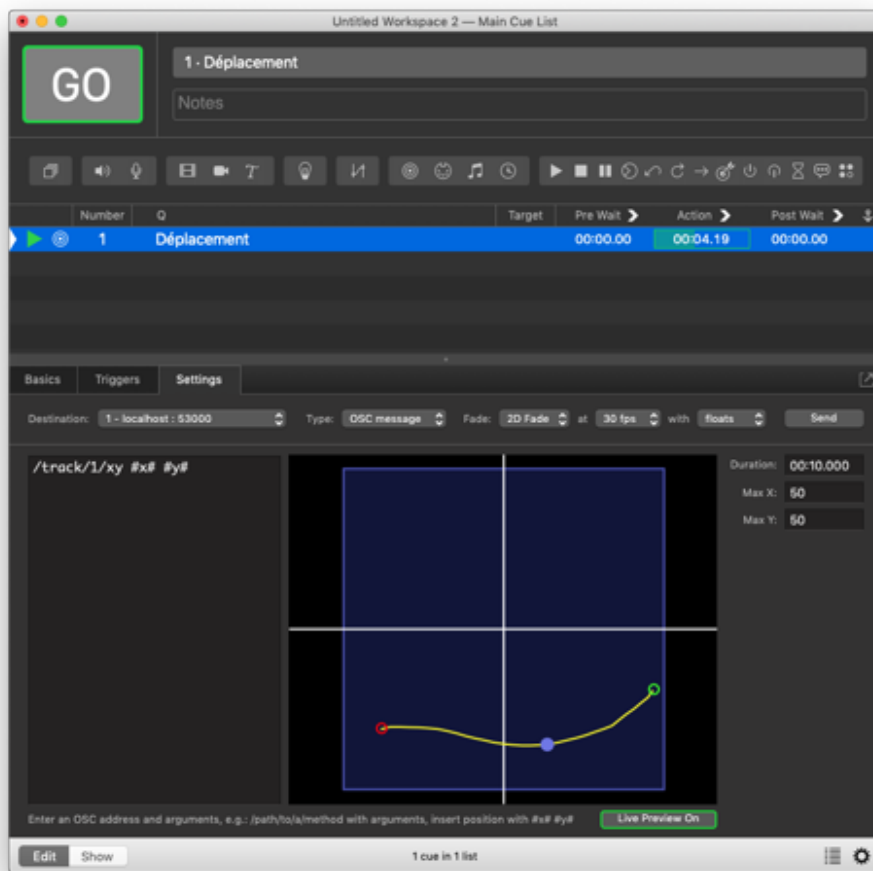


FIGURE 4.3 – Automatisation de déplacement dans QLab

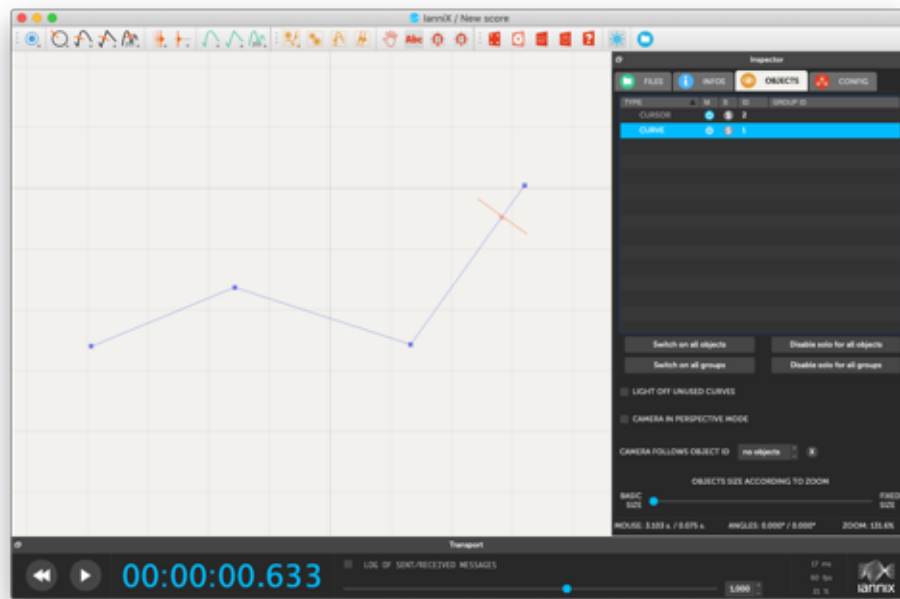


FIGURE 4.4 – Automation de déplacement dans Iannix

exemple de sauvegarder cette scène dans une liste d'évènements (CUE list), ou sous forme d'un *snapshot* qui doit pouvoir être rappelé instantanément.

Si le processeur est intégré à une console, il devrait être possible d'enregistrer les positions dans la console comme tout autre paramètre de la console. Dans le cas contraire il faudra requérir à un logiciel de commande externalisé permettant d'enregistrer des commandes.

4.2.3 Conclusion sur l'utilisation d'automations

Si l'on repose sur l'enregistrement de positions, il faut être attentif à implémenter un système qui permette d'enregistrer simplement les positions de tous les objets dans différentes mémoires qui pourront être rappelées instantanément ou avec des *crossfade*, et qui soient placées dans une liste de commande ou *cue liste* respectant l'ordre du spectacle (logique séquentielle, mais non linéaire).

Certains processeurs intègrent toutes ces fonctions, c'est notamment le cas de TiMax, qui permet d'enregistrer des scènes et de définir des déplacements.

4.3 Suivi automatique de la position

Alors que les premières matrices ou processeurs de spatialisation sonore voient le jour au milieu des années 1990, les systèmes de suivi d'interprètes sur scène commencent à apparaître au milieu des années 2000. L'on compte

par exemple le constructeur TTA, ou Outboard UK (processeur TiMax). La localisation des interprètes permet alors d'automatiser la position des sons dans l'espace. Avec l'augmentation de la précision de tels systèmes, atteignant une précision de l'ordre de 15 cm (TiMax) voir 2 cm (TTA Stage Tracker).

Cette solution est très avantageuse, car elle libère les mains de l'opérateur, automatisant la localisation des sons, qui est asservie sur le système de suivi. Si l'on voulait suivre les déplacements des interprètes avec la même précision et la même simultanéité que permettent les systèmes de suivi, il faudrait presque la moitié du nombre d'interprètes en opérateur de suivi, ce qui est très certainement questionnable, même inimaginable d'un point de vue économique. Utiliser un système de suivi est donc la seule solution permettant d'atteindre un tel niveau de précision et de flexibilité.

4.3.1 Caractéristiques des systèmes de localisation

Dans notre environnement les systèmes de localisation sont monnaie courante. À l'intérieur notre poche ou notre voiture avec un GPS, dans les magasins avec des systèmes de localisation des clients, dans l'industrie avec des systèmes de suivi permettant de localiser des items dans une manufacture ou sur une chaîne d'assemblage, ou bien dans la défense et la navigation civile à l'aide de radars, enfin au cinéma d'animation avec le *motion tracking* ou dans les sports télévisés pour asservir les mouvements de caméra sur la position du ballon ou d'une balle de golf par exemple.

Les systèmes de localisation présentent différentes caractéristiques en fonction de leur application, qui peuvent les rendre plus ou moins adaptés à leur application au spectacle vivant. Nous passons ici en revue ces différentes caractéristiques avant d'étudier le principe de fonctionnement des principaux systèmes que l'on peut trouver dans un théâtre.

Localisation ou suivi de déplacement

Deux approches pour connaître la position d'un élément peuvent être mises en œuvre.

La première approche correspond aux techniques de localisation, qui permettent de connaître la position d'un objet dans une zone définie, sans connaître la position initiale de l'objet. Ces techniques nécessitent l'utilisation de capteurs positionnés autour de la zone de localisation, la position de ces capteurs devant être connue.

L'autre approche possible est le suivi de déplacements, en général effectué à l'aide d'une technologie telle que les centrales inertielles. Avec cette méthode, il faut connaître la position initiale des personnes, le système ne

mesurant que la direction et la valeur du déplacement. La position est déduite de ces données. L'inconvénient de cette méthode est qu'elle est sujette à un fort risque de dérive de la mesure, nécessitant de recalibrer régulièrement la position suivie sur la position réelle.

Un système tel que *BlackTrax* exploite une combinaison des deux techniques, avec un système de localisation par « Motion tracking » (analyse visuelle/vidéo), ainsi qu'une centrale inertielle. Si la cible vient à sortir du champ de vision du système, la centrale inertielle prend le relais, transmettant les déplacements par radio au système de localisation.

Localisation de surveillance ou de navigation

Les méthodes de localisation (telles que la triangulation, la multilatération, l'analyse vidéo) peuvent fonctionner dans deux sens d'utilisation. La navigation consiste à calculer, localement, la position d'un élément, à partir de repères fixes dans l'environnement. C'est le cas des GPS par exemple, la position est calculée au sein de l'appareil de guidage à partir d'échanges de signaux avec des satellites à la position déterminée et connue.

La surveillance consiste à relever la position d'une cible à partir d'un ensemble de bases de mesure fixes dont la position est préalablement connue.

La surveillance semble donc la plus adaptée à notre application. Mettre en place un système de navigation supposerait de calculer la position de chaque interprète dans un appareil dont il serait équipé, puis de la transmettre à un système centralisé en régie. Notre application devant se faire en temps réel, nous devons être très attentifs à la latence. Il est donc inutile d'ajouter un transport de l'information, la surveillance nous permettant d'accéder à la position des interprètes directement en régie.

Surveillance coopérative ou non-coopérative

Dans les systèmes coopératifs, l'élément suivi émet un signal, ou est porteur d'un repère servant de base à la localisation de l'élément, mais pouvant également permettre au système de localisation de l'identifier de manière unique.

Les systèmes non coopératifs permettent quand à eux de localiser un élément sans pour autant requérir un équipement du côté de la cible que l'on cherche à localiser.

Dans notre application, nous devons associer chaque point à un canal audio. Il faut donc identifier qui est la source. En surveillance non coopérative il est compliqué d'identifier les sources de manière complètement fiable.

Lorsque l'on utilise un système de suivi pour le spectacle vivant, il n'est pas possible d'envisager d'utiliser un système présentant des risques d'er-

reur pouvant entraîner des inversions de rôles ou des non-détections. Si de telles erreurs peuvent être corrigées dans d'autres corps de métier, au spectacle la représentation doit suivre son cours. Nous cherchons donc par tous les moyens de minimiser les risques de défaillance. Le choix se portera donc sur des systèmes coopératifs, ce que proposent d'ailleurs exclusivement les systèmes disponibles sur le marché. Les artistes devront donc porter un émetteur, appelé « tag », que l'on pourrait traduire par balise.

4.3.2 Les méthodes de localisation

Deux approches se distinguent dans la localisation de cible coopérative. D'une part, nous avons la localisation par mesure de certaines caractéristiques d'un signal émis par la cible. Il peut s'agir selon la méthode choisie de mesurer par exemple le temps de propagation, l'atténuation du signal, son angle d'arrivée, *etc.* D'autre part, nous avons les techniques qui reposent sur l'analyse vidéo, souvent appelées *motion tracking*.

Si la localisation optique (via l'analyse vidéo) permet une plus grande précision, elle présente entre autres l'inconvénient de requérir une ligne de vue entre la cible et le système. Cette technique ne sera pas forcément la plus adaptée à un spectacle comprenant de nombreux décors par exemple.

L'approche la plus fréquente est basée sur l'émission d'un signal radio-fréquence ultra large bande (mais pouvant fonctionner également avec des ultra-sons, une émission infrarouge, *etc.*). Cette technique ne nécessite pas de ligne de vue directe entre l'émetteur et le récepteur, mais sera sensible aux réflexions multiples; cependant un fenêtrage adéquat temporel du signal reçu permet de surmonter ce problème.

Ces différents concepts représentent ce que l'on appellera les *techniques* de localisation, et reposent sur des *méthodes*, présentées ci-dessous, qui permettent de déduire la localisation à partir des données reçues.

Les *méthodes* présentées ici mettent en œuvre différentes équations dont la résolution permet de déterminer la localisation des objets. Ces équations ne sont pas utiles à l'ingénieur du son : elles ne servent qu'au constructeur pour concevoir et perfectionner son système. Nous ne présenterons donc pas ces équations, qui sont par ailleurs disponibles dans plusieurs ouvrages et articles. Cependant, savoir sur quels concepts se basent ces *méthodes* peut s'avérer utile à l'ingénieur du son pour mettre en place son système, c'est pourquoi nous présentons leur principe de fonctionnement de manière détaillée et accessible, mais sans aller au-delà.

Triangulation

La triangulation est certainement la méthode de localisation la plus connue du grand public, bien qu'elle ne soit en réalité pas la plus exploitée. Cette méthode suppose de posséder deux bases dont on connaît la position géographique, et repose sur la mesure des angles entre les bases et le point à localiser. En connaissant les angles formés entre l'axe de deux bases et l'objet, à l'emplacement de chaque base, il n'existe qu'un seul et unique point d'intersection entre ces deux axes. Grâce à une résolution trigonométrique, les coordonnées de la cible peuvent être déterminées.

Multilatération (native ou différentielle)

La multilatération repose sur la connaissance de la distance séparant la cible de plusieurs bases de réception.

Si l'on considère une seule base, l'ensemble des solutions possibles pour l'emplacement de la cible connaissant la distance base-cible, est une sphère (en 3D) ou un cercle (en 2D) entourant la base, dont le rayon correspond à la distance mesurée. En ajoutant une seconde base, l'intersection des solutions pour chaque base réduit les positions possibles à un cercle (en 3D), et à deux points (en 2D) [fig. 4.5 page suivante]. Dans ces deux points, l'un représente la solution réelle et l'autre la solution ambiguë, sans qu'ils puissent être distingués par l'algorithme.

L'ajout d'une troisième base permet de résoudre l'ambiguïté des deux solutions 2D et d'arriver à un unique point, et donne deux points en 3D. Quatre bases permettront de n'obtenir qu'un unique point en 3D [fig. 4.6 page ci-contre]. Les troisièmes et quatrièmes bases ne devront pas former une ligne droite avec les premières, car dans ce cas de figure elles ne permettront pas de réduire le nombre de solutions possibles. L'ajout de plusieurs bases permet également d'ajouter en précision.

Cette méthode, basée connaissance de la distance entre chaque base et la cible est appelée trilatération ou « true range multilateration », que l'on choisira de traduire ici par *multilatération native*.

La mesure de la distance peut s'avérer coûteuse, mais est indispensable pour réaliser une multilatération native. L'un des procédés pour connaître la distance est de mesurer le temps de propagation d'un signal, connaissant sa vitesse de propagation dans le milieu considéré. Ceci nécessite d'avoir des horloges synchronisées entre les bases et la cible, pour mesurer très précisément le temps écoulé entre l'émission et la réception. Souhaitant avoir une cible mobile, il est compliqué d'avoir une synchronisation parfaite à distance, là où de simples câbles suffisent à synchroniser parfaitement les bases. Cela

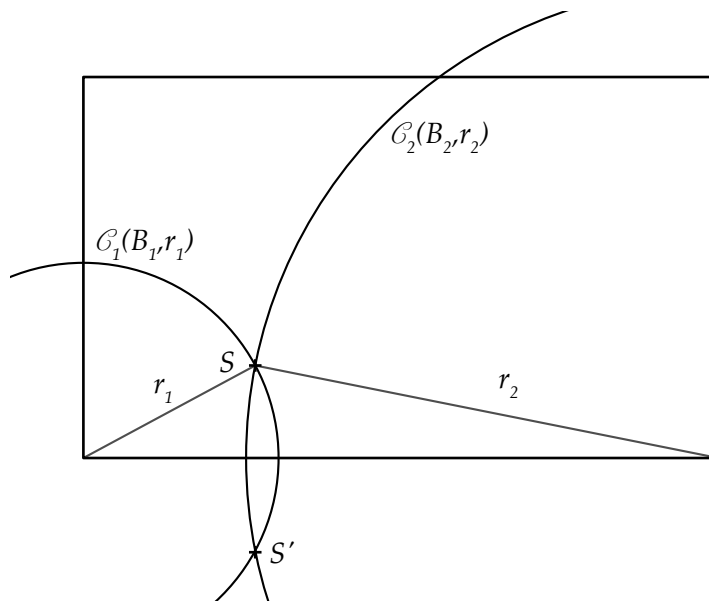


FIGURE 4.5 – Deux solutions S et S' lors d'une multilatération native avec deux bases. La mesure des distances r_1 et r_2 permet de tracer les cercles \mathcal{C}_1 et \mathcal{C}_2 et de déterminer les solutions à leurs intersections.

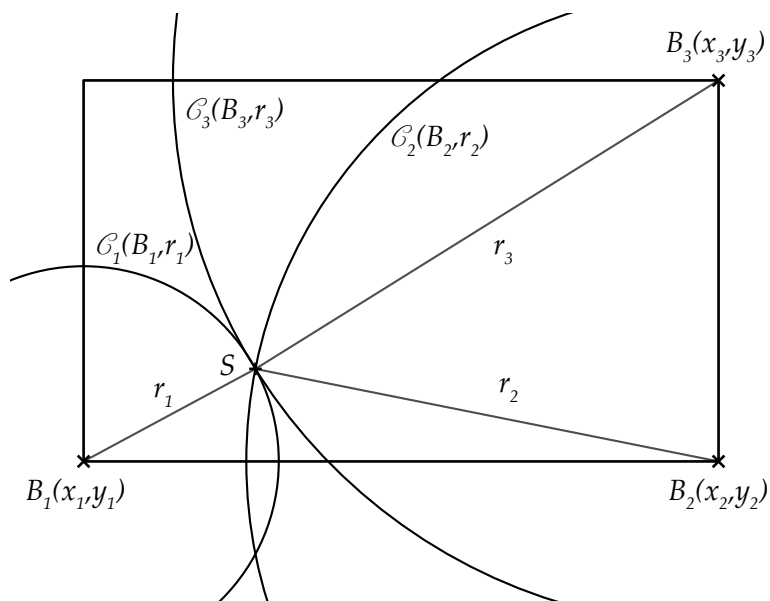


FIGURE 4.6 – Solution S unique lors d'une multilatération native avec trois bases. La mesure des distance r_1 , r_2 et r_3 permet de tracer les cercles \mathcal{C}_1 , \mathcal{C}_2 et \mathcal{C}_3 , et de déterminer la solution unique à leur intersection.

suppose d'utiliser des horloges extrêmement régulières et précises, comme les horloges atomiques par exemple, ce qui rend cette méthode couteuse.

La solution est alors non pas de mesurer les temps d'arrivée à chaque base, mais les *différences* de temps d'arrivée⁵ entre les bases. Cette technique est appelée « pseudo-range multilateration », traduisible par *multilatération différentielle*.

Si l'on considère une seule base, il n'est pas possible de mesurer des différences (de temps ou d'intensité, voir la précédente note de bas de page). Avec deux bases, l'on obtient une différence de temps ; en connaissant les propriétés du signal émis, il est facile de convertir cette différence de temps en une différence de distance. Nous connaissons alors la position de deux bases, et la différence de distance d'un point à ces deux bases. La courbe de l'hyperbole représente par définition « L'ensemble des points d'un plan dont la valeur absolue de la différence des distances à deux points fixes (foyers [nos bases, ndr]) de ce plan est constante »⁶.

La solution des positions possibles pour la cible correspond donc à un hyperboloïde (en 3D) ou à une hyperbole (en 2D). En ajoutant une base l'on arrive à deux différences de temps utiles, réduisant la solution à l'intersection des hyperboloïdes formés par les différences de temps deux couples de bases (par exemple 1-2 et 2-3) [fig. 4.7 page suivante]. Enfin, l'ajout d'une quatrième base permet de réduire le champ des possibles à une ou deux solutions selon les cas. Une dernière base permettra de lever toute ambiguïté sur la position de la cible [fig. 4.8 page ci-contre]. Tout comme pour la multilatération native, considérer le problème en 2D permet de réduire le nombre de bases nécessaire.

4.3.3 Les technologies de localisation

Un système discret

Le choix des technologies de localisation se porte sur des systèmes qui peuvent être cachés des yeux du public. Par exemple, pour le suivi dans le cinéma d'animation, les acteurs utilisent des combinaisons spécifiques portant de nombreux repères visibles. Dans le spectacle vivant, il n'est pas possible de retoucher l'image après coup, il faut donc que les balises puissent se dissimuler dans des costumes, que les récepteurs puissent s'installer hors de la

5. A travers la différence de temps d'arrivée, c'est la différence de distances qui est recherchée. D'autres méthodes telles que la mesure de l'atténuation du signal ou des différences d'atténuation du signal permettent de connaître la distance et les différences de distance. Pour ne pas perdre le lecteur avec des explications trop généralistes, nous gardons dans ce texte le cas des différences de temps, car il s'agit de l'implémentation la plus fréquente. Il suffit de garder en tête que d'autres méthodes existent.

6. Définition donnée par le dictionnaire Larousse.

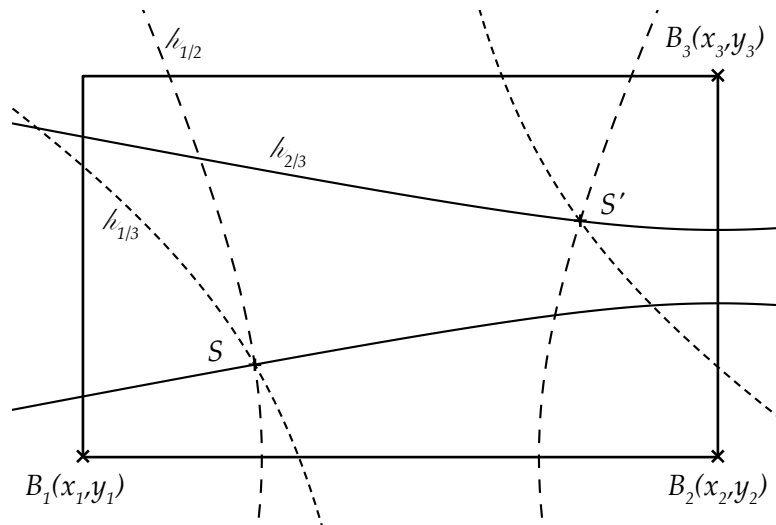


FIGURE 4.7 – Deux solutions S et S' lors d'une multilatération différentielle avec trois bases. À partir des différences de distance mesurées entre 1/2, 2/3 et 1/3, on a pu tracer les hyperboles $h_{1/2}$, $h_{1/3}$ et $h_{2/3}$ et déterminer les solutions à leurs intersections. On voit bien que seulement deux hyperboles sont utiles pour déterminer les intersections.

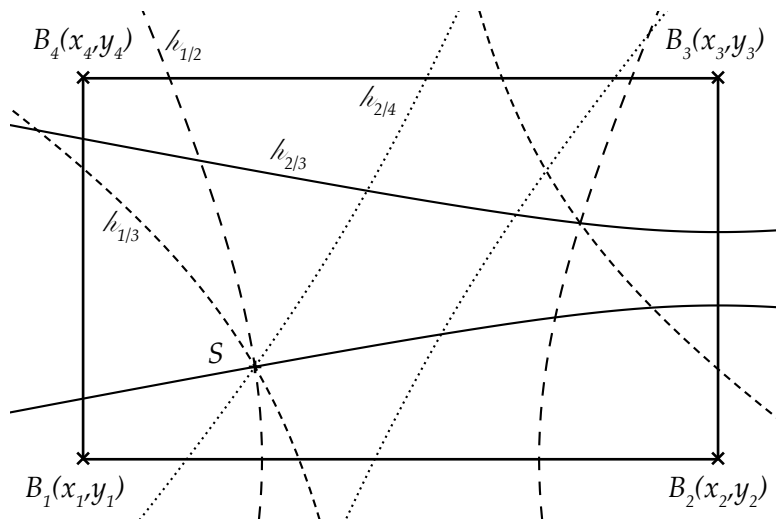


FIGURE 4.8 – Solution S unique lors d'une multilatération différentielle avec quatre bases. À partir des différences de distance mesurée 1/2, 2/3, 1/3 et 2/4, on a pu déterminer la solution unique S à l'intersections des hyperboles.

vue du public, et que l'ensemble fonctionne avec des signaux en dehors de nos champs de perception visuelle ou auditive⁷.

La technologie choisie doit également s'intégrer avec les autres présentes sur scène. C'est le cas des émetteurs radiofréquence, qui doivent pouvoir cohabiter avec les autres émissions électromagnétiques en présence (microphones et systèmes d'intercom sans fil, wifi). L'analyse vidéo basée sur une émission infrarouge devra aussi être attentive à la présence de projecteurs infrarouge sur scène, utilisés pour la surveillance du plateau lors de déplacements automatisés de décors par exemple, ou émise par des éléments du décor (comme des néons par exemple).

Angle d'incidence (AOA, *Angle of Arrival*)

L'angle d'incidence, nécessaire pour effectuer une triangulation peut être mesuré de deux manières : par réception d'un signal fréquence (généralement un signal radio, dit RF) ou bien avec un système optique. La mesure de l'angle d'arrivée d'un signal RF se fait⁸ avec une série d'antennes regroupées, en mesurant les différences de phase du signal capté, entre les différentes antennes, les mesures de phases étant mesurées à l'aide de différences de temps d'arrivée (TDOA).

Différences de temps d'arrivée (TDOA, *Time Differences of Arrival*)

Cette technologie permet d'obtenir des données permettant de calculer la position d'une source par multilatération différentielle. Lorsqu'un signal est émis, son temps de propagation fait qu'il arrive aux points de mesure à des instants différents. Ceci s'applique avec des signaux dont on connaît la vitesse de propagation : elle peut s'effectuer avec du son (des ultrasons), mais lui seront généralement préférés les signaux radiofréquences. Les différences de temps sont calculées par corrélation croisée : cette technique de traitement du signal mesure corrélation (c'est à dire la correspondance) de deux signaux en en faisant varier un par rapport à l'autre. La valeur du délai est donnée pour la valeur maximale de la fonction de corrélation croisée. Sur l'illustration 4.9, pour un même signal reçu d'abord par le récepteur P0 puis le récepteur P1, la corrélation croisée atteint son maximum à 5 ms, ce qui correspond bien au décalage observable à l'œil entre les deux signaux.⁹

7. Ou bien olfactive, mais de tels systèmes ne sont pas à l'ordre du jour...

8. Wikipedia, Contributeur de, *Angle of arrival*, janvier 2019 (URL: https://en.wikipedia.org/w/index.php?title=Angle_of_arrival&oldid=881149505) – visité le 2019-05-15.

9. Wikipedia, Contributeurs de, *Multilateration*, février 2019 (URL: <https://en.wikipedia.org/w/index.php?title=Multilateration&oldid=882224853>) – visité le 2019-05-14.

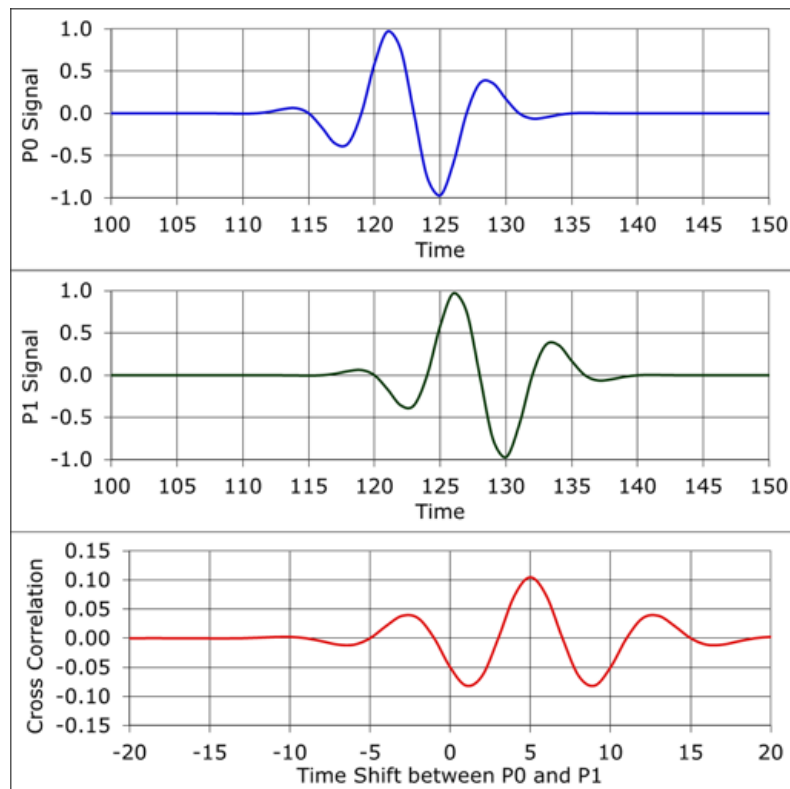


FIGURE 4.9 – Fonction de corrélation croisée pour deux signaux P0 et P1 permettant de calculer leur décalage temporel. (Source : Wikipedia, *Multilateration* ; illustration par TinyPebble, CC BY-SA 3.0)

En utilisant les formules de propagation du signal utilisé, il est possible de transformer les différences de temps en différences de distance et ainsi de chercher les coordonnées d'une source.

Ce système est sensible aux réflexions multiples que peut subir un signal avant d'arriver au récepteur, et ainsi troubler la corrélation croisée, qui trouvera plusieurs solutions à son analyse. Effectuer un fenêtrage temporel du signal reçu permet de pallier à ce problème.

Puissance du signal reçu (RSS, *Received Signal Strength*)

Lorsqu'un signal est émis, sa puissance diminue avec la distance¹⁰ : pour un émetteur omnidirectionnel, de puissance $P(r_0)$ mesurée à la distance r_0 , sa puissance à une distance r peut être exprimée par l'équation (4.3.1) :

$$P(r) = P(r_0) - 10 \log_{10} \left(\frac{r}{r_0} \right) \quad (4.3.1)$$

avec P en décibel, et r en mètres. En connaissant la puissance émise $P(r_0)$, il est donc possible de connaître la distance r séparant l'émetteur du récepteur. Cette mesure, combinée sur plusieurs récepteurs permet de procéder à une multilatération native pour déterminer la position de la cible.

Analyse vidéo

Les systèmes optiques permettent de localiser les points en se basant sur des techniques d'analyse d'images, en suivant un point lumineux situé en dehors du domaine de perception humaine, généralement une émission dans le domaine infrarouge. Le signal vidéo se situant en dehors du domaine de couleur mesuré sera filtré, ne laissant que le point d'intérêt dans l'image, dont les coordonnées dans le point sont données par la position de ses pixels. La caméra permet de positionner le point mesuré sur un plan (le plan vidéo). En recoupant les différents plans, il est possible de connaître la position de la cible.

4.3.4 Les systèmes disponibles en 2019

Les techniques de suivi automatique ont fait leur apparition très tôt dans les techniques scéniques, moins d'une dizaine d'années après les premiers processeurs de son immersif. Aujourd'hui dans le spectacle vivant nous avons réussi à avoir connaissance de quatre systèmes dont un que l'on considèrera à la marge en raison de ses performances. Dans un article de 2013 à l'AES, Hill

10. Hill, Adam J. *et al.*, "Live event performer tracking for digital console automation using industry-standard wireless microphone systems", *Audio Engineering Society*, 2013.

*et al.*¹¹ mentionnent également l'existence d'anciens systèmes, n'étant plus produit, et destinés à l'origine au domaine de l'éclairage : *Wybron Autopilot*, et *Martin Lighting Director*. En plus des systèmes que nous allons mentionner, il existe aussi des systèmes de suivi conçu pour la lumière, en particulier pour les applications de poursuite (*follow spot*) à distance. La plupart reposent sur la présence d'opérateurs derrière des systèmes de pointage. S'il est certainement possible d'exploiter leurs données pour localiser nos sons, leur étude ne nous semble pas justifiée dans ce mémoire, étant donné qu'ils sont principalement utilisés dans des productions plutôt de type concert, sur des scènes de grande envergure, et ne sont pas à notre connaissance fréquemment utilisés dans les théâtres, objet de notre étude. D'autre part, ils visent à localiser un personnage soliste, alors que nous souhaitons *a priori* localiser l'ensemble de la troupe.

Le système le plus ancien à perdurer aujourd'hui est le *TiMax Tracker*, fonctionnant avec le processeur *TiMax* de Outboard UK¹². Basé sur l'émission d'un signal radio ultra large bande dans la bande 6 GHz-8 GHz, il offre une localisation à la précision de 15 cm, avec un mélange de mesure des TDOA et des AOA. Les positions peuvent être transmises de 0,6 à 10 fois par seconde.

Le constructeur TTA dévoile en 2018 une nouvelle itération mise à jour de son ancien système de localisation, désormais plus performant : *Stagetracker II*¹³. Il est basé sur des badges émettant un signal RF ultra-large bande émettant dans la zone 4,9 GHz-5,9 GHz. Il mesure les positions avec une précision de 2 cm, via la mesure des TDOA. Les émetteurs transmettent leur position 100 fois par seconde, ainsi que des informations d'orientation grâce à l'intégration d'un gyroscope, qui transmet ces informations mesurées en interne via l'émission d'un signal radio.

Le constructeur Eliko¹⁴ propose également un système de localisation, mais dont l'utilisation semble plus prévue pour l'industrie, là où ses concurrents cités ici se concentrent sur l'approche spectacle vivant, et proposent des systèmes plus performants. Celui-ci repose également sur une mesure des TDOA, via l'émission par les balises d'un signal ultra-large bande situé dans la zone 3,3 GHz-4,7 GHz. La précision peut varier de 5 cm à 30 cm. La fréquence de calcul de la position détermine le nombre de balises qui peuvent être utilisées. Ainsi, on ne pourra suivre que 4 personnes à 10 positions par

11. Hill, Adam J. *et al.*, "Live event performer tracking for digital console automation using industry-standard wireless microphone systems", *Audio Engineering Society*, 2013.

12. *TiMax Tracker Outboard*, Site web (URL: <http://www.outboard.co.uk/timax.html#timax-tracker>) – visité le 2019-05-20.

13. *TTA Stagetracker II*, Site web (URL: <https://www.tta-sound.com/>) – visité le 2019-05-20.

14. *Eliko KIO RTLS*, Site web (URL: <https://www.eliko.ee/products/kio-rtls/>) – visité le 2019-05-20.

seconde (10 Hz), 10 personnes à 4 Hz, et 40 à 1 Hz, n'autorisant donc pas les mouvements rapides. Tous ses concurrents permettent d'utiliser bien plus de tags, à une fréquence tout de même élevée.

Enfin, il existe le système *BlackTrax* de CAST Soft¹⁵, qui repose sur une localisation vidéo de petites boules émettant un signal infrarouge. Celles-ci peuvent être cachées sous un costume. En cas de perte de vue des émetteurs par le système, une centrale inertielle incluse dans l'émetteur permet de prendre le relais le temps nécessaire, en transmettant les déplacements de l'interprète par signal radio. En raison du fonctionnement sur infrarouge, l'utilisation de ce système n'est pas possible en plein jour, et il faut faire attention à son utilisation avec des lumières infrarouges autres, qui sont très fréquemment utilisées dans les théâtres dès lors qu'il y a de la machinerie, pour offrir des vidéos de surveillance aux opérateurs de machinerie. Les boules infrarouges émettant un signal distinctif, le système peut néanmoins les repérer dans ce contexte, mais il faudra se prémunir de toute lumière infrarouge lors de la calibration du système. Le nombre de positions mesurées par secondes est de 100, et le système transmet également l'orientation de l'interprète via un signal radio. Ce système possède la particularité de permettre de suivre trois points indépendants pour un seul émetteur (les trois petites sphères émettant la lumière infrarouge sont connectées à l'émetteur). Ceci permet par exemple de dessiner une silhouette, en localisant les deux mains et la tête de manière distincte, et offre de nombreuses perspectives créatives.

4.4 Interfaçage (protocoles de communication)

4.4.1 Open Sound Control (OSC)

L'OSC (pour Open Sound Control) est un standard de transmission de données en réseau conçu pour le contrôle en temps réel, exploitant des protocoles UDP ou TCP. Le développement de l'OSC a débuté au centre de recherche « UC Berkeley Center for New Music and Audio Technology », aussi connu sous le nom de CNMAT. Souvent présenté comme le successeur de la norme MIDI, qu'il n'a pourtant pas supplanté, l'OSC présente plusieurs avantages par rapport à celle-ci, et notamment sa conception pour le réseau. Contrairement à la norme MIDI, qui est définie autour de la structure musicale, l'OSC permet de nommer les messages comme souhaité, et supporte des valeurs numériques à très haute résolution, là où le MIDI ne se contente que de 128 valeurs (256 valeurs pour le *pitch bend*).

Un paquet OSC se compose d'une destination (l'opérateur renseigne l'adresse IP et le port de destination), et du message : une chaîne de caractère, l'adresse

15. *Blacktrax*, Site web (URL: <http://blacktrax.cast-soft.com>) – visité le 2019-05-20.

(décrivant le paramètre à contrôler) et un argument (la valeur à transmettre ou définir). Les adresses s'articulent sous une logique d'arbres, dont les ramifications sont exprimées par des slash, comme toute adresse en informatique.

Par exemple, un message pour définir la position selon l'axe x de la source 1 des processeurs Holophonix (Amadeus) et Soundscape (d&b audiotechnik) sont rédigé de la forme :

```
/track/1/x 0.489
```

```
/dbaudio1/positionning/source_position_x/1 0.489
```

La communication en OSC supporte aussi l'envoi de données comprenant plusieurs variables, permettant donc de définir simultanément la position sur les trois axes dans un même paquet :

```
/track/1/xyz 0.489 0.64 0.257
```

```
/dbaudio1/positionning/source_position/1 0.489 0.64 0.257
```

Certains constructeurs nouent des partenariats qui permettent à l'opérateur de ne pas avoir à définir les fonctions lui-même.

Il est coutume, dans l'implémentation de l'OSC, que l'appareil récepteur définisse une liste de messages permettant de le piloter, et qu'il soit possible dans l'appareil émetteur de renseigner pour chaque paramètre l'adresse définie par le récepteur, à laquelle le message doit être envoyé.

La transfert d'informations depuis le processeur de spatialisation vers système de localisation peut être intéressante pour des fonctionnalités d'ergonomie. Par exemple acquérir le nom des sources dans le processeur et les afficher dans l'interface du système de localisation en regard du tag auquel la source a été affectée, la couleur choisie, *etc.*

4.4.2 Real Time Tracking Protocol - Motion (RTTrPM)

Le *Real Time Tracking Protocol* (RTTrP)¹⁶ a été développé au sein de la société CAST Soft, qui proposent parmi leurs produits le système de suivi BlackTrax. Il constitué de deux protocoles de communication en réseau destinés à transmettre les données de suivi : le RTTrPM (Motion) et le RTTrPL (Lighting). Bien que développé par une société privée, ce protocole est sous licence libre (License MIT), assurant la possibilité de modifier, distribuer et utiliser le protocole librement pour des usages priés ou commerciaux. Sa structure a néanmoins été pensée autour des spécificités du système Black-

16. Cochrane, Marty et Kuriakose, Robbins, *Real-Time Tracking Protocol*, Site web (URL: <https://rttrp.github.io/RTTrP-Wiki/index.html>) – visité le 2019-05-15.

Trax, aussi il semble peu probable que d'autres systèmes de suivi intègrent ce protocole à l'avenir. La spécificité du système de BlackTracx est qu'en plus de pouvoir suivre la position d'un individu, il est également possible d'y ajouter trois balises supplémentaires, et donc d'acquérir d'autres positions comme celle des mains, des pieds, *etc.*

Le protocole RTTrPL est adapté pour le contrôle de la lumière dans un environnement DMX, tandis que le RTTrPM fournit des données de positionnements utilisables pour de nombreuses autres utilisations telles que l'audio.

Les paquets RTTrPM sont des paquets UDP (User Datagram Protocol). L'entête de paquet est commun pour les paquets RTTrPL et RTTrPM, puis est suivi d'un entête spécifique RTTrPL ou RTTrPM, pouvant contenir ou non un horodatage. Un paquet RTTrPM est ensuite composé de plusieurs modules, en peut en contenir jusqu'à 80. Chaque module contient un ensemble de données de localisation :

- la position du barycentre (coordonnées cartésiennes 3D);
- l'orientation (Euler et Quaternion);
- la position de points de suivi individuels (coordonnées cartésiennes 3D);
- l'accélération et la vitesse du barycentre;
- l'accélération et la vitesse de points de suivi individuels.

4.4.3 MIDI Show Control (MSC)

Le protocole MIDI (Musical Instrument Digital Interface) a été créé dans les années 1980 pour permettre l'échange de données musicales entre différents appareils tels que des instruments, des séquenceurs, *etc.* Les messages qu'il permet d'envoyer sont structurés autour du fonctionnement de la musique et des synthétiseurs. En outre, il définit un certain nombre de notes (128, numérotées de 0 à 127) toutes espacées d'un demi-ton. Un message MIDI comprend les informations : note-on, note-off, vitesse, pression *after-touch*, *control change*, *program change*, *pitch-bend*. D'autres messages peuvent être transmis, entre autres le volume et le panoramique, messages qui peuvent être intéressants pour l'implémentation de ce standard dans une console.

Afin de permettre la personnalisation des messages transmis via le MIDI, le standard inclut la possibilité de transmettre des messages propriétaires, personnalisables par les constructeurs, appelés *Systeme Exclusive (SysEx)*. Quelques sous standards (en réalité des extensions du *SysEx*) ont été développés pour permettre d'interconnecter d'autres types d'équipements en prévoyant des messages plus adaptés. Ont ainsi été créés le *MIDI Show Control (MSC)* dans les années 1990 et le *MIDI Machine Control (MMC)*, qui standardisent un certain nombre de commandes courantes dans les systèmes de *show*

control, comme déclencher un évènement.

Les processeurs de spatialisation sonore actuellement disponibles sur le marché n'intègrent pas de commandes MIDI pour commander la position des sources sonores, leur préférant des protocoles plus adaptés, comme l'OSC. Si le concepteur sonore utilise des processeurs de spatialisation sonore, il y a de fortes chances qu'il n'ait pas à exploiter la norme MIDI pour spatialiser les voix.

Néanmoins, s'il décide de ne pas utiliser de processeur de spatialisation sonore, c'est qu'il choisit de venir piloter une matrice, ou bien des envois d'AUX dans une console. Comme nous l'avons vu dans la première partie, concevoir ce type de sonorisation avec un placement dynamique des sources sonores suppose de concevoir soi-même une couche logicielle pilotant automatiquement les paramètres d'envoi d'AUX ou de la matrice en fonction du positionnement de l'interprète. Or, la plupart des consoles actuellement sur le marché ne sont pilotables qu'à travers la norme MIDI, à l'exception de quelques-unes (comme les consoles DiGiCo), intégrant l'OSC. L'utilisation de la norme MIDI reste donc restreinte à un cas d'utilisation particulier.

4.5 Corrélation

La précédente section était dédiée aux protocoles permettant de transmettre l'information du positionnement d'un interprète à un système de spatialisation sonore. Nous avons vu que les protocoles transmettent les coordonnées d'un objet, sans autre information. La dernière étape est d'associer ces données de localisation à une source dans un processeur de spatialisation sonore, en s'assurant de décrire la position des objets de manière uniformisée, pour ne pas avoir de décalage.

L'orientation de l'espace dans le monde du théâtre

Comme en navigation, l'univers du théâtre a développé un vocabulaire spécifique pour se repérer dans l'espace. Connaître ce vocabulaire permet de se comprendre et de ne pas confondre les directions, que l'on soit dans le public en regardant la scène, ou bien sur scène en regardant vers le public. Nous donnons ici également les équivalents en anglais, la plupart des processeurs n'étant disponibles que dans cette langue.

Côté Jardin, Côté Cour : lorsque l'on fait face à la scène depuis le public, Jardin correspond à la gauche de la scène, et cour à la droite. En anglais, côté cour se dit *stage left* ou *prompt side*, et côté jardin se dit *stage right* ou *opposite prompt*. Pour se situer dans l'auditorium, l'on parle de *house right* côté cour et de *house left* côté jardin.

Avant scène : ce nom est assez transparent. L'on peut parfois parler de la face, du proscenium, ou bien faire référence à l'orchestre (situé dans une fausse à l'avant-scène). En anglais, on dit *downstage*.

Lointain : aussi fond de scène, arrière-scène. En anglais cela se traduit par *upstage*.

Grill ou cintres : désigne l'espace situé au-dessus de la scène. C'est un espace technique qui comprend perches et ponts permettant d'y accrocher les équipements lumineux ainsi que les éléments de décor.

Façade : ce terme est spécifique à l'univers du son, et désigne la régie son principale, souvent positionnée dans le public et faisant face à la scène. On parle en anglais de *FOH* (*Front of House*).

Repère et système de coordonnées

Les coordonnées d'un point dans l'espace sont décrites à l'aide d'un système de coordonnées. Pour qu'il n'y ait pas d'erreur dans l'interprétation des coordonnées, il faut que les deux appareils fonctionnent autour des mêmes critères :

1. Le système de coordonnées (cartésien, sphérique, cylindrique, *etc.*);
2. La position de l'origine du repère;
3. L'orientation des axes dans l'espace;
4. L'échelle de chacun des axes.

Deux approches se distinguent pour la transmission des données de position, en ce qui concerne la référence des unités. La première consiste à transmettre les données de position suivant les unités du système international : les distances seront exprimées en mètres, et les angles en degrés. Les valeurs transmises correspondront à leur valeur selon l'unité considérée. Mais il existe une autre approche, qui est généralement utilisée uniquement en coordonnées carésiennes. Elle consiste à définir un rectangle dans le plan horizontal (x, y) , définissant la surface sur laquelle les sources sont positionnées. Les positions des sources sont définies entre -1.0 et $+1.0$ sur x et sur y , la position $(1, 1)$ représentant le coin supérieur droit du rectangle, $(-1, -1)$ le coin inférieur gauche, et $(0, 0)$ son centre. Cette approche tient certainement de la coutume d'exprimer en OSC la valeur des paramètres entre 0 et 1, ce qui présente l'avantage de s'affranchir des unités. Néanmoins, ici il est généralement possible de dépasser -1 ou 1 , et donc de dépasser du cadre délimité.

On s'assurera que le système de localisation possède la même approche que le système de spatialisation sonore, le plus simple étant généralement de s'en tenir aux recommandations et partenariats annoncés par le constructeur.

Chapitre 5

Reproduire le mouvement

Introduction

Les processeurs et logiciels de spatialisation offrent une approche simple à la spatialisation des sources dans l'espace : il suffit d'indiquer la position d'une source, et le processeur calcule automatiquement l'audio que doit reproduire chaque haut-parleur pour que l'auditeur localise le son à cet emplacement. Nous avons vu comment capter le mouvement, nous nous préoccupons maintenant de comment le restituer.

Les « anciennes » méthodes furent décrites plus haut, nous nous concentrons maintenant sur l'utilisation de processeurs et logiciels de mixage objet. ceux-ci mettent en place des algorithmes, qui sont analysés dans ce chapitre. Braxton Boren¹ distingue les techniques de spatialisation sonore selon deux catégories : les techniques basées sur le champ acoustique (WFS et Ambisonie), et les techniques multicanal (VBAP et autres).

Dans le contexte du spectacle vivant, il est très souvent préféré par les concepteurs d'avoir recours à un processeur plutôt qu'à un logiciel de spatialisation sonore. L'avantage du processeur est qu'il est prêt à l'emploi : il possède la puissance de calcul nécessaire, dispose des connectiques appropriées à ce secteur d'activité, *etc.* Le problème des solutions logicielles est qu'elles supposent d'avoir un ordinateur. Ceci n'est pas gênant pour des programmes ne demandant pas ou peu de puissance de calcul, mais ces algorithmes sont très gourmands. Il faut donc un ordinateur taillé pour la tâche, ce qui demande au concepteur un temps supplémentaire de préparation très conséquent.

1. Boren, Braxton, "History of 3D Sound", *Immersive Sound* 2018.

5.1 Algorithmes à synthèse de champs sonores

5.1.1 Wave Field Synthesis, W.F.S

La synthèse de front d'onde a été développée à la fin des années 1980 à l'université de Delft aux Pays-Bas par A.J. Berckout dans ce qui était à l'origine des recherches sur les ondes sismiques². Ce procédé est extrêmement séduisant pour notre application dans la mesure où il promet de synthétiser un champ acoustique réaliste. C'est-à-dire qu'avec ce procédé d'holophonie, il serait possible de positionner un son à un point précis de l'espace, d'en simuler l'émission en ce point, et ainsi d'en avoir le champ acoustique comme s'il avait été vraiment émis en ce point.

Dès 1997, Evert Walter Start, doctorant à l'université de Delft sous la direction de A.J. Berckout, s'intéresse à l'utilisation de la WFS dans un contexte de sonorisation. Il étudie dans sa thèse³ les principes scientifiques à l'œuvre dans la WFS, et consacre deux chapitres à la conception et l'évaluation de systèmes de sonorisation par WFS.

L'étude de cette technologie continue dans le monde de la recherche, avant de percer le marché des solutions industrielles, en 2004 avec la solution de Isono, entreprise aujourd'hui rachetée par Barco, suivi quelques années après par Sonic Emotion, vers 2010, Astro Spatial Audio en 2017 et Amadeus en 2018⁴.

De par son application relativement évidente à un contexte de sonorisation, la WFS a fait l'objet d'un certain nombre d'études, relativement complètes, quand à son utilisation pour la sonorisation. Malheureusement, les autres algorithmes n'ont pas fait l'objet de recherches si prolifiques pour un contexte de sonorisation.

Principe scientifique

Le fonctionnement de la WFS repose sur le principe de Huygens, qu'il publie en 1690 dans son *Traité de la lumière*. Celui-ci dit en somme qu'un front d'onde émis par une source que l'on dira primaire, peut être considéré comme la somme des contributions d'un ensemble de sources secondaires, émettant des ondelettes qui, conjointement, contribuent à générer un front d'onde secondaire, en tout point identique au front d'onde primaire. En cal-

2. Sporer, Thomas *et al.*, "Wave Field Synthesis", *Immersive Sound* 2018.

3. Start, Evert Walter, *Direct sound enhancement by wave field synthesis*, Thèse de doctorat, Delft University of Technology, Delft, Netherlands, 1997.

4. TiMax fonctionne sur une matrice de délais et de niveaux, ce n'est pas de la WFS, bien que cela en soit proche par certains aspects. Ceci est également valable pour Soundscape qui fonctionne avec le processeur DS100 est une matrice, en outre, le constructeur ne parle pas de WFS. Enfin, L-ISA ne donne aucune précision, mais il semblerait que ce système n'intègre aucunement des délais.

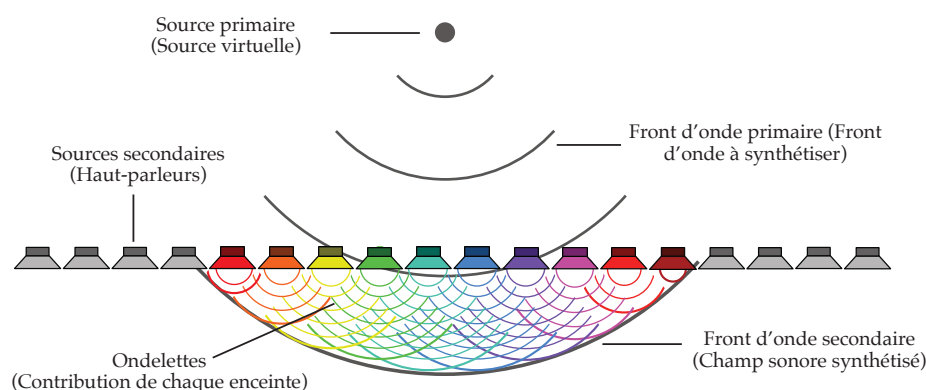


FIGURE 5.1 – Principe de fonctionnement de la WFS.

culant la contribution que chaque source secondaire doit apporter, il est possible de recréer le front d'onde en aval des sources secondaires, bien que la source primaire soit absente.

Selon le modèle scientifique décrit dans plusieurs travaux et ouvrages^{5 6 7}, pour obtenir une synthèse parfaite du champ sonore, les sources secondaires doivent être d'une part réparties de manière continue et infinie, et d'autre part constituer une surface fermée autour de la zone de restitution, ou bien une ligne infinie face à la zone de restitution.

Dans son application à la sonorisation, la source primaire est notre source que l'on souhaite sonoriser (comme un acteur sur scène par exemple), et les sources secondaires sont des haut-parleurs. Le front d'onde primaire est le signal capté par le microphone, et les contributions des sources secondaires *alias* les haut-parleurs, sont calculées par un algorithme hébergé sur un processeur.

Les haut-parleurs synthétisent donc le front d'onde sur la zone d'audience tel qu'il aurait été émis par la source primaire. C'est en cela que réside la particularité de cet algorithme. Là où les algorithmes multicanaux ne permettent de positionner les sources que *sur* les lignes d'enceintes, cette technologie permet de simuler une source derrière les enceintes (en amont du banc) mais également devant les enceintes (en aval du banc), l'on parlera dans ce cas de source en focalisation.

Il semble évident dans ce cadre qu'il est impossible d'avoir une distribution continue et infinie de haut-parleurs : nous avons dans la réalité une distribution discrète et finie de haut-parleurs, c'est-à-dire qu'ils forment un

5. Start, Evert Walter, *Direct sound enhancement by wave field synthesis*, Thèse de doctorat, Delft University of Technology, Delft, Netherlands, 1997.

6. Caulkins, Terence, *Caractérisation et contrôle du rayonnement d'un système de Wave Field Synthesis pour la situation de concert*. Thèse de Doctorat Université Pierre et Marie Curie - Paris 6, Paris, 2007.

7. Sporer, Thomas *et al.*, "Wave Field Synthesis", *Immersive Sound* 2018.

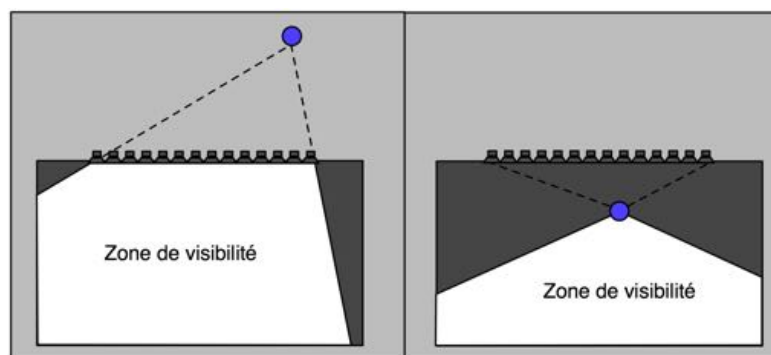


FIGURE 5.2 – Effet de fenêtre pour une source et source focalisée (Source : Caulkins, 2007)

segment (certaines configurations permettent néanmoins de faire un cercle autour des auditeurs), et que les haut-parleurs sont espacés entre eux. Ces caractéristiques ne sont pas rédhibitoires pour synthétiser un front d'onde, mais ils ont deux conséquences : le fenêtrage de la scène sonore, et de repliement spatial. D'autres conséquences résultent de cette approximation par rapport au modèle théorique, elles sont en outre clairement énoncées par Caulkins dans sa thèse de 2007. Nous nous intéresserons ici aux conséquences qui impactent surtout l'ingénieur du son dans son travail de sonorisation, les autres relevant principalement de la conception des processeurs de WFS.

Fenêtrage de la scène sonore

Le banc de haut-parleurs étant fini, il possède des bords qui agissent comme une fenêtre, qui conditionne la visibilité des sources en fonction du positionnement de celles-ci (voir figure 5.2). En tirant deux droites entre la position de la source et les bords du banc de haut-parleurs, l'on définit les limites au-delà desquelles la localisation du son ne sera plus correctement perceptible. Ces zones sont appelées « zone d'ombre » (voir Caulkins). Le son sera toujours audible cependant (à moins que les auditeurs soient également en dehors de la zone de couverture des enceintes, auquel cas le son sera aussi détimbré). Il est donc nécessaire d'avoir un banc WFS assez large pour pouvoir couvrir toute la largeur de la zone d'audience convenablement. Néanmoins, ceci ne change pas beaucoup par rapport à la sonorisation « classique », où l'on utilise des enceintes ponctuelles pour « déboucher » les zones d'audiences mal couvertes par le système principal.

Échantillonnage spatial

Bien connu en traitement du signal, le phénomène d'échantillonnage a lieu dès lors que l'on transpose une description continue en une description

discrète. Comme nous l'avons vu, nous utilisons une distribution discrète d'enceintes (c'est-à-dire un nombre limité d'enceintes), il se produit donc un phénomène dit de « repliement spatial ». L'espacement entre les enceintes ne permet pas de reproduire des phénomènes de petite dimension, et introduit en conséquence des interférences. En conséquence, au-dessus d'une certaine fréquence le front d'onde ne sera plus correctement reproduit. Cette fréquence f_r varie⁸ avec l'espacement entre les enceintes Δx , l'angle d'incidence α du front d'onde virtuel/primaire sur le banc de haut-parleurs, et dépend de la vitesse du son c :

$$f_r = \frac{c}{2\Delta x \sin(\alpha)}$$

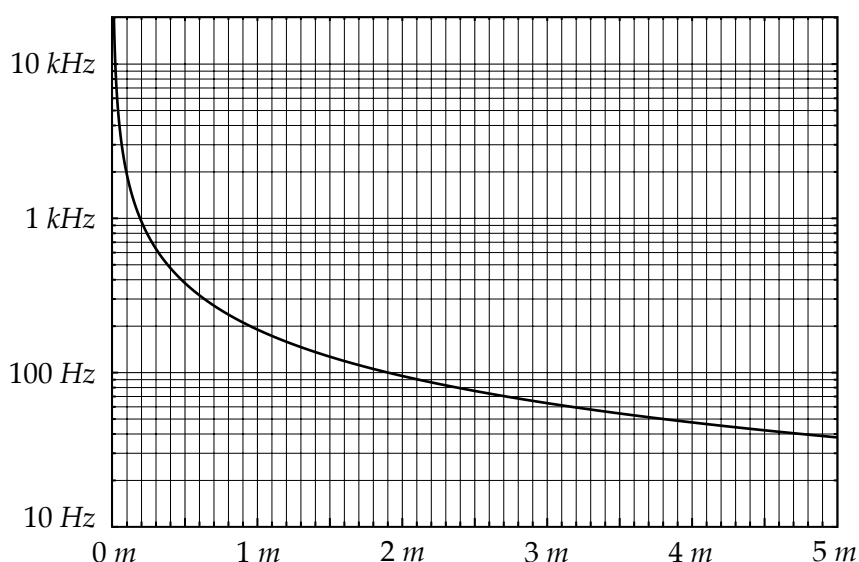


FIGURE 5.3 – Valeur de f_r en fonction de Δx pour $\alpha = 90^\circ$

Les déformations qui interviennent consistent en des baisses de niveau⁹ sur des bandes de fréquence spécifiques, à des emplacements précis de l'espace. Sporer et *al.* expliquent que dans les basses fréquences (donc proche de f_r), ces creux de niveau sont légers et large bande, et dans les hautes fréquences sont plus fortes, mais à bande étroite. Notre audition n'est pas aussi sensible au creux de niveau à bande étroite dans les hautes fréquences que dans les basses fréquences, ce qui a pour avantage de minimiser l'audition de ces déformations. Il est alors évident que plus la fréquence de repliement spatial sera basse, plus les forts creux bande étroite interviendront dans des plus basses fréquences, ou notre audition serait susceptible d'y être plus sensible à ceux creux.

8. Sporer, Thomas *et al.*, "Wave Field Synthesis", *Immersive Sound* 2018.

9. *Ibid.*

Étant donné que cet artefact dépend de l'angle d'incidence du front d'onde virtuel sur le banc de haut-parleurs, cet artefact sera audible sur des sources en mouvement. La sonorisation de sources statiques pourrait donc se contenter d'enceintes plus espacées que la sonorisation de sources mobiles.

Des chercheurs¹⁰ ont proposé plusieurs solutions permettant de limiter cet effet, dont l'une est simplement de ne spatialiser les sons en WFS qu'en dessous de la fréquence de repliement, en utilisant uniquement des différences d'intensité sonore au-dessus. L'efficacité de cette solution est souvent justifiée par le fait que, comme nous l'avons vu dans le chapitre 2, notre oreille est sensible aux différences de temps pour localiser les sons en dessous de 1700 Hz, mais est surtout sensible aux différences d'intensité au-dessus. Si l'on se fie à ce critère, il faut avoir un espacement de 10 cm environ entre les centres acoustiques des enceintes. En réalité, nous avons vu que notre audition est toujours sensible aux différences de temps après 1700 Hz, simplement l'évaluation ne se fait plus sur le signal lui-même, mais à l'aide de son enveloppe.

Cette méthode a pour avantage de n'induire aucun artefact, mais Sporer et al. décrivent bien son inconvénient : les sons en dessous de f_r seront correctement spatialisés, tandis que les sons au-dessus seront perçus comme provenant du banc de haut parleurs. D'autres méthodes sont proposées visant à réduire l'effet des artefacts sur une zone d'écoute.

Dans les faits, la plupart des productions utilisent des systèmes WFS avec des enceintes beaucoup plus espacées, atteignant parfois plusieurs mètres pour les lieux de représentation les plus grands.

Filtrage

La WFS implémente également un filtrage dépendant de la position de la source, permettant de compenser différents artefacts de l'algorithme, tel que l'augmentation des basses fréquences créé par l'échantillonnage¹¹. Pour alléger les calculs de l'algorithme et réduire sa latence, de nombreuses implémentations calculent les caractéristiques des filtres en amont, pour un nombre déterminé de positions.

5.1.2 Ambisonie

L'ambisonie est une technique de spatialisation sonore qui a été développée par Michel Gerzon à l'université d'Oxford dans les années 1970. Bien qu'elle continue d'être très étudiée dans le monde universitaire, l'Ambisonie a rapidement connu des applications commerciales ou orientées grand

10. Sporer, Thomas et al., "Wave Field Synthesis", *Immersive Sound* 2018.

11. *Ibid.*

public.

Les premières applications de l'Ambisonie se trouvent principalement du côté de la post production, des technologies telles que la vidéo dite « VR » et l'écoute binaurale, ainsi que les concerts de musique contemporaine avec spatialisation sonore. La grande popularité de cette technique fait que les études et ouvrages traitant du sujet sont nombreux, mais ils se concentrent généralement sur les situations d'écoute individuelle, qu'il s'agisse de l'écoute au casque (lorsque cette technique est hybridée avec le binaural), ou bien l'écoute sur haut-parleurs, supposant que l'auditeur est au centre du dispositif, seul, ou en tout petit groupe. En effet, dans son fonctionnement de base, c'est à cet endroit que doit se placer l'auditeur pour avoir une bonne reproduction du champ sonore, nous faisant retrouver le problème « sweet spot » de la stéréophonie.

Les études portant sur l'utilisation de cette technique dans un contexte de sonorisation pour de larges audiences sont elles beaucoup plus moins nombreuses. Néanmoins grâce aux évolutions permises par la recherche scientifique, il est désormais possible de toucher des zones d'audiences plus larges que le simple centre du dispositif.

Les produits de sonorisation exploitant l'algorithme Ambisonics sont rares. L'on trouve de nombreux outils publiés par l'Ircam ou bien (Paris 8, l'ICST), mais ces solutions nécessitent un temps de développement qui ne conviens pas pour un certain nombre d'utilisations. L'une des premières solutions commerciales intégrant l'Ambisonie est Ircam Spat de l'éditeur Flux. Bien qu'il s'agisse d'un produit « clef en main », il n'en demeure pas moins une solution logicielle, ce qui présente également un certain nombre d'inconvénients pour l'utiliser dans certains contextes de production. Holophonix, lancé en 2018 par Amadeus, est à notre connaissance le premier processeur à intégrer nommément cette technologie.

Principe scientifique

Analogie avec la prise de son M/S

L'Ambisonie est à l'origine une technique de prise de son puis de reproduction du son, qui fonctionne de manière analogue à la prise de son M/S.

Dans la prise de son stéréophonique, l'on utilise deux microphones de directivité identique (généralement omnidirectionnels ou cardioïdes) formant un couple, dont on ajuste l'angulation et l'espacement afin d'ajuster la captation de la scène sonore au résultat souhaité. Pour écouter la prise de son, le signal de chaque microphone est destiné à un haut-parleur d'un système stéréo.

Dans la prise de son M/S, l'on utilise un microphone cardioïde (M pour

middle) et un microphone bidirectionnel (S pour *side*) pour enregistrer des scènes stéréophoniques. Les deux capsules doivent être coïncidentes, et le lobe positif du capteur bidirectionnel doit pointer vers la gauche. À l'inverse de la prise de son stéréo classique, ces deux signaux bruts ne sont pas destinés à être reproduits par des enceintes : il faut auparavant dé-matricer ces signaux. Les deux lobes du microphone bidirectionnel (S) étant en opposition de phase, si l'on somme son signal au micro cardioïde (M), le signal capté par le lobe négatif de (S) sera annulé. En inversant la polarité de (S) et en le sommant de nouveau à (M), c'est le lobe positif qui est annulé. Par cette opération de dé-matrigage, l'on obtient les signaux d'alimentation des enceintes gauche et droite d'un système stéréo :

$$\begin{aligned}L &= M + S \\R &= M - S\end{aligned}$$

Sur le même principe que le couple M/S, l'Ambisonie utilise un canal « W », analogue à (M) issu d'une captation omnidirectionnelle ainsi que trois canaux « X », « Y » et « Z », analogues à (S), issus de captations bidirectionnelles, chacune pointant vers une direction du système de coordonnées cartésiennes. Pour reproduire le champ sonore ainsi capturé, il suffit de dé-matricer ces différents signaux de manière adaptée à la position de chaque haut-parleur utilisé.

Ambisonie et HOA (Higher Order Ambisonics)

L'Ambisonie repose donc sur la décomposition de l'espace en plusieurs dimensions, appelées harmoniques sphériques. Ces fonctions spatiales (les canaux W, X, Y, Z évoqués plus haut) peuvent être considérées comme des lobes de directivité en 3D, chacune pointant dans une direction de l'espace, permettant ainsi de recomposer le champ sonore environnant le point de captation (dans le cas d'un enregistrement) ou le point de restitution (dans le cas d'une scène générée virtuellement). Dans sa forme de base telle que décrite par Gerzon en 1973, l'Ambisonie fonctionne avec les quatre canaux W, X, Y, Z, qui correspondent aux harmoniques sphériques d'ordre zéro (pour W) et de premier ordre (pour X, Y, Z).

Des formes plus évoluées de l'Ambisonie font intervenir des harmoniques sphériques supplémentaires, donc d'ordre supérieur, dont la directivité dans l'espace s'affine. L'ordre de l'Ambisonie est lié au nombre d'harmoniques avec lesquels l'espace est décomposé et donc à l'ordre de ces harmoniques. Plus l'ordre est élevé, plus le nombre d'harmoniques est grand, plus la résolution spatiale est précise. Au-delà de l'ordre un, on parle d'*Ambisonie d'Ordre Supérieur* ou de *HOA* (pour *Higher Order Ambisonic*). L'ordre 1 compte quatre harmoniques sphériques, l'ordre 2 en compte six, l'ordre 3 en compte huit,

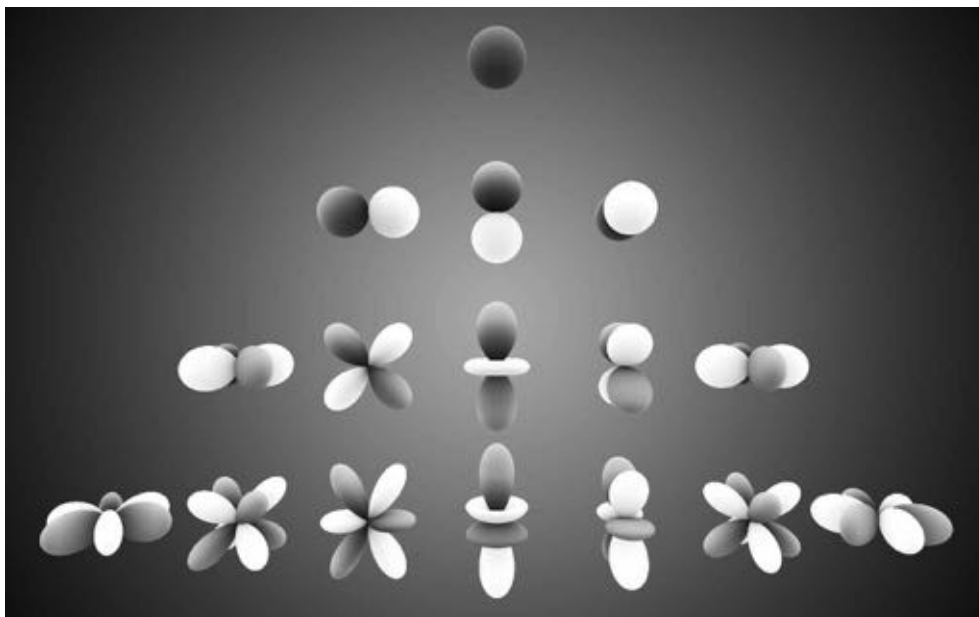


FIGURE 5.4 – Représentation des harmoniques sphériques jusqu'à l'ordre 4. Les lobes sombres sont positifs, les lobes clairs sont négatifs. (Source : Wikipedia, « Harmoniques Sphériques » ; illustration par Inioquilez, CC BY-SA 3.0)

etc. L'illustration 5.4 montre les figures de directivité des harmoniques sphériques d'ordre un à trois.

Ces fonctions de directivité tiennent leur nom *d'harmoniques* sphériques de l'analogie qui peut être faite avec la décomposition en séries de Fourier pour les sons. De même que n'importe quel signal peut être considéré comme la somme de fonctions sinusoïdales d'amplitudes et de fréquences différentes (ses harmoniques), l'espace peut être découpé en fonctions spatiales qui, une fois regroupées, permettent de décrire l'ensemble de l'espace.

Encodage, décodage

À chacun de ces « axes » de captation dans l'espace que sont les harmoniques sphériques est associé un canal audio, cet encodage est créé virtuellement par les processeurs. Nous ne nous sommes intéressés ici qu'à l'encodage de scènes de synthèse, mais il existe des microphones permettant d'enregistrer au format Ambisonique.

Cet ensemble de canaux doit ensuite être décodé pour chacun des haut-parleurs. Chaque enceinte reçoit/reproduit une combinaison pondérée de l'ensemble de ces canaux.

Décodage et système de restitution

Nombre et disposition des haut-parleurs

Le nombre minimal de haut-parleurs pour restituer une scène Ambisonique (non HOA) est de trois en 2D et quatre en 3D. Néanmoins, les systèmes avec si peu de haut-parleurs présentent le désavantage que l'auditeur ne peut être placé qu'au centre du dispositif s'il souhaite percevoir une reproduction correcte du champ acoustique (Daniel 2001).

Afin de permettre plus de souplesse dans la position de l'auditeur et d'élargir la zone d'audience, il est nécessaire de travailler avec plus d'enceintes, et donc de travailler en Ambisonie d'ordre supérieur. Pour reproduire au mieux les contenus Ambisoniques, il est nécessaire d'observer la règle suivante : le nombre de haut-parleurs doit être supérieur ou égal au nombre de canaux Ambisoniques. Soit, si l'on désigne par m l'ordre Ambisonique, et par n le nombre d'enceintes, il faut¹² :

$$\begin{aligned} n &\geq 2 \times m + 1 && \text{en 2D} \\ n &\geq (m + 1)^2 && \text{en 3D} \end{aligned}$$

De manière générale les processeurs ou logiciels ambisonique permettent de décoder une scène sur un nombre d'enceintes inférieur ou supérieur au nombre requis. Si le nombre de haut-parleurs est inférieur au nombre requis, la scène est décodée sans artefact, mais avec une moins bonne résolution spatiale. Lorsqu'une scène est sur un nombre plus important d'enceintes que normalement requis, il peut se produire des artefacts similaires à du filtrage en peigne¹³, il faut alors que l'algorithme ignore les haut-parleurs supplémentaires qui lui ont été déclarés.

Précisons qu'en Ambisonie, quelle que soit la position d'une source, toutes les enceintes contribuent à la restitution du champ sonore. Il sera donc nécessaire d'avoir une disposition de haut-parleurs entourant la zone d'audience, nous verrons plus loin que des méthodes de décodage permettent d'avoir des répartitions de haut-parleurs inhomogènes, laissant donc la possibilité de disposer de plus grande densité de haut-parleurs au cadre de scène, qu'autour du public.

Les haut-parleurs doivent être positionnés en cercle (2D) ou en dôme (3D) autour de l'auditeur, mais leur positionnement est relativement libre par rapport à des dispositifs multicanaux classiques.

Amélioration de la surface d'écoute

La technique de décodage de base de la HOA, appelé décodage *basic*, offre une bonne reproduction du champ sonore au centre du dispositif, pour les basses fréquences. Différentes techniques de décodage ont été introduites

12. Nicol, Rozen, "Sound Field", *Immersive Sound* 2018.

13. *Ibid.*

pour permettre d'assurer une meilleure synthèse du champ sonore sur une zone d'écoute étendue.

Le décodage dit *max r_E* ou *maximum r_E* permet de maximiser la localisation en hautes fréquences en maximisant le vecteur énergie r_E , qui correspond au « taux de concentration de l'énergie » dans la « direction moyenne de provenance de l'énergie »¹⁴. Pour l'exprimer plus simplement, cette méthode accroît en quelque sorte la contribution du haut-parleur le plus proche de la localisation de la source. En outre, plus le paramètre r_E est proche de 1, plus la direction perçue pour une source virtuelle sera stable en s'éloignant du centre du dispositif.

Malham propose en 1992 le décodage *in phase*, qui parvient également à assurer une localisation plus robuste pour les utilisateurs excentrés. Dans ce procédé où l'ensemble des enceintes émettent en phase, plus la direction d'un haut-parleur s'éloigne de celle de la source virtuelle considérée, plus sa contribution sera diminuée, jusqu'à atteindre une contribution nulle pour l'enceinte opposée à la source.

Comme indiqué par Jérôme Daniel¹⁵, pour optimiser le décodage en fonction de l'étendue de l'auditoire à l'intérieur du dispositif de restitution, il est possible d'utiliser des combinaisons de ces décodeurs, chacun traitant une sous-bande du domaine fréquentiel. Pour une position d'écoute centrée l'on recommande une combinaison *basic/max r_E* , pour un public étendu *max r_E /in phase*, et enfin si les spectateurs sont à proximité des enceintes, le décodage *in phase*. L'auteur ne précise pas de fréquence de coupure optimale pour chaque sous-bande, laissant aux lecteurs le soin de le déterminer de manière empirique.

Adaptation à des dispositions de haut-parleurs non homogènes

Deux méthodes classiques de décodage sont traditionnellement utilisées et implémentées dans les dispositifs Ambisoniques : le décodage *direct sampling* et le décodage *mode matching*. Ces deux méthodes offrent de bons résultats de localisation des sources pour un ensemble de haut-parleurs respectant une répartition sphérique et homogène dans l'espace, mais fonctionnent mal sur des répartitions non uniformes d'enceintes. En effet, la différence d'espacement des enceintes provoque une variation d'énergie en fonction de la position lorsqu'une source est en mouvement. Deux méthodes introduites au début des années 2010 permettent d'utiliser des dispositions non uniformes.

La méthode *energy preserving*, de Zotter, Pomberger et Noisternig¹⁶, per-

14. Daniel, Jérôme, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Thèse de doctorat, Université Pierre et Marie Curie, Paris 6, Paris, 2001.

15. *Ibid.*

16. Zotter, Franz, Pomberger, Hannes et Noisternig, Markus, "Ambisonic decoding with

met comme son nom l'indique de maintenir une énergie constante pour une source, quelle que soit sa position.

La méthode *all-round* de Zotter et Frank¹⁷ offre quand à elle la particularité de combiner les trois précédentes méthodes en tirant avantage du VBAP. En effet, celle-ci utilise un ensemble de haut-parleurs virtuels dont la disposition (difficilement exécutable dans le réel, mais offrant d'excellents résultats) est rendue sous forme de sources virtuelles avec l'algorithme VBAP.

5.2 Algorithmes multicanaux

Les algorithmes que nous appelons ici « multicanaux » pour les distinguer de techniques de synthèse du champ acoustique ont des méthodes de fonctionnement beaucoup plus simple. Ceci a pour avantage que ces techniques soient beaucoup moins exigeantes en puissance de calcul. Nous ne présentons ici que les algorithmes au fonctionnement le plus spécifique : le DBAP, VBAP/VBIP et le LBAP. Il existe d'autres algorithmes, comme l'Angular 2D, ou le KNN, mais leur fonctionnement est si simple qu'il n'est pas nécessaire de s'y attarder.

Le problème de la parallaxe

Ces types d'algorithmes spatialisent les sons dans le plan des enceintes, puisqu'ils fonctionnent avec des sources fantômes. Ainsi nous considérons qu'ils sont en ce point analogue à des sonorisations L/R ou L/C/R, apportant simplement tout de même l'avantage de permettre de travailler avec plus d'enceintes, limitant les problèmes de décalage pouvant être rencontré avec la stéréo, puisque l'espacement entre les enceintes est fortement réduit. Ces algorithmes poseront néanmoins un problème de parallaxe pour les auditeurs excentré. En effet, puisque le son est localisé *entre* les enceintes, si l'acteur se tient à distance de celles-ci, les personnes ayant une position excentrée entendront un décalage entre la localisation sonore et la localisation réelle de l'interprète.

5.2.1 Distance Based Amplitude Panning (DBAP)

Le DBAP¹⁸ a été présenté en 2011 par Lossius, Blatazar et de la Hogue. Cet algorithme est celui qui se rapproche plus de la logique que nous avons

and without mode mathing : a case study using the hemisphere", *Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics* 2010.

17. Zotter, Franz et Frank, Matthias, "All-Round Ambisonic Panning and Decoding", *Journal of Audio Engineering Society* 60 2012, Nr. 10.

18. Lossius, Trond, Baltazar, Pascal et Hogue, Théo de la, "DBAP - Distance-Based Amplitude Panning", 2011.

pu avoir jusqu'à présent puisqu'il se rapproche d'une logique matricielle, permettant d'exploiter n'importe quelle disposition de haut-parleurs. En pratique, l'algorithme évalue la distance de la source à chaque haut-parleur, et prend en compte l'atténuation dans l'air pour calculer le niveau sonore auquel chaque enceinte doit reproduire le son. Toutes les enceintes contribuent, il est néanmoins possible de définir des pondérations d'enceintes pour permettre de limiter le son à un sous-ensemble de haut-parleurs.

5.2.2 Vector Base Amplitude & Intensity Panning (VBAP, VBIP)

Le VBAP¹⁹ a été introduit à l'origine par Ville Pulkki dans le journal de l'AES en 1997. Cette technique présentait l'avantage non négligeable de permettre une spatialisation sonore demandant peu de puissance de calcul et n'imposant pas une disposition d'enceintes fixe comme l'imposent les systèmes 5.1 et autres. Cependant les enceintes devront être Cet algorithme repose sur des différences d'amplitude (VBAP) ou d'intensité (VBIP) sonore entre les enceintes.

Pour reproduire l'image d'une source en un point, ces algorithmes utilisent les trois haut-parleurs encadrant la source (en 3D) ou les deux haut-parleurs encadrant la source (en 2D). Chacun des trois haut-parleurs reproduit le son de la source avec un niveau différent. Ces haut-parleurs forment avec le point d'écoute une base vectorielle utilisée pour le calcul du gain appliqué à la source sur chacune des enceintes. Les haut-parleurs doivent être équidistants.

5.2.3 Layer Based Amplitude Panning (LBAP)

Le LBAP²⁰ est une nouvelle technique, introduite en 2016 par Ivica Ico Bukvic à la Virginia Tech aux États-Unis. Son objectif principal est définir une approche du panoramique d'amplitude optimisée pour les dispositions d'enceintes non homogènes. Il offre d'excellents résultats de spatialisation pour des dispositions 2D ou 3D qui ont des zones où les enceintes sont plus nombreuses que d'autres. Cela peut être par exemple une disposition surround concentrant plus d'enceintes devant la scène, ou encore une sphère ou un cube d'enceintes ayant plus de haut-parleurs sur le cercle inférieur que les cercles supérieurs.

L'algorithme sélectionne les deux couches (*layers*) d'enceintes encadrant la position de la source, et évalue d'abord la distance séparant la source à

19. Pulkki, Ville, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", *Journal of Audio Engineering Society* Vol. 45 June 1997, Nr. 6.

20. Bukvic, Ivica Ivo, "3D Time-Based Aural Data Representation Using D4 Library's Layer Based Amplitude Panning Algorithm", *Proceedings of the 22nd International Conference on Auditory Display - ICAD 2016*, juillet 2016.

la couche basse et la couche haute. Une amplitude est associée à chacune des deux couches en utilisant la règle de panoramique sinusoïdale. Puis sur chaque couche sont sélectionnées les deux enceintes autour de la source. Les amplitudes sont calculées pour ces deux paires d'enceintes à partir de l'azimut de la source, et avec la loi de panoramique \cos/\sin .

5.2.4 Panning d'amplitude et de délais basé sur la distance

A notre connaissance, il n'existe pas d'algorithme formalisé autour du concept de panning d'amplitude et de délais, basés sur la distance. Cette approche est similaire dans sa conception à *TiMax*, à la différence que celui-ci calcule sa matrice pour des positions clef (*image definition*). Le panning d'amplitude et de délais consiste quand à lui à utiliser dans la matrice des délais variables. Elle est également très similaire à la WFS, mais elle n'intègre pas de filtrage ou d'autres optimisations.

Néanmoins cette solution est intégrée par des constructeurs tels que d&b audiotechnik ou dans des solutions logicielles telles que celle proposée par Pierre-Olivier Boulant²¹.

21. Boulant, Pierre-Olivier, *Wave Field Synthesis DIY*, site web (URL: <https://wfs-diy.net>) – visité le 2019-05-15.

Chapitre 6

Mixer en conséquence

6.1 Précision de la localisation par traitement du signal

Les différentes techniques de spatialisation sonore de la voix que nous avons étudiées permettent principalement de préciser la localisation des sources dans l'axe azimutal (gauche-droite), en ce qui concerne notre application. Certains (en particulier la WFS) peuvent être en capacité de restituer la distance, mais ce n'est pas le cas d'autres algorithmes, basé sur la création de sources fantômes. Comme nous l'avons vu section §1.1, ce qui nous permet de préciser la localisation dans d'autres dimensions que l'azimut, ce sont surtout les variations de niveau, la modification du spectre (filtrage), et l'acoustique (rapport premières réflexions/champ diffus).

Certains processeurs ou logiciels de spatialisation sonore proposent une série de traitements automatisés permettant de préciser la localisation, voire de donner des sensations de distance, en se basant sur les paramètres évoqués ci-dessus.

6.1.1 Traitement de la source : délai, atténuation et filtrage

Délais

Le délai n'est nécessaire que pour faire un effet Haas, et garder en cohérence la sonorisation avec le son émis sur scène. Certains algorithmes comme la WFS intègrent nativement un délai qui correspond au temps de propagation du son jusqu'aux enceintes, dans la logique de reconstruction d'un front d'onde. Comme pour la spatialisation sonore matricielle, il existe un délai différent pour chaque source sur chaque enceinte. Certains constructeurs, comme d&b audiotechnik avec *Soundscape*, permettent de supprimer ce délai, tout en conservant le même front d'onde¹ : le délai différentiel entre

1. Cette fonctionnalité permet de s'adapter aux situations où un délai pourrait être indésirable entre certaines sources.

les enceintes sera conservé, mais pas le délai de distance de la source.

D'autres algorithmes ne proposent nativement pas de délai basé sur la distance de la source. Deux approches peuvent alors être choisies pour mettre en place un effet Haas en surplus de la spatialisation sonore. La première méthode consiste à appliquer un délai fixe à l'ensemble du système². Cette solution favorise néanmoins l'apparition de positions critiques ou un effet d'écho pourra se faire entendre. La deuxième méthode en ce sens est moins sujette à ce problème, puisqu'elle consiste à personnaliser le délai pour chaque source. Par là, elle se différencie de la spatialisation par matrice ou WFS, puisqu'elle n'associe pas un délai unique pour chaque source sur chaque enceinte. Le délai est appliqué uniquement sur la source, avant traitement dans le bus de spatialisation sonore, qui n'intègre pas de délai spécifique pour chaque enceinte. Selon l'outil, l'on pourra opter pour un délai fixe (par exemple dans une console), que l'on pourra programmer sur des valeurs différentes selon l'emplacement prévu de l'interprète, ou bien, puisque nous disposons d'outils de mixage objet, utiliser un délai dynamique, mis à jour en permanence selon la position de l'interprète. C'est ce que propose par exemple l'Ircam dans leur logiciel de mixage immersif *Panoramix*. Un délai individuel calculé en fonction de la position de la source peut être appliqué ou non.

Le mixage s'effectuant à l'aide d'outils numériques, ils impliquent une latence inhérente au système. Il faudrait en théorie estimer la latence globale du système sonore (*round trip latency*) et pouvoir la déduire du délai de distance.

Atténuation et filtrage

Le niveau sonore est un indice de distance que l'on évalue en partie de manière comparative ou différentielle : deux sources de niveau différent, ou une source dont le niveau baisse par exemple. Cet effet peut être aisément reproduit par le régisseur, mais est intégré dans certains processeurs ou logiciels de mixage immersif. Lorsque l'algorithme prend en charge l'asservissement du niveau sur la position, cela demandera à l'opérateur souhaitant mixer les voix (pour quelque raison que ce soit³) d'adapter ses gestes par rapport à ses anciennes habitudes de mixage : entre autres il ne faudrait pas compenser la baisse de niveau apportée par le système.

Filtrer un signal permet également d'accentuer la sensation d'éloignement, en reproduisant l'atténuation spectrale de l'air. Néanmoins, cet effet est plus complexe à mettre en œuvre que le changement de niveau si le pro-

2. Nous parlons ici du système principal. Il est évident que les sous-systèmes de renfort sonore devront avoir des délais appropriés.

3. Nous y reviendrons plus loin, mais il peut exister plusieurs raisons de vouloir mixer les voix : problèmes de filtrage en peigne évoqués dans la première partie du mémoire, clarifier la scène sonore en présence de nombreux microphones, etc.

cesseur ne propose pas une telle fonction, puisqu'il faut assigner l'évolution des paramètres d'un filtre à la position de la source. Certaines solutions de mixage comme le logiciel *Panoramix* de l'Ircam intègrent la possibilité d'enclencher ce type de traitements. Pour ceux qui ne le permettent pas, une solution peut permettre de le mettre en œuvre, mais elle suppose de pouvoir commander les paramètres d'une console via un protocole quelconque, ou bien de pouvoir faire passer l'audio souhaité par un ou des traitements informatiques au sein d'un ordinateur, dont les paramètres peuvent être pilotés de manière externalisée (c'est-à-dire autrement que par l'interface proposée par le développeur). L'on pourra alors, par exemple, définir un filtrage maximal, s'appliquant à une certaine distance, et un filtrage minimal (a priori *pas de traitements*) lorsqu'il n'y a pas d'éloignement. Il faut encore définir une fonction régissant l'évolution du paramètre de gain de chaque point de filtrage en jeu, en fonction de la distance à l'origine. Si l'on choisit une fonction linéaire telle que $f(x) = x$, lorsque la source est à 10% de la position maximale, les paramètres de gain de chaque point de filtrage seront à 10% de leur valeur maximale. Jouer avec cette courbe permet de modifier la perception de l'éloignement, et en utilisant d'autres types de courbes personnalisées, cela peut permettre par exemple d'accélérer la sensation d'éloignement lorsque la source arrive en fond de scène, et ainsi de forcer la perspective sonore, comme on le ferait avec des décors⁴.

Sans passer par un traitement externe, il est également possible de générer cet effet intéressant en termes de création. Il suffit de positionner les sources plus loin qu'elle ne sont réellement, dans l'interface du processeur. Mais si nous utilisons un système de tracking, alors il n'est pas possible de truquer la position manuellement.

Une première solution est de modifier le rapport de profondeur lors de la corrélation entre le système de localisation et le système de spatialisation. En donnant une échelle falsifiée pour l'axe dans la profondeur, l'on pourra faire en sorte que le processeur interprète à notre guise l'information de distance. Cette méthode est fixe et s'appliquera à tous les interprètes suivis, pendant tout le spectacle, alors qu'il est très probable que le concepteur préfère avoir la possibilité de jouer avec ce paramètre de manière plus fine.

Une deuxième solution s'offre à nous. Il faudra faire passer les positions transmises par le système de localisation par un programme permettant de truquer les positions détectées par le système avant de les transmettre au processeur. Comme proposé dans le paragraphe précédent, il est intéressant de définir des fonctions liant la position réelle à la position fictive. La fonction

4. La technique de perspective forcée est bien connue des décorateurs et scénographes de théâtre : elle consiste à accentuer l'illusion de profondeur des objets et des décors en modifiant leur forme normale pour leur donner de la perspective. Un décor peut ainsi sembler plus profond qu'il n'est en réalité.

$f(x) = x$ maintiendra la distance réelle, tandis qu'une fonction $f(x) = 2x$ doublera la distance. Il existe bien évidemment d'autres fonctions, laissées à l'imagination et aux besoins du concepteur.

6.1.2 Traitement auxiliaire : réverbération

Comme nous l'avons vu dans la section 1.1, la localisation est précisée par les informations que nous donne l'acoustique, en particulier les premières réflexions, et le ratio champ direct/champ réverbéré. Pour un réalisme accru, il faudrait reproduire l'acoustique de la salle dans laquelle le spectacle a lieu, mais ce n'est pas nécessairement souhaitable en particulier dans les lieux de représentation n'ayant pas une acoustique favorable. D'autre part, la réverbération est un élément créatif majeur dans le processus de mixage. Actuellement, dans le mixage « classique », les ingénieurs du son utilisent soit des réverbérations internes à la console de mixage, soit des réverbérations externes, en général choisies pour leurs qualités estimées supérieures à celles proposées par la console. Ces réverbérations correspondent au format de diffusion choisis permis par la console : en général stéréo (2.0) ou multicanal (5.0). Puisque nous fonctionnons avec des algorithmes faisant voler en éclat les formats classiques tels que la stéréo, les réverbérations doivent être également réadaptées.

Plusieurs tendances se distinguent actuellement : soit les réverbérations natives, algorithmiques ou à convolution, proposée directement au format souhaité de l'algorithme, soit l'utilisation de réverbérations externes, pas nécessairement adaptées au format de spatialisation sonore.

La question de la réverbération est longue et complexe, nous ne pouvons donc l'évoquer ici qu'en surface, en présentant les outils actuellement à disposition, et les solutions qui peuvent et sont mises en œuvre aujourd'hui.

Réverbération algorithmique « native »

À notre connaissance, cette solution est proposée par l'Ircam dans ses solutions *Panoramix* et *Spat Revolution*, par L-Acoustics avec *L-ISA*, et chez Ama-deus dans *Holophonix*. L'approche alors proposée est de n'avoir qu'un seul algorithme de réverbération que l'on peut façonner à sa guise, en jouant sur le filtrage et les différents paramètres (temps de réverbération global ou par bandes de fréquence, filtrage des réflexions précoces et du champ diffus, etc.).

Réverbération à convolution

A notre connaissance, seul le constructeur allemand d&b audiotechnik propose l'implémentation d'une telle solution à l'heure actuelle, dans son

processeur *Soundscape*. Les réponses impulsionnelles de plusieurs salles ont été enregistrées à de nombreux emplacements dans la salle, en excitant plusieurs points (trois sur scène, et un dans la salle). Les réponses impulsionnelles à utiliser sont choisies automatiquement par le processeur en fonction du nombre de haut-parleurs utilisés.

Réverbération externe

Lorsque le processeur n'intègre pas de système de réverbération, il faudra alors utiliser des réverbérations externes. D'autres raisons peuvent pousser à utiliser une réverbération externe : par exemple pour avoir plus de choix concernant sa sonorité, ou bien pour des facilités d'intégration⁵. Les canaux de sortie de la réverbération seront alors positionnés dans le processeur comme de sources fixes dans l'espace, ou pourront être assignés directement à des haut-parleurs selon l'effet recherché et les possibilités permises par le processeur. Pour les réverbérations externes actuellement disponibles sur le marché, il faut fournir l'audio au format de la réverbération (donc deux canaux pour les réverbérations stéréo, cinq pour les réverbérations surround, etc.). L'objectif étant de reproduire des informations spatiales, l'on ne peut se contenter d'envoyer l'audio de chaque source en mono à la réverbération, car il n'y aura alors aucune variation en fonction de la position de l'objet. Il faut donc alors définir un certain nombre de sorties audio vers lesquelles le son sera automatiquement spatialisé : autrement dit, dans la logique de ces processeurs, il faut définir des haut-parleurs, vers lesquels l'audio sera naturellement spatialisé. Ces sorties audio nourriront les différentes entrées audio des réverbérations. Les retours de réverbération seront positionnés dans l'espace, et affectés à l'algorithme de spatialisation sonore souhaité pour la sonorisation.

Par exemple, l'on pourrait utiliser deux moteurs de réverbération stéréo, dont les retours seront positionnés en fond de scène. L'on déclarera alors quatre haut-parleurs au processeur à l'emplacement de ces retours, et dont on affectera les sorties audio à l'entrée des moteurs de réverbération. L'on affectera ensuite les sources souhaitées au bus de spatialisation sonore comprenant ces enceintes. Si le processeur ne permet pas d'affiner quelles sources sont affectées à quelles enceintes, toutes les sources seront alors affectées par la réverbération.

5. Dans de nombreuses réverbérations externes, il est par exemple possible, comme dans les consoles, de mettre en mémoire les réglages choisis et de les rappeler le moment venu lors du spectacle. Effectuer la même chose avec un processeur peut s'avérer fastidieux s'il ne permet pas d'enregistrer des scènes en mémoire et de les rappeler instantanément (la plupart des processeurs ne le permettent pas actuellement). Il faudra enregistrer tous les paramètres de la réverbération (sans en omettre un seul) dans un logiciel externe tel que QLab, pour les rappeler le moment souhaité.

Dans une approche créative, l'on pourrait imaginer un système où l'on souhaite avoir plusieurs acoustiques à des emplacements différents sur scène. L'on définira alors autant de points de réverbération que souhaité. En fonction de son emplacement sur le plateau, l'acoustique changera. Un tel effet requerrait auparavant de modifier les envois de réverbération et de les mémoriser en fonction de la mise en scène. Ici, cette solution ne demande pas de mises en mémoire, favorisant donc des formes de spectacle plus libres, proches de l'improvisation.

6.2 Vers le mixage 100% objet

Nous l'avons vu à de nombreuses occasions au cours de ce mémoire, les processeurs de spatialisation sonores changent le paradigme du mixage. Dans ce nouvel univers, il peut être complexe d'y retrouver ses petits, et notamment les (bonnes) habitudes de mixage développées au cours du temps. Il est donc nécessaire d'observer quels sont les techniques de mixage actuellement employées et leurs objectifs. Il faut que les processeurs de spatialisation sonore offrent des outils aux ingénieurs du son qui leur permettent de maintenir un ensemble de fonctionnalités qui leur sont utiles, tout en conservant une logique de mixage objet. D'autre part, ces nouveaux outils ouvrent la voie à de nombreuses possibilités créatives, qui verront à leur tour naître des habitudes de mixage et des méthodes simplifiant le travail de régie. Si l'on peut proposer des outils sortis de vues de l'esprit, seule la pratique par différents concepteurs dans différents cadres de création permettra de générer un ensemble de pratiques et méthodes de mixage.

6.2.1 Habitudes de mixage à conserver

Le mixage des voix pour le spectacle vivant a amené à développer un ensemble de techniques très spécifiques, que l'on ne retrouve pas ou peu dans d'autres univers de mixage. Cet ensemble de procédures, d'habitudes et de fonctionnalités doit être préservé. Nous en dressons ici un inventaire (non exhaustif) à la Prévert des fonctionnalités principales, d'autres ont préalablement été évoquées dans le mémoire, nous ne les rappelons pas ici (mémoires de scènes et *snapshots*, réverbération, *etc.*).

Le rapport source/sonorisation

L'ingénieur du son mixe les niveaux sonores pour apporter l'équilibre le plus approprié entre le son émis sur scène et sa sonorisation. En comédie musicale, par exemple, ce rapport est très dynamique avec des musiques actuelles : lors des chansons, la sonorisation pourra se rapprocher d'une sono-

risation de concert, couvrant la voix naturelle des interprètes, tandis que la sonorisation prendra sûrement une place de « renfort sonore » très discrète lors de scènes parlées. Ce rapport source/sonorisation est un outil qui permet également de jouer avec l'attention du public. Dans un spectacle de théâtre, l'on pourra diminuer le son des voix pour favoriser le son direct des interprètes, et ainsi augmenter l'attention du public lors des scènes intimes ou fragiles par exemple.

La conséquence de ceci est que l'ingénieur du son, pour ces raisons, peut vouloir gérer le niveau de chaque comédien comme il le faisait auparavant sur une console de mixage. Nous discutons de ceci plus en détail dans la section Le fader.

La gestion du rapport source/sonorisation est également importante dans un autre cas de figure : lorsqu'un interprète projette suffisamment sa voix pour qu'elle soit audible de l'audience, sans nécessiter une amplification aussi forte que lorsqu'il chuchote. Ceci concerne en particulier les chanteurs d'opéra, qui peuvent projeter de forts niveaux sonores. Dans les systèmes de sonorisation classique, le système est optimisé pour être le plus homogène en niveau sonore sur toute la salle. Ainsi, en baissant le *fader* de niveau de l'interprète, il est simple d'ajuster le niveau de sonorisation afin d'éviter d'assourdir les spectateurs en ajoutant un niveau sonore trop fort au niveau déjà suffisant produit par l'interprète. Or, dans les systèmes que nous avons étudiés, on cherche à maintenir une spatialisation sonore cohérente dans toute la salle, ce qui implique un son différent en tout point de la salle. Sans entrer plus en détail dans les considérations de conception des systèmes, il est aisé de comprendre le problème que cela pose : lorsque l'interprète qui produit un fort niveau sonore est proche d'une enceinte, il n'est pas nécessaire de l'amplifier outre mesure à l'emplacement où il se trouve, mais il faut conserver l'amplification sur les autres enceintes, pour maintenir l'image sonore, et une bonne intelligibilité dans le reste de la salle, moins proche de l'interprète. Il faut donc un moins puissant renfort sonore à l'emplacement de l'interprète. Pierre Olivier Boulant dans son logiciel *DIY WFS* et Outboard UK avec *TiMax* prennent en compte cette problématique. Par exemple dans *DIY WFS*, une atténuation peut être automatiquement appliquée au haut-parleur le plus proche de l'interprète lorsque celui-ci se rapproche du système de sonorisation.

Le fader

Nous venons de le voir, dans certains cas, il est souhaitable de pouvoir contrôler le niveau de chaque source comme on le fait dans le mixage actuellement. Or les processeurs de spatialisation sonore ne proposent actuellement aucune surface de contrôle.

Dans la majorité des cas cela n'est pas nécessaire, les concepteurs sonores préférant utiliser une console en amont du processeur. Dans ce cas, la sortie de chaque tranche de console est affectée au processeur. La console n'assure plus la sommation des signaux, mais permet d'effectuer l'ensemble des traitements et d'enregistrer les scènes dans ses mémoires. Sa surface de contrôle permet alors d'utiliser ses *faders*. Mais certains concepteurs souhaitent aujourd'hui ne plus utiliser de console de mixage, préférant tout faire à l'aide du processeur. Il suffit donc d'avoir une surface de contrôle proposant *a minima* des *faders*. Pour l'instant, les constructeurs semblent préférer travailler à l'intégration avec les consoles de mixage : aucun ne propose de surface de contrôle. Néanmoins, les processeurs étant pilotables en OSC, il est possible de se construire de tels outils, mais il est regrettable qu'ils n'existent pas dans une forme prête à l'emploi.

Les groupes et DCA

Les VCA à l'origine sur les consoles de mixage analogique, sont des amplificateurs commandés en tension (*Voltage-Controlled Amplifier*). Ils servaient à contrôler plusieurs *fader* à l'aide d'un unique *fader*. Au passage aux consoles numériques la fonctionnalité a été conservée, et nommée DCA (*Digitally-Controlled Attenuator* pour respecter la logique numérique) ou bien plus simplement « groupes ». Il est ainsi possible de regrouper sous un unique *fader* la commande de niveau de plusieurs *faders*, tout en conservant leurs niveaux relatifs (donc leur balance).

Cette fonctionnalité est très utilisée dans la technique de mixage anglosaxonne appelée *line-by-line mixing*⁶ (mixage réplique à réplique) ou *line-crunching*⁷, et est documentée en détail dans l'ouvrage de Shannon Slaton (2019). Le principe est simple : face à un grand nombre de microphones à mixer, éparpillés sur la surface de la console, on utilise un nombre limité de DCA (8 à 12 en général) pour y regrouper les microphones à mixer. En fonction de la scène, l'on affecte aux DCA les tranches des interprètes présents sur scène, dans l'ordre facilitant le plus le mixage en fonction des répliques. On évitera par exemple d'espacer trop deux interprètes interagissant beaucoup l'un avec l'autre. Tout l'art de cette technique repose également dans la gestion des enchaînements d'une scène à l'autre afin de faciliter la transition. Un DCA peut piloter plusieurs *faders*, on les utilise donc également pour regrouper plusieurs interprètes lorsqu'ils sont en chœur, facilitant ainsi leur mixage avec le soliste. Les affectations aux DCA changent selon les scènes.

Dans les faits, il est très probable que les spectacles utilisant cette tech-

6. Slaton, Shannon, *Mixing a Musical*, 2^e édition. Routledge, 2019.

7. Fry, Gareth, *Sound Design for the Stage*, The Crowood Press, avril 2019.

nique de mixage continueront d'utiliser des consoles avec le processeur, mais intégrer aux processeurs un tel outil si standard dans les consoles de mixage serait intéressant pour les concepteurs souhaitant se passer d'une console, ce qui serait tout à fait envisageable avec les processeurs.

Double microphonie

Sur de nombreuses productions, il est très fréquent d'équiper les interprètes principaux avec deux microphones. Les microphones sont fragiles, et les défaillances sont fréquentes. Pour l'ensemble des interprètes, il suffit d'avoir un opérateur en coulisses capable d'effectuer des changements de microphone ou d'émetteur si nécessaire lors des sorties de scène. Le problème avec les rôles principaux est qu'ils sont souvent sur scène. Si un problème intervient sur leur microphone, il faut attendre longtemps avant leur sortie de scène, et ils peuvent ne pas avoir le temps d'effectuer un changement sur le micro en coulisses à cause de changements de costumes nécessaires, ou autres. On leur installe donc deux microphones, en se donnant la possibilité au niveau de la console de mixage de passer d'un microphone à l'autre instantanément, sans que cette opération soit audible du public.

Dans la plupart des consoles utilisées au théâtre, il est possible de définir deux entrées pour une tranche de console : entrée A et entrée B. Il suffit ensuite de programmer sur les boutons personnalisables de la console une fonction permettant de passer d'une entrée à l'autre. Aucun processeur ne permet pour le moment de définir deux entrées pour une source. Si l'on utilise un processeur sans console de mixage, il faut donc pouvoir reproduire cette fonctionnalité. Il faut donc définir deux sources pour chaque interprète possédant deux micros, et asservir leur position au même interprète, et enfin reproduire sur les deux sources les mêmes traitements, et penser à le faire au moment où l'on effectue des changements sur l'une. Avec une commande OSC, l'on pourra définir une source sur *mute* et l'autre non, avec une commande permettant d'inverser le réglage. Si l'on utilise des DCA pour le mixage avec une surface de contrôle connectée au processeur, il faudra intégrer à chaque fois les deux sources au même DCA. S'il est possible de changer avec une commande les affectations (*routing*) d'entrée, et ce de manière fiable et sécurisée, cette solution sera préférée, car il n'est alors plus nécessaire de définir deux sources.

Il est évident qu'il serait beaucoup plus simple que les constructeurs intègrent cette possibilité de définir une entrée A et une entrée B pour chaque source au sein de leur processeur, mais dans les faits la plupart des productions continuent d'utiliser des consoles de mixage, qui peuvent assurer ces fonctions.

Traitements sonores

Lorsque l'on sonorise des voix, l'on utilise un ensemble de traitements, pouvant inclure : des filtres (coupe-bas, égalisateurs), des traitements de la dynamique (compresseur), et des techniques hybrides (dé-esseur). Lorsque l'on utilise une console en amont du processeur, tous ces réglages sont disponibles. Lorsque l'on n'utilise pas de console, il faut s'assurer que ces traitements soient possibles, ces possibilités varient selon le constructeur du processeur. Ces traitements peuvent également être externalisés : ils sont parfois possibles au sein des convertisseurs analogiques-numériques par exemple.

Il est également fréquent d'effectuer des traitements sur des bus. L'on retrouve fréquemment des bus spécifiques regroupant les voix d'hommes et de femme par exemple. On y applique alors des traitements plus génériques (généralement tenant plutôt du filtrage). Même si l'on utilise une console, il n'est plus possible de le faire de la même manière si l'on utilise un processeur de spatialisation sonore. En effet, le traitement s'effectue sur le bus, qui est même format que le système de diffusion (stéréo, ou mono s'il est destiné au point central d'un système L/C/R). Dans ce bus, l'ensemble des sources sont sommées, elles ne ressortent plus de manière indépendante, or nous avons besoin de canaux distincts pour alimenter les sources dans le processeur de spatialisation. Actuellement, à notre connaissance, il n'existe pas de solution à ce problème, seule la solution logicielle *Panoramix* de l'Ircam offrant la fonctionnalité d'appliquer des traitements sur les bus de spatialisation sonore.

Si les constructeurs de consoles souhaitent rester indispensables même avec l'utilisation de processeurs, il faut qu'ils modifient le fonctionnement de leurs consoles pour permettre une telle fonctionnalité ; si les processeurs veulent pouvoir se faire valoir en tant qu'outil suffisant de mixage, il faut proposer des traitements sur les bus de spatialisation sonore, et donner la possibilité d'avoir plusieurs bus différenciés (c'est par exemple le cas de *Holophonix*).

6.2.2 Possibilité offertes par le mixage objet

Le mixage objet ouvre également la porte à de nouvelles possibilités, dont les limites ne sont que les fonctionnalités offertes par le processeur utilisé. Chaque paramètre des processeurs étant généralement pilotable en OSC, le concepteur peut à sa guise prendre le contrôle des fonctionnalités offertes pour inventer de nouveaux usages et de nouvelles possibilités.

Retournement de l'acteur

L'une des fonctionnalités les plus intéressantes pour aboutir le mixage réaliste des voix est de prendre en compte le retournement de l'acteur, lorsqu'il s'oriente dos au public. Actuellement, à notre connaissance, il n'existe pas de processeur intégrant cette fonctionnalité, pourtant les systèmes de localisation proposent tous cette donnée. Néanmoins, la plupart des paramètres permettant de jouer avec sont actuellement à disposition dans certains processeurs ou logiciels (en particulier *Holophonix* de Amadeus, *Panoramix* et *Spat Revolution* de l'Ircam, et *Soundscape* de d&b audiotechnik).

Lorsqu'un acteur est de dos, il se passe plusieurs phénomènes. Le rapport son direct/son réverbéré change à la faveur du son réverbéré, puisqu'il se retourne vers le mur du fond, et le son direct se retrouve filtré. La contribution du champ réverbéré peut être facilement contrôlée sur les processeurs, notamment via le paramètre de largeur des premières réflexions. Il est actuellement difficile de modifier le rapport champ direct/champ réverbéré puisque ces deux paramètres sont automatisés en fonction de la position du comédien, et il n'existe pas de possibilité d'affecter ces paramètres de manière relative (c'est-à-dire de leur ajouter un différentiel sur leur valeur automatique), l'intégration de ceci dépendra donc des possibilités et des impératifs imposés par le matériel choisi. Concernant le filtrage, il est possible de procéder de manière analogue à l'éloignement.

Création sonore

Sur le plan de la création sonore le mixage objet offre de nombreuses possibilités, qui sont souvent facilitées par l'accès à la position de l'acteur sur scène. L'on peut ainsi imaginer des interactions entre la position de l'acteur et la composition musicale, en particulier dans le cadre de la musique contemporaine, qui fonctionne souvent de manière paramétrique, où l'on peut aisément imaginer ce type d'interactions.

Nous avons également évoqué le jeu avec les espaces acoustiques via la réverbération, mais beaucoup d'autres choses peuvent être inventées, ce que racontera l'histoire, et les volontés du concepteur sonore, du metteur en scène, du compositeur, et dorénavant pourquoi pas du chorégraphe, du concepteur lumières.

Chapitre 7

Description de la partie pratique du mémoire

La partie pratique du mémoire aura lieu au Théâtre National de Chaillot. Depuis le début de l'année 2019, ce théâtre est équipé du processeur de spatialisation *Holophonix* de Amadeus, et du système de localisation *Stagetacker II* de TTA. Utiliser cette installation sera l'occasion de comprendre les opportunités de développement de la voix dans l'espace.

Les observations faites lors de cette partie pratique seront présentées lors de la soutenance, à l'aide notamment d'un petit système de spatialisation sonore (mais sans tracking) et seront ajoutées au mémoire en annexe.



FIGURE 7.1 – La ligne d'enceintes de rappels au Théâtre National de Chaillot

L'équipement disponible à Chaillot

L'installation disponible au Théâtre National de Chaillot est unique en France. Si de plus en plus de théâtre et salles de concert s'équipent de systèmes de spatialisation sonore, il est le seul à posséder aujourd'hui un sys-

tème de localisation des interprètes sur scène. De plus, le théâtre était auparavant équipé du processeur *Wave I* de Sonic Emotion.

Le système de sonorisation se compose d'un banc de douze enceintes en haut du cadre de scène, d'un bandeau niché dans le nez de scène opérant en *front fill* et d'une ligne de rappel composée également de douze enceintes. Une série d'enceintes opérant en surround sont également disponibles sur les côtés.

Le processeur *Holophonix* propose d'exploiter de nombreux algorithmes (WFS, Ambisonie HOA, VBAP, LBAP, KNN, *etc.*) et offre des possibilités de configuration souple. Il intègre une réverbération native dans tous les formats de spatialisation sonore, permettant de contrôler distinctement les premières réflexions et le champ diffus.

Le système de localisation repose sur l'utilisation de deux capteurs positionnés de part et d'autre du cadre de scène. Ils permettent de localiser les balises portées par les comédiens sur l'ensemble du plateau et presque la totalité de l'avant-scène. Leur calibration permet d'atteindre une précision de ± 5 cm dans la localisation.

Les observations que nous effectuerons

Les observations nous effectuerons seront faites de manière empirique en nous positionnant en tant que gens de métier. Il n'en demeure pas moins qu'elle seront faites avec rigueur.

- ▷ Le volume de la salle et les dimensions du plateau sont conséquents. Nos essais viseront à observer les effets de bords qui ocurrent pour les positions les plus extrêmes du public, et ce qu'il en est de la qualité de la localisation des sons pour ces positions.
- ▷ Puisque nous avons à disposition plusieurs algorithmes, nous effectuerons des comparaisons entre ceux-ci. Nous nous intéresserons en particulier à la WFS, la HOA et un algorithme de panning d'amplitude comme le VBAP.
- ▷ Le processeur *Holophonix* offre une grande variété de paramètres pour ces algorithmes. Nous chercherons à caractériser leur utilisation en faisant varier ces différents paramètres, entre autres :
 - Evaluer la gêne que peut représenter le filtrage WFS en modulant d'un essai à l'autre leur contribution dans l'algorithme (c'est un paramètre offert par le processeur)
 - Evaluer l'apport du *blur* et du *spread* dans les Algorithmes HOA et VBAP. Ces paramètres permettent de produire une image sonore moins nette, surement favorable à l'auditeur. En effet, des sources localisées

trop précisément sur le banc de haut-parleurs peuvent encore maintenir une sensation de délocalisation

- ▷ Nous évaluerons les possibilités de jouer sur l'effet de profondeur, en développant un micrologiciel permettant de truquer la position de l'acteur en profondeur.
- ▷ Nous évaluerons les possibilités de jouer sur l'effet de retournement de l'acteur, en développant un micrologiciel permettant de modifier le ratio champ direct/premières réflexions et d'appliquer un filtrage.
- ▷ Nous testerons enfin différents paramètres dans le système de localisation, en particulier les filtrages permettant de moyenner la position des acteurs, afin de trouver lesquels sont les plus adéquats face à des mouvements rapides et lents.

Conclusion

Nous l'avons vu dans l'introduction de ce mémoire, le principal grief qui est fait à la sonorisation de la voix est sa délocalisation. Les acteurs se meuvent sur scène, mais leur voix reste bloquée dans un ou deux haut-parleurs, à une position fixe en haut du cadre de scène. Ces systèmes standards de sonorisation de la voix ont été traités dans la première partie du mémoire, afin d'offrir un panorama des techniques généralement mises en œuvre, et permettre au lecteur qui découvre ce champ d'application de la sonorisation de se familiariser avec des systèmes présents sur de nombreux spectacles. En parallèle de ces techniques de sonorisation s'en sont développées d'autres, nommées alors outre-manche *Source Oriented Reinforcement*, dont le but était de maintenir une localisation sonore des voix des interprètes cohérente avec leurs positions sur scène. À ses débuts, ces techniques pouvaient être presque aussi variées que les spectacles eux-mêmes, mais de grandes tendances se dessinaient. Nous avons donc saisi l'opportunité de les présenter au lecteur. Ces techniques, bien qu'anciennes, n'en sont pas moins pertinentes encore aujourd'hui : lorsque l'on ne dispose pas des moyens de mettre en œuvre des techniques plus modernes, il est toujours possible d'utiliser celles-ci, puisqu'elles demandent généralement peu de matériel, ou du matériel aujourd'hui standard et peu coûteux.

Le lecteur aura ensuite pu constater par la chronologie que nous avons effectuée que les processeurs de spatialisation sonore se sont multipliés au cours des dernières années, offrant des solutions plus simples à mettre en œuvre pour spatialiser les voix. Ce mouvement a également été accompagné au cours des six dernières années du perfectionnement des outils de localisation d'interprètes pour le spectacle vivant. En couplant les deux, il devient possible d'obtenir un outil performant, permettant de spatialiser les voix en accord avec la position des acteurs sur scène. Bien que *TiMax* ait proposé dans le courant des années 2000 un système similaire, cette conjecture est réjouissante pour de telles applications, puisqu'elle offre un plus grand choix dans la conception du système de sonorisation des voix, là où une solution unique, certes intéressante, était l'unique possibilité avant.

C'est cette utilisation conjointe d'un processeur de spatialisation sonore

et d'un système de localisation qui a guidé notre étude tout au long de la deuxième partie du mémoire.

Nous avons relevé les différentes problématiques de l'acquisition des positions des interprètes sur scènes de manière manuelle ou préenregistrée, qui montrent l'intérêt que revêt l'utilisation d'un système de localisation automatisée. Les principes de fonctionnement de ces systèmes ont ensuite été présentés au lecteur, afin d'appréhender au mieux leur installation le cas échéant. Le même procédé a ensuite été utilisé pour présenter les algorithmes de spatialisation sonore utilisés dans les processeurs de spatialisation, dans la même logique de favoriser les questions pratiques avant les explications scientifiques. Une attention particulière a été portée au rapport au mouvement de ces algorithmes, et nous espérons que le lecteur pourra trouver dans ces chapitres des éléments lui permettant de porter son choix vers l'un ou l'autre des algorithmes proposés, selon les nécessités du spectacle qu'il prépare.

Enfin nous nous sommes interrogés sur les outils de mixage proposés dans le cadre de ce type de sonorisation, et porté plusieurs observations concernant l'ergonomie de mixage. Il en ressort que l'utilisation la plus simple aujourd'hui est d'utiliser une console de mixage en amont du processeur, ce qui est la solution choisie par les sonorisateurs dans la quasi-intégralité des cas aujourd'hui. Il est possible d'envisager utiliser les processeurs seuls, sans console, mais l'on fera attention à bien avoir tous les outils nécessaires. En outre ce type d'utilisation n'est pas possible avec tous les processeurs, ceux-ci devant intégrer un minimum de traitements de source, et d'outils de contrôle du niveau par exemple.

Dans le cadre de la partie pratique liée à ce mémoire, nous organiserons une série d'essais sur la sonorisation de la voix au Théâtre National de Chaillot, en utilisant leur processeur *Holophonix* et leur système de localisation *TTA Stagetracker II*. Ces essais auront pour but de dire si notre préférence personnelle se porte en faveur d'un algorithme plutôt qu'un, tout en relevant les différences que nous aurons pu remarquer. Nous évaluerons aussi l'optimisation des paramètres de corrélation entre le système de localisation et le processeur. Ces conclusions personnelles seront présentées à l'occasion de la présentation orale du mémoire, avec un petit démonstrateur de spatialisation sonore installé dans les murs de l'école pour illustrer nos propos.

Ces systèmes de sonorisation offrent aujourd'hui une solution relativement simple pour spatialiser les voix, et relativement complète. S'agissant d'une méthode rompant avec les systèmes habituels L/C/R et L/R, il en ressort que certains sujets méritent encore que l'on s'y intéresse, mais sont en bonne voie d'être résolus dans quelques courtes années. Il y a entre autres la question de la réverbération, où l'on doit pouvoir proposer plus de choix et

plus d'outils au sonorisateur. Se pose également la question de l'optimisation des systèmes de sonorisation (appelé « calage système »), qui doit être étudiée de manière approfondie pour des applications au spectacle vivant.

Bien que nous présentions ici des solutions qui peuvent sembler par certains aspects « figés », il ne faut pas oublier que la conception d'une sonorisation doit toujours rester libre et dynamique. C'est la chance qu'offre le spectacle vivant : chaque nouveau spectacle présente ses spécificités et ses défis, auxquels le sonorisateur a l'opportunité fantastique de répondre de manière créative. En somme, ces outils de spatialisation sonore ne sont pas une finalité absolue, il ne faut pas s'empêcher d'avoir recours à d'autres procédés de multidiffusion bien connus au théâtre (enceintes dans le décor, en coulisses, en hauteur), qui viendront compléter le système principal. D'autre part la spatialisation sonore peut être au besoin envisagée de manière dynamique : est-il nécessaire, dans une comédie musicale de musiques actuelles, de spatialiser la voix lors du solo d'un chanteur, sonorisé de manière proche d'un concert ? Le sonorisateur peut effectuer ce choix en analysant l'œuvre et la mise en scène qui en est produite. La spatialisation sonore (aujourd'hui performante et plus simple à mettre en œuvre) se présente finalement comme un outil parmi d'autres à la disposition du sonorisateur.

Bibliographie

- Blacktrax*, Site web \langle URL: <http://blacktrax.cast-soft.com> \rangle – visité le 2019-05-20.
- Eliko KIO RTLS*, Site web \langle URL: <https://www.eliko.ee/products/kio-rtls/> \rangle – visité le 2019-05-20.
- TiMax Tracker Outboard*, Site web \langle URL: <http://www.outboard.co.uk/timax.html#timax-tracker> \rangle – visité le 2019-05-20.
- TTA Stagetracker II*, Site web \langle URL: <https://www.tta-sound.com/> \rangle – visité le 2019-05-20.
- ASTRO-SPATIAL-AUDIO, *Sara II object rendering engine*, site web \langle URL: <http://www.ravepubs.com/astro-spatial-audio-debut-sara-ii-object-rendering-engine-ise-2017/> \rangle – visité le 2019-11-26.
- BARCO, Communiqué de presse, *Barco gains the team and expertise from IOSONO GmbH*, \langle URL: <https://www.barco.com/en/News/Press-releases/Barco-gains-the-team-and-expertise-from-IOSONO-GmbH.aspx> \rangle – visité le 2019-05-16.
- BLAUERT, Jens, *Spatial hearing : the psychophysics of human sound localization*, Revised édition. Cambridge: MIT Press, 1997.
- BOREN, Braxton, "History of 3D Sound", *Immersive Sound* 2018.
- BOTTE, Marie-Claire *et al.*, *Psychoacoustique et perception auditive*, Cachan [France] : Paris: EMinter; INSERM/SFA/CNET, 1989.
- BOULANT, Pierre-Olivier, *Wave Field Synthesis DIY*, site web \langle URL: <https://wfs-diy.net> \rangle – visité le 2019-05-15.
- BREEBAART, Jeroen et FALLER, Christof, *Spatial Audio Processing : MPEG Surround and Other Application*, John Wiley Sons, 2007.
- BUKVIC, Ivica Ivo, "3D Time-Based Aural Data Representation Using D4 Library's Layer Based Amplitude Panning Algorithm", *Proceedings of the 22nd International Conference on Auditory Display - ICAD 2016*, juillet 2016, p. 36–45.
- BURSTON, Jonathan, "La technologie audio, le corps chantant reconfiguré et les megamusicals", *Théâtre/Public* 2011, Nr. 199.
- CAULKINS, Terence, *Caractérisation et contrôle du rayonnement d'un système de Wave Field Synthesis pour la situation de concert*. Thèse de Doctorat Université Pierre et Marie Curie - Paris 6, Paris, 2007.

- COCHRANE, Marty et KURIAKOSE, Robbins, *Real-Time Tracking Protocol*, Site web \langle URL: <https://rttrp.github.io/RTrP-Wiki/index.html> \rangle – visité le 2019-05-15.
- DANIEL, Jérôme, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Thèse de doctorat, Université Pierre et Marie Curie, Paris 6, Paris, 2001.
- DESHAYS, Daniel, "L'impossible façade : l'usage du microphone et son incidence sur la scénographie des espaces", *Le son du théâtre : XIXe - XXIe siècle* 2016.
- FISCHETTI, Antonio, *Initiation à l'acoustique*, Paris: Belin, 2003.
- FREED, Adrian et SCHMEDER, Andy, Features and Future of Open Sound Control version 1.1 for NIME,, p. 5.
- FRY, Gareth, *Sound Design for the Stage*, The Crowood Press, avril 2019.
- GRANT, Mark N., *The Rise and Fall of the Broadway Musical*, Boston: Northeastern University Press, 2004.
- HILL, Adam J. *et al.*, "Live event performer tracking for digital console automation using industry-standard wireless microphone systems", *Audio Engineering Society*, 2013, p. 13.
- IDMT, *Spatialsound wave*, site web \langle URL: <https://www.idmt.fraunhofer.de/en/institute/projects-products/spatial-sound-wave.html#tabpanel1-3> \rangle – visité le 2019-11-26.
- JOSEPHS, Andrew et PENNY, Ryan, "Back to basics : multiple sound sources", *The Echo, the ASD magazine (Association of Sound Designers)* septembre 2016, Nr. 12.
- LARRUE, Jean-Marc, "Son, présence et ontologie", *Le son du théâtre : XIXe - XXIe siècle* 2016.
- LARRUE, Jean-Marc et MERVANT-ROUX, Marie-Madeleine, *Le son du théâtre : I. Le passé audible, Théâtre/Public* 197 [2010].
- Larrue, Jean-Marc et Mervant-Roux, Marie-Madeleine (éd.)**, *Le son du théâtre XIXe - XXIe siècle : histoire intermédiaire d'en lieu d'écoute moderne*, Paris: CNRS éditions, 2016.
- LARRUE, Jean-Marc, MERVANT-ROUX, Marie-Madeleine et GUINEBAULT-SLAMOWICZ, Chantal, *Le son du théâtre : II. Dire l'acoustique, Théâtre/Public* 199 [2011].
- LEONARD, John A., *Theater Sound*, A C Black, 2001.
- LOSSIUS, Trond, BALTAZAR, Pascal et HOGUE, Théo de la, "DBAP - Distance-Based Amplitude Panning", 2011, p. 4.
- MERCIER, Denis et CASTELLENGO, Michèle, *Le livre des techniques du son, tome 1 : Notions fondamentales*, 4^e édition. 2010.
- NICOL, Rozen, "Sound Field", *Immersive Sound* 2018.
- PULKKI, Ville, "Virtual Sound Source Positioning Using Vector Vase Amplitude Panning", *Journal of Audio Engineering Society* Vol. 45 June 1997, Nr. 6.

- RUMSEY, Francis, *Spatial Audio*, Focal Press, 2001.
- SLATON, Shannon, *Mixing a Musical*, 2^e édition. Routledge, 2019.
- SPORER, Thomas *et al.*, "Wave Field Synthesis", *Immersive Sound* 2018.
- START, Evert Walter, *Direct sound enhancement by wave field synthesis*, Thèse de doctorat, Delft University of Technology, Delft, Netherlands, 1997.
- TAGG, Philip, "Analysing popular music : theory, method and practice", *Popular Music* 1982, Nr. 2.
- WANÈGUE, Jean-José, "Révolution de palais dans la Cour d'Honneur : La WFS s'invite à Avignon", *Réalisation* octobre 2010, Nr. 62.
- WIKIPEDIA, Contributeur de, *Angle of arrival*, janvier 2019, Page Version ID : 881149505 (URL: https://en.wikipedia.org/w/index.php?title=Angle_of_arrival&oldid=881149505) – visité le 2019-05-15.
- WIKIPEDIA, Contributeurs de, *Multilateration*, février 2019, Page Version ID : 882224853 (URL: <https://en.wikipedia.org/w/index.php?title=Multilateration&oldid=882224853>) – visité le 2019-05-14.
- WIKIPEDIA, Contributeurs de, *Triangulation (en)*, avril 2019, Page Version ID : 894171239 (URL: <https://en.wikipedia.org/w/index.php?title=Triangulation&oldid=894171239>) – visité le 2019-05-14.
- WIKIPEDIA, Contributeurs de, *True range multilateration*, mai 2019, Page Version ID : 895924871 (URL: https://en.wikipedia.org/w/index.php?title=True_range_multilateration&oldid=895924871) – visité le 2019-05-14.
- WRIGHT, Matt, *The Open Sound Control 1.0 Specification*, Site web, mars 2002 (URL: http://opensoundcontrol.org/spec-1_0) – visité le 2019-05-14.
- ZOTTER, Franz et FRANK, Matthias, "All-Round Ambisonic Panning and Decoding", *Journal of Audio Engineering Society* 60 2012, Nr. 10.
- ZOTTER, Franz, POMBERGER, Hannes et NOISTERNIG, Markus, "Ambisonic decoding with and without mode matching : a case study using the hemisphere", *Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics* 2010.