



Proposition d'un nouvel outil de spatialisation sonore pour le cinéma

Louis Crétel

Mémoire de Master
Spécialité Son

Directeur interne :	Jean-Loup Pecquais
Directeur externe :	Cyril Holtz
Référent académique universitaire :	Laurent Millot
Coordinateur des mémoires :	Corsin Vogel
Examineur :	Alan Blum

Promotion 2025

Résumé

Ce mémoire de recherche est né d'une interrogation sur la spatialisation sonore au cinéma. Nous avons déjà toutes et tous ressentis une instabilité de la scène sonore qui penche d'un côté ou de l'autre selon notre position dans la salle de cinéma. Cet effet est dû à la présence unique des différences d'intensités dans nos outils de spatialisation. Les différences de temps, impactant tout autant notre système perceptif, sont utilisées dans certaines techniques de spatialisation et présentent de nombreux avantages dans les domaines tels que la création sonore ou la sonorisation du spectacle vivant, mais sont très peu utilisées au cinéma.

Notre étude se base ainsi sur la réalisation d'un outil de spatialisation qui permettrait d'allier les différences d'intensités et de temps, qui sont deux mécanismes perceptifs majeurs du système auditif humain. Ce travail de recherche vise donc à saisir l'importance de ces mécanismes dans le système auditif humain, en évaluant si l'outil développé permet d'explorer une nouvelle approche de spatialisation. Afin de répondre à ces questions, nous avons réalisé deux tests : un premier test perceptif afin de comprendre l'utilité et l'intérêt de ces mécanismes dans notre perception en les comparant individuellement ou ensemble, puis un second test pratique qui nous permet d'évaluer la pertinence de cette nouvelle approche de spatialisation auprès de professionnels.

Les résultats montrent un meilleur ancrage de la source dans l'espace sonore et une facilité à la localiser lorsque des différences de temps et d'intensité sont combinées. Toutefois, chaque spatialisation induite par un mécanisme individuel s'est révélée intéressante pour produire certains effets sonores spécifiques.

Mots clés : Spatialisation, localisation, différence de temps, panoramique, perception, post-production, système multicanal, différence interaurale, différence interaurale de temps, arbre multicanaux

Abstract

The idea for this research arose from a question about sound spatialization in cinema. We have all experienced the instability of a sound scene that leans to one side or the other depending on our position in the cinema. This effect is due to the unique presence of intensity differences in our spatialization tools. Time differences, which have just as much impact on our perceptive system, are used in certain spatialization techniques and offer numerous advantages in fields such as sound creation or live performance sound reinforcement, but are scarcely used in cinema.

Our study is therefore based on the development of a spatialization tool that would make it possible to combine differences in intensity and time, which are two major perceptive mechanisms of the human auditory system. The aim of this research is to understand the importance of these mechanisms in the human auditory system, by assessing whether the tool developed can be used to explore a new approach to spatialization. In order to answer these questions, we carried out two tests: a first perceptual test, which enabled us to understand the usefulness and interest of these mechanisms in our perception by comparing them individually or together. Secondly, we carried out a practical test to assess the relevance of this new approach to spatialization with professionals.

The results show a better anchoring of the source in sound space and a better localization of the source when differences in time and intensity are combined. However, each spatialization induced by an individual mechanism proved interesting for the aesthetic effects sought.

Keywords : Spatialization, localisation, time difference, panoramic, perception, post-production, multichannel system, interaural difference, ITD, microphone array design

Remerciements

Je remercie chaleureusement Cyril Holtz et Jean-Loup Pecquais, mes directeurs de mémoire, pour leur accompagnement sans faille, leur disponibilité, leur engagement et la richesse de leurs conseils tout au long de ce travail. Je tiens également à remercier Corsin Vogel et Armand Schwarz de m’avoir guidé dans la méthodologie.

Merci à Michael Williams pour son accueil bienveillant, pour cette belle matinée d’échanges et le bon repas partagé autour de ce projet.

Merci à Antoine Martin pour ses conseils à la fois pertinents et précieux concernant la réalisation de l’outil.

Merci à mes amis et à mes camarades de la promotion son 2025 qui ont su me soutenir et trouver les mots justes dans les moments de doute, pour leur bonne humeur et compagnie plus que rassurante.

Un immense merci à ma famille, qui ne saisit pas tous les détails de ce mémoire, mais sans laquelle je n’aurais pas pu avancer aussi sereinement et avec certitude, grâce à leur soutien indéfectible.

Merci à toutes les personnes qui ont participé aux différents tests et prêté leurs oreilles affûtées.

Merci à Polyson pour leur accueil chaleureux et pour m’avoir permis de mener une partie des tests dans leurs locaux.

Merci à Anaëlle Dreyer et Thibaud Carcy pour leur aide attentive et précieuse dans la mise en forme des résultats.

Table des matières

Résumé	2
Abstract	3
Remerciements	4
Introduction	8
1 La spatialisation sonore au cinéma	10
1.1 Évolution des systèmes d'écoute au cinéma	10
1.1.1 Premier système multicanal	11
1.1.2 Dolby Stéréo	11
1.1.3 Expansion du multicanal	13
1.1.3.1 <i>Dolby Digital</i>	13
1.1.3.2 <i>DTS</i>	13
1.1.3.3 <i>Sony Dynamic Digital Sound</i>	14
1.1.3.4 <i>Dolby Surround 7.1</i>	14
1.1.4 Mixage orienté objet	15
1.2 Approche physiologique	16
1.2.1 Différence interaurale d'intensité	17
1.2.2 Différence interaurale de temps	18
1.2.3 Différence interaurale de phase	20
1.2.4 Différence spectrale	21
1.3 Variétés d'approches pour créer des scènes sonores multicanales	21
1.3.1 Création d'un espace multicanal dès la prise de son	22
1.3.1.1 <i>Préambule sur la directivité des microphones</i>	22
1.3.1.2 <i>Arbres multicanaux</i>	23
1.3.2 Création d'un espace multicanal en postproduction	30
1.3.2.1 <i>Techniques de spatialisation</i>	30
1.3.2.2 <i>Quelques lois de panoramique</i>	31
1.3.2.3 <i>Upmixage</i>	33
1.4 Systèmes utilisant des différences temporelles	33
1.4.1 <i>Wave Field Synthesis</i>	33

1.4.2	ViMiC	35
1.5	Perception spatiale en salle	37
1.5.1	Configuration de la salle et variabilité inter-salles	37
1.5.2	Le <i>sweet-spot</i>	41
1.5.3	L'utilisation des différences de temps en complément de celles d'intensités	42
2	Conception de l'outil	44
2.1	Présentation générale de l'outil	44
2.1.1	Idée de l'outil	44
2.1.2	Cahier des charges	45
2.1.3	Décomposition en blocs fonctionnels	45
2.2	Calcul de la distance et de l'angulation source/microphone	46
2.3	Conception du module de modélisation microphonique	47
2.4	Une palette variée de configurations	48
2.4.1	Arbre multicanal à sept capsules omnidirectionnelles	48
2.4.2	Prédominance des différences de temps sur les enceintes frontales	50
2.4.3	Prédominance des différences de temps sur les enceintes " <i>surround</i> "	54
2.5	Réflexion autour d'une interface graphique	57
2.5.1	Quel contrôle souhaite-t-on accorder à l'utilisateur ?	57
2.5.2	Présentation des réglages	57
2.5.2.1	<i>Mécanisme perceptif</i>	57
2.5.2.2	<i>Configuration de l'arbre</i>	58
2.5.3	Exploration de l'apparence de l'interface	58
3	Méthodologie pour un protocole de tests expérimentaux	61
3.1	Test perceptif : évaluation de l'impact des mécanismes perceptifs à travers un système multicanal	61
3.1.1	Paramètre et perception spatiale	61
3.1.1.1	<i>Indices de localisation</i>	61
3.1.1.2	<i>Localization et Locatedness</i>	62
3.1.1.3	<i>Choix de la configuration de l'outil</i>	63
3.1.1.4	<i>Sources sonores</i>	64
3.1.2	Mise en place du test	66
3.1.2.1	<i>Environnement de l'expérience</i>	66
3.1.2.2	<i>Répartition des sources et mécanismes perceptifs par session</i>	66
3.1.2.3	<i>Positionnement des sources dans l'outil</i>	68
3.1.3	Protocole de test	69

3.1.3.1	<i>Choix des participants, participantes et répartition par groupe</i>	69
3.1.3.2	<i>Réalisation du questionnaire</i>	69
3.2	Test pratique : Utilisation de l’outil par des professionnels du son	70
3.2.1	Mise en place du test	70
3.2.1.1	<i>Environnements de l’expérience</i>	70
3.2.1.2	<i>Préparation de notre interface graphique</i>	70
3.2.2	Protocole de test	71
3.2.2.1	<i>Présentation de nos participants et participantes</i>	71
3.2.2.2	<i>Réalisation du questionnaire</i>	72
4	Exploitation des résultats	73
4.1	Test perceptif	73
4.1.1	Statistiques de l’échantillon	73
4.1.2	Analyse par questions	76
4.1.2.1	<i>Q1 : Localisation</i>	76
4.1.2.2	<i>Q2 : Direction</i>	80
4.1.2.3	<i>Q3 : Étalement</i>	82
4.1.2.4	<i>Q4 : Préférence</i>	85
4.1.3	Discussion autour des analyses	86
4.2	Test pratique	88
4.2.1	Analyse par critères	88
4.2.1.1	<i>Ergonomie</i>	88
4.2.1.2	<i>Perception</i>	88
4.2.1.3	<i>Remarques générales et suggestions d’améliorations</i>	90
4.2.2	Discussions autour des analyses	91
	Conclusion	92
	Bibliographie	95
	Table des Annexes	102
	Annexe A : Frise des systèmes sonores au cinéma, PÉRIAUX <i>et al.</i> , 2015	103
	Annexe B : Abaques de Michael Williams	104
	Annexe C : Arbres microphoniques supplémentaires	108
	Annexe D : Lois de Panoramiques, “SPAT Revolution - Panning Algorithms”, 2024	115
	Annexe E : Questionnaires des différents tests	118
	Annexe F : Pouvoir séparateur selon différents plans, BLAUERT, 1996	128

Introduction

La spatialisation sonore est essentielle dans la sensation d’immersion d’un public, particulièrement en salle de cinéma. De nos jours, l’expérience sensorielle d’une spectatrice ou d’un spectateur repose sur des technologies avancées de l’image et du son. Les vidéoprojecteurs sont de plus en plus performants et le nombre de canaux de nos systèmes de diffusion sonore ne cesse de croître. Grâce à ces évolutions, il est possible de fabriquer des scènes sonores de plus en plus élaborées afin de soutenir une idée esthétique, d’appuyer ou de nuancer une idée exprimée à l’image.

Cependant, la spatialisation peut gêner le spectateur ou la spectatrice, par exemple à cause de l’*exit effect*. Cet effet est défini par la perception d’un son qui semble provenir du dehors de la salle, voir à l’opposé de la direction de l’écran (vers le panneau de sortie), créant ainsi une sensation de rupture ou de distanciation par rapport à la scène sonore frontale (HOLMAN, 2008). Cet effet peut être voulu pour désorienter l’auditeur ou l’auditrice mais, dans la plupart des cas, il résulte d’un défaut de spatialisation.

Nous avons été frappés par un autre défaut de spatialisation lors du visionnage du film *Il reste encore demain* de la réalisatrice Paola Cortellesi sorti en 2023. Nous nous trouvions bien au centre de la salle sans être trop proche ou trop loin de l’écran, notre position était “idéale”, néanmoins nous avons été désorientés par l’instabilité de certaines sources sonores. Un simple basculement de tête d’un côté ou de l’autre faisait basculer la position de certaines sources du côté correspondant. Cette instabilité était due à un fort niveau sonore des sources dans les canaux gauche et droit du système, mais aussi à la présence unique des différences d’intensité comme mécanisme de localisation dans nos outils de spatialisation.

Cette recherche s’appuie sur l’exploration et la proposition d’un nouvel outil qui combinerait plusieurs mécanismes de localisation. Nous chercherons ainsi à saisir l’importance des différents mécanismes perceptifs (différence de temps, d’intensité, de phase et spectrale) et à en implémenter quelques-uns dans l’architecture de notre outil. Il se

basera sur la simulation d'un système de captation sonore qui permet de moduler l'action de certains mécanismes selon la configuration spatiale choisie. Grâce à ses configurations multiples, il en découlera plusieurs types de spatialisation.

L'idée de cet outil est d'intégrer à son architecture des différences temporelles en plus de celles d'intensités. Des systèmes qui s'appuient sur des différences de temps, afin de spatialiser des sources sonores existent déjà et permettent une meilleure reproduction d'une image sonore, c'est le cas de la WFS (*Wave Field Synthesis*). Ce système, fondé sur la reconstruction des caractéristiques physiques du champ sonore, présente de nombreux avantages comme la reproduction fidèle d'un champ sonore et l'élargissement de la zone d'écoute.

Ainsi, posons-nous la question suivante : que pourrait apporter un outil de spatialisation qui combinerait différences de temps et d'intensité à une scène sonore au cinéma ? La perception de l'espace en serait-elle impactée ?

La première partie de notre étude présentera la mise en commun d'écrits abordant la spatialisation d'un point de vue historique, perceptif et technique qui nous permettra de comprendre au mieux les enjeux et possibilités des systèmes d'aujourd'hui et de mener au mieux la conceptualisation de notre outil. Ensuite, nous détaillerons dans une deuxième partie le cheminement et l'élaboration de l'outil en proposant plusieurs configurations. Nos deux dernières parties interrogeront l'importance des mécanismes perceptifs dans la perception et leur utilité dans l'outil créé lors de cette recherche. Ces expériences prendront deux formes distinctes : d'une part, des sessions d'écoute où plusieurs sujets seront invités à comparer différentes restitutions spatiales et, de l'autre, des rencontres auprès de professionnels afin de comprendre la pertinence d'un tel outil dans nos manières de travailler et l'apport que celui-ci pourrait amener. La préparation de ces expériences sera décrite dans la troisième partie et les résultats dans la dernière.

Partie I

La spatialisation sonore au cinéma

Un rapide historique du son spatialisé au cinéma nous amènera à comprendre l'évolution des systèmes d'antan à ceux que nous utilisons aujourd'hui. Ensuite, nous chercherons à comprendre les différents mécanismes qui permettent à l'humain de localiser des sons dans un environnement quelconque en abordant par la suite l'élaboration et la création de scènes sonores multicanales. Pour finir, nous présenterons quelques systèmes multicanaux qui utilisent les différences de temps dans leur conception, puis nous discuterons de la perception spatiale au cinéma.

1.1 Évolution des systèmes d'écoute au cinéma

■ *C'était en octobre 1987 lors de la réunion SMPTE. Les gens se demandaient "combien de canaux doit-il y avoir pour le format standard du son au cinéma ?" Certaines personnes disaient deux...d'autres quatre...un autre disait huit. J'ai levé ma main et dit, "cinq point un". Tout le monde se demandait, "Mais de quoi parle-t-il ?" ■*

– Tomlinson Holman, ingénieur du son et développeur du système THX, (KERINS, 2011).

La question de la spatialisation au cinéma a toujours été omniprésente. Cependant, les décisions qui ont pu être prises tout au long de son histoire privilégiaient un but commercial et prônaient la facilité plutôt que la qualité d'un système en lui-même, comme le démontre notre citation¹. Avec *The Jazz Singer* d'Alan Crosland, sorti en 1927 (considéré comme le premier film parlant bien qu'il n'inclue que quelques scènes de dialogue synchronisé), est né le vitaphone. Ce système a vu le jour grâce aux frères Warner et lisait la piste sonore du film sur un disque de dix minutes, lu par le phonographe d'Edison, en parfaite synchronisation avec l'appareil de projection.

¹Le système "5.1" est en effet une configuration assez particulière qui limite les choix de spatialisation que nous développerons plus tard dans la partie 1.5.1.

1.1.1 Premier système multicanal

L'un des premiers systèmes multicanaux discrets² était le Fantasound qui a vu le jour lors de la sortie du film *Fantasia* en 1940. Ce système multicanal développé par RCA et Walt Disney Studios utilisait deux nouvelles avancées techniques : premièrement, il n'y avait pas seulement une ou deux enceintes derrière l'écran de projection mais trois, avec chacune une piste sonore distincte. La deuxième avancée était l'utilisation d'un seul canal "surround" à travers quatre-vingt-seize haut-parleurs. Ce canal "surround" était utilisé principalement pour des effets sonores particuliers (KERINS, 2011).

Le système Fantasound n'a pas connu de succès commercial à cause d'un système d'enregistrement d'orchestre trop complexe et d'un mixage laborieux. De ce fait, il a vite été supplanté par d'autres systèmes comme ceux de *Dolby* qui ont, eux, plutôt axé leurs produits sur la portabilité et la flexibilité au détriment de la qualité sonore.

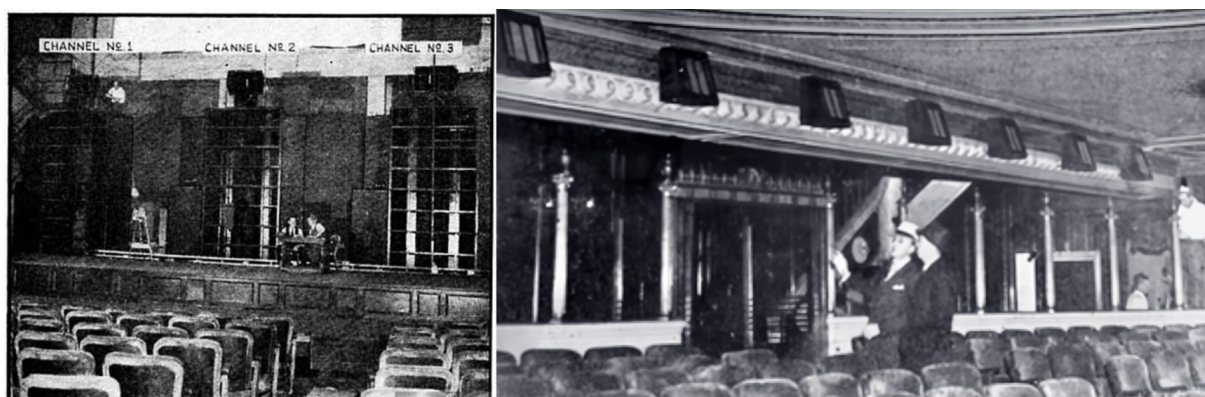


Figure 1.1 : Photos du système Fantasound installé en salle, www.pmamagazine.org/fr/comment-disneys-fantasound-a-introduit-le-son-surround-a-hollywood-en-1940/

1.1.2 Dolby Stéréo

Contrairement au Fantasound, le système Dolby Stéréo fut un succès commercial international. Ce système n'est pas un système dit stéréophonique, c'est-à-dire un système discret à deux canaux retranscrit sur quatre haut-parleurs. À l'aide de deux canaux audios LtRt, il se décode à partir d'une matrice 4-2-4³ afin d'obtenir quatre canaux LCRS pour être retranscrit sur un système à plus de deux haut-parleurs (PÉRIAUX *et al.*, 2015). Ce qui le définit comme un système matriciel. Le système Dolby Stéréo présentait donc trois canaux

²Dispositif qui comprend plusieurs canaux audio distincts retranscrits respectivement sur le nombre de haut-parleurs correspondants

³Le terme 4-2-4 correspond à la chaîne de codage et de décodage. Nous avons à l'origine 4 canaux que nous codons sur 2 canaux afin de restituer les 4 canaux lors du décodage.

frontaux : gauche, centre, droit respectivement retranscrits sur trois haut-parleurs et un canal dit “*surround*” retranscrit sur plusieurs haut-parleurs.

Une des forces de ce système résidait dans le fait que la bande son était inscrite directement en optique sur la pellicule même du film (Figure 1.2), ce qui le rendait portable et ne nécessitait pas de périphériques externes comme le Fantasound et d’autres systèmes. Sa réponse fréquentielle et dynamique était meilleure que celle des systèmes concurrents de cette époque grâce aux avancées du Dolby SR en 1980 qui offrait une meilleure réduction de bruit. Cependant, beaucoup de professionnels se plaignaient d’une spatialisation sonore limitée et imprécise dû au matriçage, et que la dynamique sonore soit restreinte en raison de sa compression optique. (KERINS, 2011) Le système a été utilisé pour la première fois sur *La Guerre des Étoiles* de Georges Lucas en 1977. Plusieurs spectateurs et spectatrices se sont dits impressionnés par la spatialisation de certains effets sonores.

■ *En effet, il est difficile d’évoquer le Dolby Stéréo sans faire référence à La Guerre des étoiles, sorti en 1977. La scène d’ouverture, dans laquelle un vaisseau spatial géant se déplace de l’arrière de la salle à l’avant d’un point de vue sonore, a mis en évidence les puissantes capacités du nouveau format audio et a propulsé le Dolby Stéréo dans la conscience du public.* ■

– KERINS (2011).

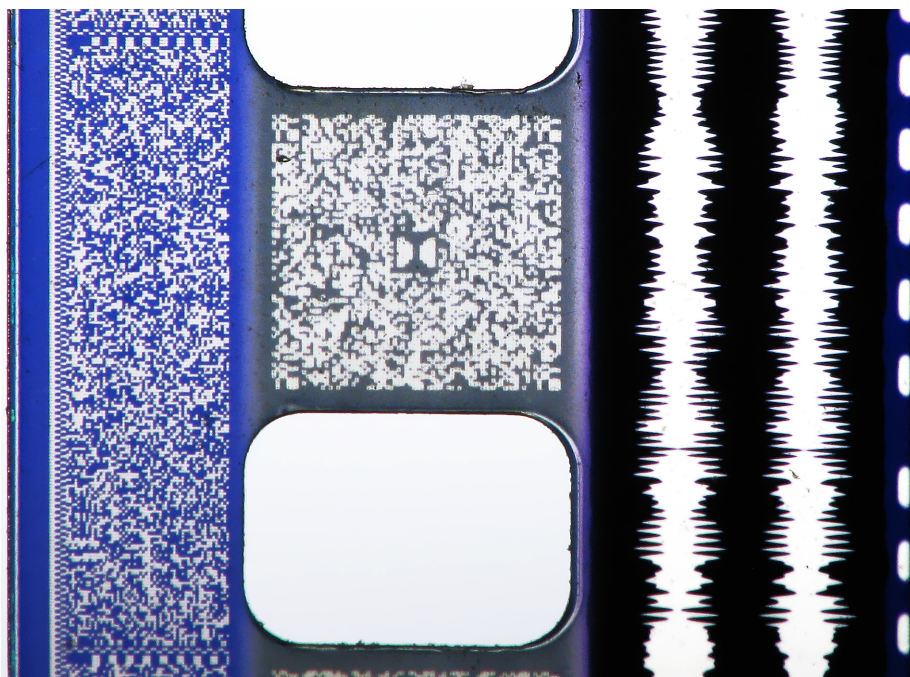


Figure 1.2 : Une photo agrandie d’une pellicule 35 mm, quatre formats sont représentés respectivement de gauche à droite : le SDDS de Sony, le Dolby Digital, le Dolby Stéréo et les informations de *timecode* pour DTS. (PECQUAIS, s. d.)

1.1.3 Expansion du multicanal

Ayant connaissance des limitations du Dolby Stéréo et des possibilités que pouvait offrir le numérique, l'industrie du cinéma était en pleine réflexion sur le format standard à appliquer en salle. Lors de la réunion SMPTE de 1987, deux décisions ont été prises. Une première obligeait le système à être discret et non pas matriciel à cause des désavantages de spatialisation engendrée. Une seconde contraignait la qualité sonore à être au minimum une qualité dite CD⁴ afin de rattraper les performances des systèmes domestiques proposant ces caractéristiques. Il ne manquait plus qu'à se mettre d'accord sur le nombre de canaux. Cité en début de partie, le système "5.1" proposé par Holman a été adopté après discussions. Celui-ci se base sur le système matriciel *Quintaphonics* (1975) qui utilisait déjà cette configuration et qui aurait pu s'imposer, mais vite supplanté par le Dolby Stéréo. *Cinema Digital Sound* (CDS) sera le premier système en 1990 à répondre aux attentes de cette réunion. L'idée était de récupérer le marché financier et de concurrencer Dolby. Cependant, ce système n'aura pas fait long feu : lors de sa deuxième utilisation en salle, le processeur qui décodait la bande sonore aurait pris feu et enflammé la salle de projection. (KERINS, 2011)

1.1.3.1 Dolby Digital

Utilisé pour la première fois sur *Batman Returns* en 1992, c'est le deuxième système qui adopte le format dit standard, introduit en 1987. Dolby Digital (anciennement *Spectral Recording Digital* (SRD)) ne s'implantera pas si rapidement à cause de son coût d'installation. Très peu de salles en étaient équipées et ce système laissera place à la concurrence comme Digital Theater Systems (DTS) ou Sony. (KERINS, 2011)

1.1.3.2 DTS

Un an après la sortie du Dolby Digital, Digital Theater System voit le jour avec la sortie de *Jurassic Park* de Steven Spielberg (1993). DTS a pu être installé dans cent fois plus de salles que son concurrent direct Dolby Digital grâce à deux avantages. Premièrement, le coût d'installation d'un système DTS était quatre fois moins cher que celui de Dolby, deuxièmement, DTS avait le soutien d'un grand studio, Universal, et d'un réalisateur populaire, Spielberg. Le réalisateur n'avait pas simplement convaincu Universal d'utiliser et de financer DTS sur *Jurassic Park*, mais aussi sur la totalité de ses prochains films. Dolby a pu démontrer la faisabilité d'un système multicanal numérique, mais c'est DTS qui a su se démarquer auprès du public. (KERINS, 2011)

⁴Échantillonnage à 44 100 Hz et 16 bits de quantification

1.1.3.3 Sony Dynamic Digital Sound

Une semaine plus tard, la sortie de *Jurassic Park*, un troisième système entre dans l'arène du marché des systèmes multicanaux numériques : le Sony Dynamic Digital Sound (SDDS) avec la sortie de *Last Action Hero* (1993). Avec son nouveau système, Sony a choisi de placer cinq enceintes frontales ainsi que deux “surround” afin de réaliser un système “7.1” (Figure 1.3) et non “5.1” requis au minimum par la SMPTE adoptée par Dolby et DTS. À l'époque, cette décision pouvait sembler farfelue, surtout quand on pense à la taille relativement petite des écrans de projection. Trois enceintes frontales étaient amplement suffisantes pour remplir l'espace sonore de ces écrans. Même si ce système n'a pas eu le même impact ni la même popularité que ses deux rivaux de l'époque, il a tout de même réussi à se faire une place en tant que concurrent viable dans la décennie qui a suivi. (KERINS, 2011)

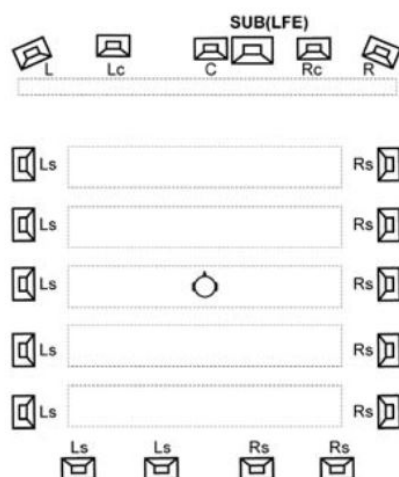


Figure 1.3 : *Format 7.1 SDDS cinéma* (PÉRIAUX et al., 2015)

Un problème persistait : comment une salle de cinéma équipée en DTS pourrait-elle diffuser un film mixé en Dolby Digital ? Plusieurs studios se sont donc mis d'accord afin d'encoder et d'inclure les différentes bandes sonores de chaque système sur la pellicule (Figure 1.2) afin que toute salle de cinéma puisse diffuser au bon format.

1.1.3.4 Dolby Surround 7.1

En 2010, Dolby a lancé son nouveau système à huit canaux, le Dolby Surround 7.1, lors de la sortie du film d'animation *Toy Story 3*. Ce système inclut trois canaux frontaux, quatre canaux “surround” et un canal réservé aux basses fréquences (Figure 1.4). L'objectif de ce nouveau système est d'amener davantage de précision spatiale dans les canaux arrière.

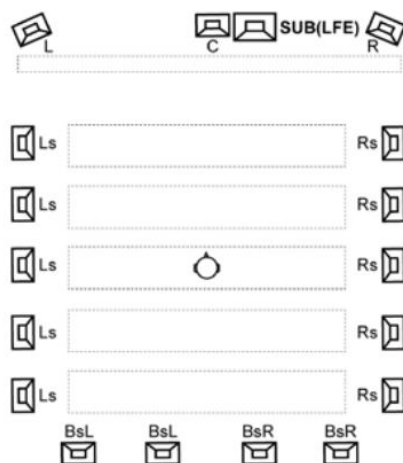


Figure 1.4 : *Format 7.1 ITU⁵ cinéma* (PÉRIAUX et al., 2015)

1.1.4 Mixage orienté objet

Une autre manière de penser les systèmes multicanaux digitaux est apparue en 2012 sur le film *Brave* de Brenda Chapman et Mark Andrews. Le mixage orienté objet cherche à transposer un espace sonore fictif vers un système physique réel. Les informations de spatialisation des sources sonores dans l'espace fictif sont inscrites dans un flux de données et retranscrites sur une configuration de haut-parleurs connus. Un processeur est donc responsable de déterminer la meilleure manière de reproduire le positionnement des sources dans la configuration définie. (PÉRIAUX et al., 2015) Lorsque Dolby a dévoilé le Dolby Atmos qui utilise cette technologie, il l'a qualifié de “son 3D”, terme qui définit une nouvelle dimension perceptive : celle de l'élévation.

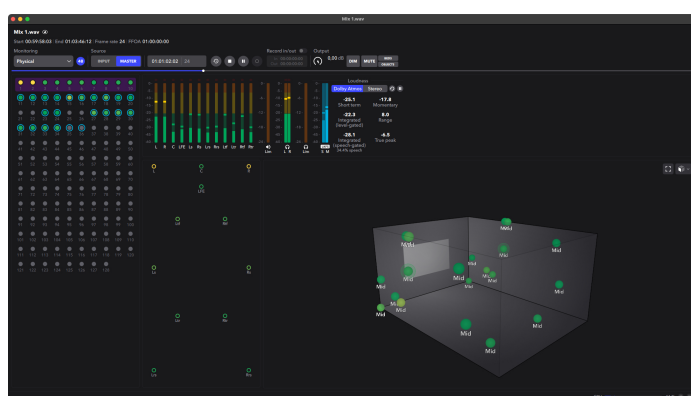


Figure 1.5 : *Logiciel “Dolby renderer”,*
<https://professional.dolby.com/product/dolby-atmos-content-creation/dolby-atmos-renderer/>

Le “Dolby renderer⁶” (Figure 1.5) permet aux mixeurs et mixeuses de visualiser le positionnement des sources dans l'espace fictif qui sera retranscrit sur un système physique que nous pouvons observer en Figure 1.6.

⁵Union internationale des télécommunications : organisation internationale jouant un rôle clé dans la normalisation des technologies de communication

⁶Moteur de rendu Dolby Atmos

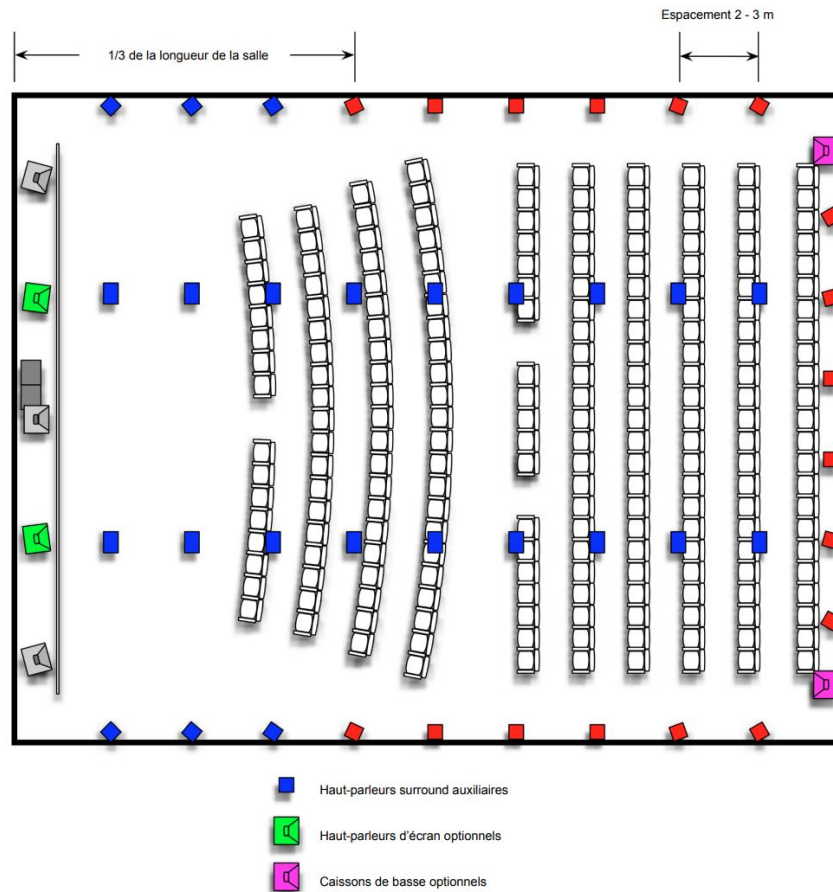


Figure 1.6 : Représentation d'une salle équipée en Dolby Atmos
(DOLBY LABORATORIES, 2012)

Une frise chronologique référençant toutes les évolutions techniques du cinéma est disponible en Annexe A : Frise des systèmes sonores au cinéma, PÉRIAUX *et al.* (2015).

1.2 Approche physiologique

Dans cette partie, nous allons explorer le fonctionnement de l'oreille humaine. En nous appuyant sur diverses recherches et publications, nous tenterons de comprendre comment un individu se repère dans son environnement.

De nombreux écrits expliquent que notre capacité à localiser des sons se décompose en deux paramètres distincts qui sont chacun prédominant dans une zone fréquentielle déterminée : il s'agit à basse fréquence de la différence de temps et à haute fréquence de la différence d'intensité [Stevens et Newman, 1936]. Ce modèle reste assez proche de notre capacité à localiser des sons, cependant, d'autres paramètres entrent en jeu. À vrai dire, nous traiterons quatre mécanismes de localisation.

1.2.1 Différence interaurale d'intensité

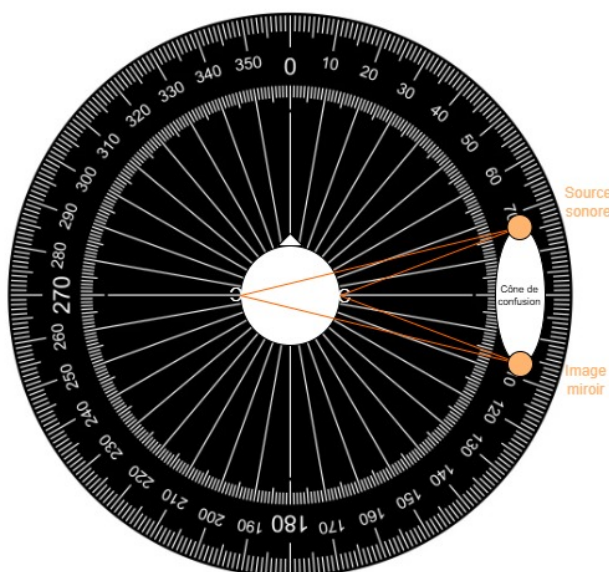


Figure 1.7 : Schéma d'une tête vue de dessus dans le plan azimutal avec la représentation d'un cône de confusion (figure de l'auteur)

Considérons une source sonore dans un environnement. Nos deux oreilles captent les variations de pression créées par cette source avec des intensités différentes, mis à part dans certains cas particuliers. En effet, si notre source est localisée à 0° ou 180° , notre système perceptif ne détecte aucune différence d'intensité entre nos deux oreilles, autrement appelée différence interaurale d'intensité (IID). Il existe également d'autres positions où ces différences sont identiques, ce sont sur les cônes de confusion (Figure 1.7). Toute source possède une image “miroir” par rapport à l'axe de symétrie formé par la droite reliant nos oreilles, et c'est dans ces conditions que nous avons affaire à des difficultés dans la localisation de sources (PULKKI *et al.*, 1999).

Ces différences d'intensités sonores sont dues à notre morphologie ; taille et forme de nos têtes et oreilles. La tête constitue un obstacle majeur pour les fréquences aiguës alors que les basses fréquences ne sont pas influencées par la morphologie de l'être humain. D'après la famille Letowski, “*nous supposons généralement que l'effet de diffraction d'une tête humaine moyenne devient négligeable en dessous de 1000 Hz et qu'à des fréquences inférieures à 1500 Hz, l'IID est trop faible pour faciliter la localisation du son*”, car les ondes acoustiques se réfléchissent sur des obstacles dont la taille est proche de la demi-longueur d'onde. (LETOWSKI et LETOWSKI, 2012).

1.2.2 Différence intéraurale de temps

Le signal émis par la source sonore se propage à une certaine vitesse et parvient à notre système auditif après un certain temps. La relation reliant ces paramètres est la suivante : $temps = \frac{distance}{célérité}$. Le temps que met l'onde acoustique à parvenir à l'une de nos oreilles est généralement différent de celui arrivant à l'autre oreille. La différence de ces temps est appelée : différence intéraurale de temps (ITD).

À partir du modèle de Woodworth (1938), nous pouvons écrire que la différence de marche Δd suit la formule simplifiée suivante où r représente le rayon d'une tête et θ l'azimut d'une source à l'infini :

$$\Delta d = r(\theta + \sin\theta) \quad (1)$$

Nous pouvons alors en déduire la différence intéraurale de temps :

$$\Delta t = \frac{r}{c}(\theta + \sin\theta) \quad (2)$$

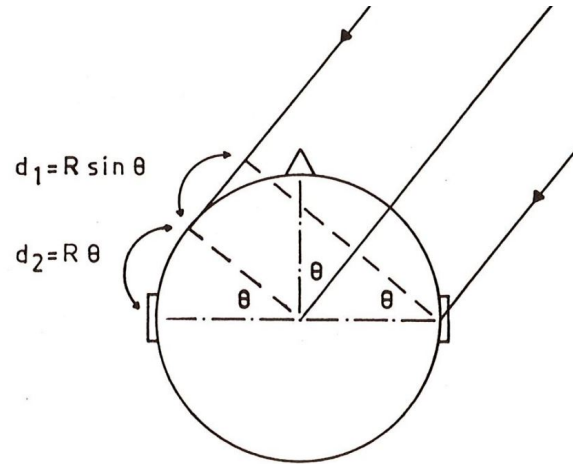


Figure 1.8 : Évaluation de la différence de marche Δd aux oreilles pour une source à l'infini d'azimut θ , Woodworth, 1938 (BOTTE et al., 1988)

Cette différence intéraurale de temps oscille entre 0 et 794 μs , cette dernière valeur étant sa valeur critique. Elle est valable pour un sujet dont le rayon de tête est de 9 cm et pour une température de 15 °C (LETOWSKI et LETOWSKI, 2012). Cette valeur critique est modulée par la température de l'environnement (T en degrés Celsius) qui affecte la célérité du son dans l'air (c en $m.s^{-1}$) à partir de la formule suivante ; $c = 331 + 0.6T$.

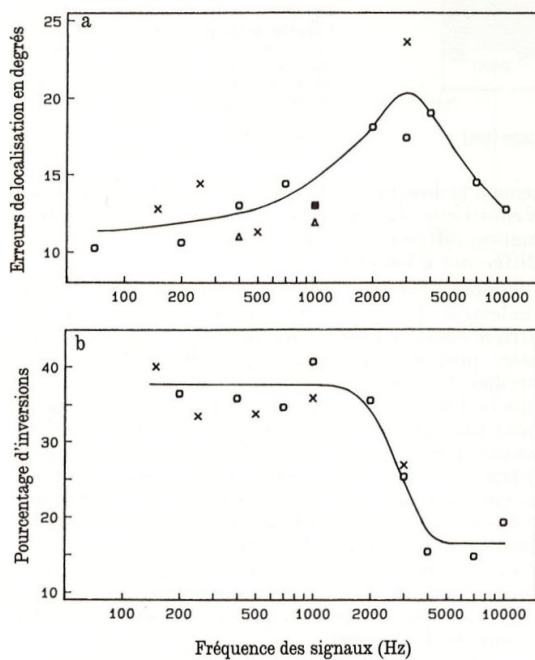


Figure 1.9 : *Limites de la précision d'identification en azimuth (Stevens et Newman, 1936), d'après BOTTE et al. (1988)*

*a - Erreurs de la localisation en fonction de la fréquence,
b - Pourcentage des confusions avant-arrière en fonction de la fréquence.*

Cette différence a un effet prépondérant dans les basses fréquences contrairement à l'IID. Les résultats des expériences menées par Stevens et Newman (Figure 1.9a), en plein air sur les toits des laboratoires de biologie de l'université d'Harvard, montrent clairement que les erreurs de localisation sont presque constantes à basse fréquence et que, vers 3000 Hz, elles atteignent une importance croissante pour ensuite décroître. Les auteurs de l'étude étaient à la fois observateurs et expérimentateurs et écoutaient principalement chacun leur tour des sons purs. Ces résultats sont en accord avec nos considérations sur les effets du temps et de l'intensité sur la localisation. "Par conséquent, en regroupant les zones fréquentielles, où peuvent intervenir soit la phase, soit l'intensité, on découvre une sorte de "zone morte", vers 3000 Hz, où ni l'une ni l'autre n'apportent d'informations suffisantes pour une localisation précise." (BOTTE et al., 1988)

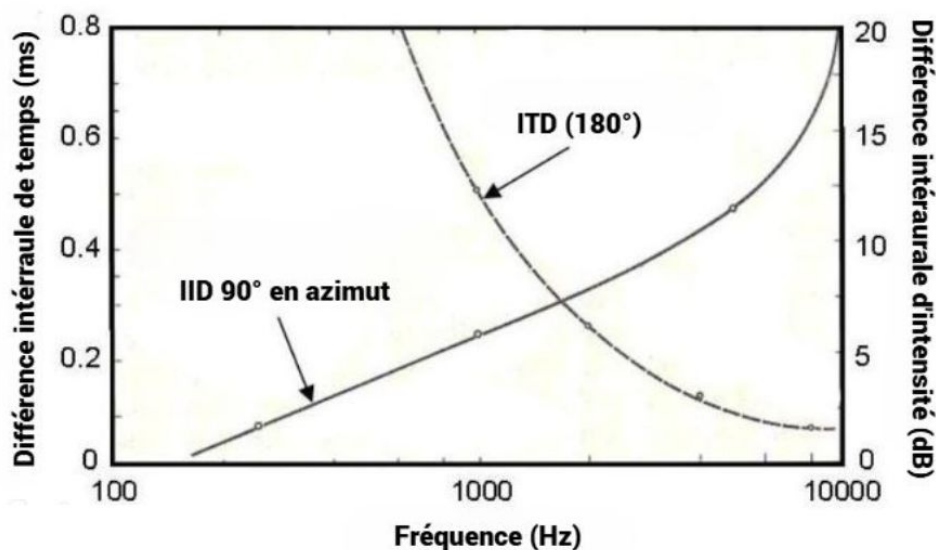


Figure 1.10 : *Effet de l'ITD et l'IID en fonction de la fréquence (LETOWSKI et LETOWSKI, 2012)*

1.2.3 Différence intéraurale de phase

La différence intéraurale de phase n'est pas à confondre avec celle de temps. Pour un signal sinusoïdal de fréquence f , la phase correspond à son décalage temporel rapporté à la période, et s'exprime en degrés. Un retard temporel τ se traduit alors par une différence de phase $\phi = 2\pi f\tau$.

Pour un signal large bande, la phase devient une grandeur locale, dépendante de la fréquence ; on parle alors de phase $\phi(f)$ et de retard de groupe lorsqu'on considère un retard global du signal.

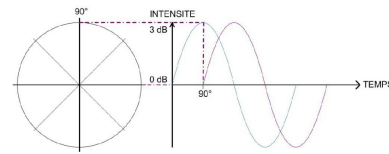


Figure 1.11 : *Représentation de la phase*

En fonction du rapport de phase, les interactions entre deux signaux de même amplitude peuvent être destructives ou constructives. Si nos signaux sont hors phase (décalage de 180°), la sommation de nos signaux est destructive : nous n'avons aucun signal en sortie. Si au contraire nos signaux sont en phase (décalage de 0° ou de 360°), la sommation est constructive et un gain de trois décibels est ajouté à notre signal de sortie. Notons que pour des signaux périodiques dont la longueur d'onde est inférieure ou égale à 17 cm (distance intéraurale humaine), la localisation d'une telle source est imprécise. En effet, au-dessus de 1500 Hz, notre système perceptif n'arrive plus à localiser des sources sonores à l'aide de la différence intéraurale de phase (IPD), car à haute fréquence, les longueurs d'ondes correspondantes sont plus courtes que la distance intéraurale humaine, ce qui entraîne une incertitude dans l'IPD.

Très peu d'études mentionnent l'IPD comme étant un mécanisme perceptif, à part celles de Koray Ozcan. Cet auteur a pu comparer ces trois paramètres séparément dans *The Significance of Phase as an Auditory Cue* (OZCAN *et al.*, 2003) et voici les résultats :

- L'IPD aurait un effet conséquent sur des signaux tonaux, mais un effet plus faible sur des signaux impulsifs tels que des transitoires. Cependant, l'ITD aurait un effet significatif sur les deux types de signaux.
- L'IPD et l'ITD se compensent mutuellement dans notre système perceptif. C'est-à-dire qu'un décalage de phase peut très bien être compensé par un décalage de temps.
- Un signal stéréophonique dont l'un de ses canaux est en avance de phase sera perçu comme provenant du côté de ce canal.

1.2.4 Différence spectrale

Les *Head Related Transfer Functions* (HRTF) correspondent à un ensemble de filtres appliqué “naturellement” au signal à certaines fréquences en fonction de la diffraction de l’onde sonore sur notre tête, y compris nos pavillons et les réflexions sur notre corps comme les épaules et le torse. Elles se traduisent par un ensemble de filtres spécifiques par rapport à l’angle de provenance de la source et désignent un ensemble de fonctions de transfert qui, à partir de la position d’une source sonore par rapport à un auditeur ou auditrice, permettent au cerveau d’interpréter les indices spectraux associés (PÉRIAUX *et al.*, 2015).

■ *Le pavillon de l’oreille fonctionne comme une cavité résonnante, il ajoute des colorations spectrales variables en fonction de la direction de la source, appelées indices spectraux. Il semble être l’élément clef dans la caractérisation de la fonction HRTF.* ■

– PÉRIAUX *et al.* (2015).

Ces phénomènes sont propres à chaque individu dû aux différentes morphologies possibles. Cette fonction peut être décrite de deux manières, soit par une courbe de réponse fréquentielle (HRTF), soit par une réponse impulsionnelle (HRIR⁷), son équivalent dans le domaine temporel. Trois variables rentrent donc en jeu : la fréquence ou le temps, la position de la source sonore en coordonnées sphériques (azimut, angle d’élévation) et la morphologie de l’individu. Ces indices spectraux sont perçus comme des indications de direction plutôt que des variations de timbre (BLAUERT, 1996).

1.3 Variétés d’approches pour créer des scènes sonores multicanales

Au cinéma, le domaine de l’audio est contraint par un élément majeur de cet art : l’image. Afin de recréer des espaces sonores qui seraient cohérents ou non à l’écran, le monde de l’audiovisuel a dû s’adapter et innover. Plusieurs types de sources sonores proviennent des différentes étapes de la conception d’un film : le tournage, le montage son, la session de bruitage et le mixage. À partir de tous ces matériaux, un espace sonore doit être monté de toute pièce. Nous nous intéresserons seulement aux prises de son multicanales s’appliquant parfaitement à notre domaine d’étude et ensuite aux outils disponibles de nos jours pour recréer des espaces sonores multicanaux.

⁷ *Head Related Impulse Response*

1.3.1 Création d'un espace multicanal dès la prise de son

La prise de son est un élément majeur dans le domaine de l'audio, c'est la première étape de la création sonore. À partir d'un transducteur, le microphone, nous obtenons un signal électrique issu d'une onde acoustique stockée sur un support de stockage.

1.3.1.1 Préambule sur la directivité des microphones

Deux types de microphones existent, ceux à pression et ceux à gradient de pression. Le microphone à pression est également appelé microphone omnidirectionnel et le microphone à gradient de pression, microphone directionnel. Les microphones directifs sont en réalité une combinaison de deux directivités. Il existe plusieurs directivités possibles, Michael Williams a pu se pencher sur la question et nous propose plusieurs nominations : omnidirectionnel, hypocardiöide, cardiöide, supercardiöide, hypercardiöide et bidirectionnel (WILLIAMS, 2004a).

L'équation polaire γ d'un microphone omnidirectionnel est égale à 1 alors que celle d'un bidirectionnel est égale à $\cos(\theta)$. À partir de l'article suivant PETERS *et al.*, 2008, nous reprenons l'équation suivante :

$$\gamma = a + b * \cos(\theta) \quad \text{avec } 0 \leq a, b \leq 1 \quad (3)$$

Voici un tableau référençant les différentes directivités pour diverses valeurs de a et b ainsi que leurs diagrammes polaires respectifs :

Directivité	a	b
Omnidirectionnel	1.0	0.0
Subcardiöide	0.7	0.3
Cardiöide	0.5	0.5
Supercardiöide	0.33	0.67
Hypercardiöide	0.3	0.7
Bidirectionnel	0.0	1.0

Nous pouvons remarquer qu'il y existe une dépendance entre les variables a et b :

$$\begin{cases} b = 1 - a \\ \gamma = a + (1 - a) \cos \theta \\ 0 \leq a \leq 1 \end{cases} \quad (4)$$

Cette réflexion nous intéressera lors de la réalisation de notre outil.

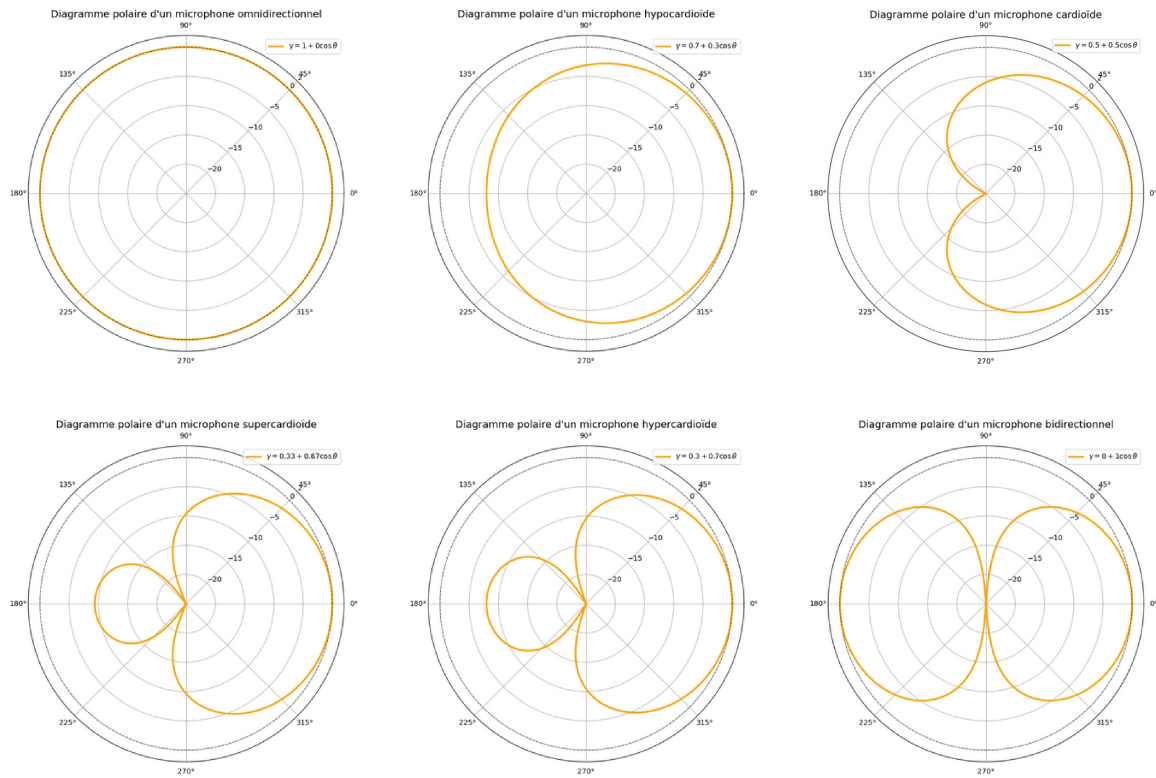


Figure 1.12 : Diagrammes polaires des différents modèles théoriques de directivité microphonique obtenues à l'aide de la librairie matplotlib de Python

1.3.1.2 Arbres multicanaux

Une manière de capter l'environnement est de rassembler plusieurs microphones afin de le retranscrire sur un système à plusieurs haut-parleurs. Le preneur ou la preneuse de son a un panel de choix à sa disposition lors de la conception de son arbre multicanal afin d'élaborer l'espace multicanal qu'il ou elle souhaite retranscrire.

De nos jours, deux types de systèmes existent afin de capter notre environnement : les systèmes matricés et non matricés. Le double MS et le système ambisonique sont deux systèmes matricés qui nécessitent un décodage après la prise de son en postproduction. Ces systèmes rendent possible *“la capacité en postproduction de changer la directivité virtuelle des micros, de récupérer un signal mono ou stéréo, de moduler l'espace.”* (MAGNIER, 2019) Contrairement aux systèmes matricés, les systèmes non matricés (double ORTF⁸, croix IRT⁹, arbre Decca¹⁰) prennent en compte un écartement des

⁸Office de Radiodiffusion Télévision Française

⁹Institut für Rundfunktechnik

¹⁰Il tient son nom de la maison de disques *Decca Records*, dont les ingénieurs ont conçu ce système microphonique dans les années 1950

microphones ainsi qu’une angulation relative, ce qui permet une excellente restitution de l’espace. Étant donné que les microphones sont espacés, il en résulte des écarts temporels entre chaque capsule.

Dans le cadre de notre étude, seuls les systèmes non matricés attirent notre attention, car nous cherchons à prendre en compte le caractère temporel présent dans ce type de système. Michael Williams réalisa plusieurs recherches sur les différentes configurations de ces systèmes. L’un d’eux porte même son nom, le MMAD¹¹ (ou Umbrella¹² de Williams) inspiré de ces recherches sur la “*Multichannel Microphone Array Design*”. Le MMAD est un système de captation qui offre plusieurs possibilités, notamment le choix de la directivité des microphones, des angles de prise de son déterminés par leur espacement et leur ouverture, ainsi que du nombre de microphones utilisés. Il peut ainsi être configuré en arbre 4.0, 5.0, 6.0, 7.0 ou 8.0, dans le respect des abaques de Williams (Annexe B : Abaques de Michael Williams). Ce système “*a l’avantage de donner un excellent rendu spatial et une image très bien localisée, c’est en revanche un système complexe à mettre en place, car l’utilisateur doit respecter les angles et les distances entre les microphones suivant l’abaque qu’il choisit.*” (PÉRIAUX *et al.*, 2015)

Comme mentionné précédemment, il existe d’autres systèmes de prise de son espacés comme l’arbre Decca. Inventé dans les années 1950, il est à l’origine, un système à trois microphones omnidirectionnels où le microphone central est replacé au centre de la stéréo (Figure 1.13). Ce système eut un succès dans l’enregistrement des musiques de film grâce à son microphone central. Il permettait d’avoir un centre physique et fut un atout particulièrement adapté aux formats de diffusion cinématographiques de l’époque, comme le Dolby Stéréo (1.1.2). (PÉRIAUX *et al.*, 2015)

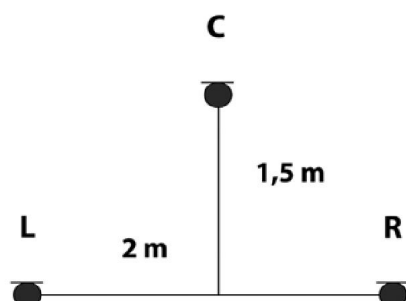


Figure 1.13 : Configuration du triangle Decca (PÉRIAUX *et al.*, 2015)

¹¹*Multichannel Microphone Array design*, signifie en anglais conception d’un arbre multi-microphonique

¹²Signifie “parapluie” en anglais

Il est possible d'ajouter à ce triangle Decca deux microphones omnidirectionnels afin d'obtenir un système de prise de son à cinq canaux. La configuration de ces deux microphones arrière est variable et ces microphones peuvent être orientés au choix vers l'orchestre ou vers le fond de la salle. En quelques mots, l'arbre Decca offre une bonne image spatiale et un bon équilibre spectral, mais souffre d'une localisation très instable et imprécise. Il nécessite également la mise en place de microphones d'appoints pour compenser cette faiblesse. (PÉRIAUX *et al.*, 2015)

D'autres systèmes sont également appréciés pour leur excellente restitution spatiale et sont donc régulièrement utilisés pour la prise de son d'ambiances. La croix IRT et le double ORTF en font partie. Dans le cas de la croix IRT, on utilise soit quatre microphones cardioïdes espacés de 20 ou 25 cm avec un angle de 90° entre chaque capsule, soit quatre microphones hypercardioïdes avec un espacement réduit à 14 cm. Quant au double ORTF, comme son nom l'indique, il est équipé de deux couples ORTF¹³ espacé de 30 cm (Figure 1.14). (PÉRIAUX *et al.*, 2015 et MAGNIER, 2019)



Figure 1.14 : *Photo de la croix IRT et d'un double ORTF (MAGNIER, 2019)*

En annexe sont référencés d'autres arbres multicanaux comme l'OCT Surround, le carré Hamasaki et l'arbre Fukada. (Annexe C : Arbres microphoniques supplémentaires)

D'après WILLIAMS (2004a), il existe cinq facteurs à prendre en compte lors de la réalisation d'un réseau microphonique, qu'il soit stéréophonique ou multicanal :

- La **courbe de réponse en fréquence** des microphones. Chaque microphone a une

¹³Le couple ORTF est un système de captation composé de deux microphones cardioïdes espacés de 17 cm et disposés selon un angle de 110° entre les deux capsules

réponse en fréquence qui lui est propre et qui définit la bonne restitution du timbre d'une source. Ces réponses en fréquences sont bien sûr liées aux directivités des microphones, particulièrement en basses fréquences.

- La **position du réseau microphonique** afin de trouver la bonne proportion de signal direct et réverbérant ainsi que le bon équilibre entre tous les éléments de la scène sonore.
- L'**image sonore** (stéréophonique ou multicanale). L'angle de prise de son (SRA¹⁴ en anglais) permet justement de moduler notre image sonore, car c'est grâce à cet angle que nous choisissons la disposition de nos éléments sonores à partir de plusieurs paramètres : espacement, angulation et directivités des microphones.
- La **distorsion angulaire**, elle est la déformation d'un espace sonore perçu lors de la prise de son. Cette déformation se manifeste quand les positions des sources réelles ne correspondent pas avec l'image sonore perçue. Elle dépend des mêmes paramètres que l'angle de prise de son.
- La **compatibilité monophonique** qui se définit par la bonne reproduction de l'image stéréophonique ou multicanale sur un système monophonique.

En plus des cinq facteurs cités ci-dessus, il existe des paramètres où plusieurs choix doivent être faits, dans les cas d'une réalisation d'un arbre multicanal MMAD seulement (WILLIAMS et LE DÛ, 2000). Lors de la réalisation d'un arbre multicanal, nous devons choisir les angles de prise de son des différents segments de notre arbre :

- En ce qui concerne les trois microphones avant, Williams nomme cette couverture **“Front Triplet Coverage”**¹⁵. *“La reproduction de cette scène sonore frontale est probablement l'étape la plus importante dans la conception d'un réseau de microphones multicanaux (MMAD). En général, le son direct de la plupart des sources sonores rencontrées sera couvert par ce triplet frontal, mais la configuration des trois microphones conditionnera aussi considérablement les difficultés rencontrées pour obtenir une “critical linking” avec les segments latéraux et arrière.”*
- L'angle de prise de son des deux microphones arrière est appelé **“Back Pair Coverage”**. Le choix de cet angle dépend de la configuration de la scène sonore. En d'autres termes, la scène sonore entoure-t-elle l'arbre multicanal ou est-elle tout simplement frontale ?

¹⁴Stereo Recording Angle

¹⁵Terme utilisé par Williams qui réunit les deux angles de prise de son des trois microphones avant (gauche, centre, droit) d'un arbre multicanal

- Les angles de prise de son des microphones latéraux (gauche/arrière gauche et droit/arrière droit) sont appelés “**Lateral Segment Coverage**” et sont déduits des deux paramètres précédents.

Ces trois paramètres sont d’une grande importance, car ils permettent de choisir, de limiter ou non, la restitution de la scène sonore frontale, et également de tirer parti de l’environnement “*surround*” (Figure 1.15). De plus, d’après WILLIAMS et LE DÛ (1999), nous devons vérifier si tout notre espace sonore est couvert par notre dispositif, les auteurs prennent en compte ce paramètre à l’aide du “**Critical Linking**”¹⁶.

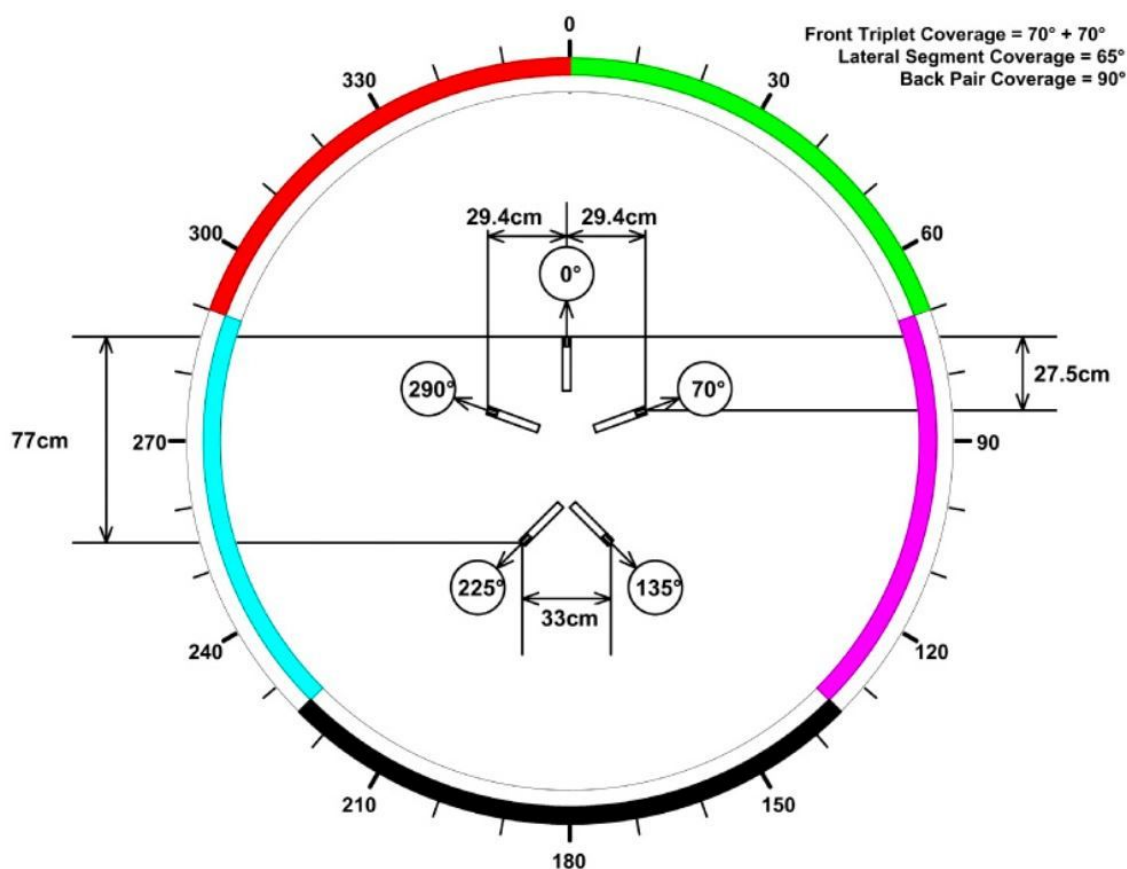


Figure 1.15 : Représentation d’un arbre multicanal à cinq microphones hypocardiïde (WILLIAMS et LE DÛ, 2004)

Concernant un arbre multicanal à sept microphones (Figure 1.16), le paramètre “**Lateral Segment Coverage**” se sépare en deux afin de pouvoir configurer nos différents angles de prises de sons latéraux ; “**Front Lateral Segment Coverage**” et “**Back Lateral Segment Coverage**” sont respectivement les angles de prises de son latéraux avant et latéraux arrière.

¹⁶Corresponds à “l’utilisation d’un décalage angulaire sur deux angles de couverture voisins, de sorte que les limites latérales de chaque angle de couverture se raccordent l’une à l’autre, sans se chevaucher ni laisser d’espace non couvert.” (WILLIAMS et LE DÛ, 1999)

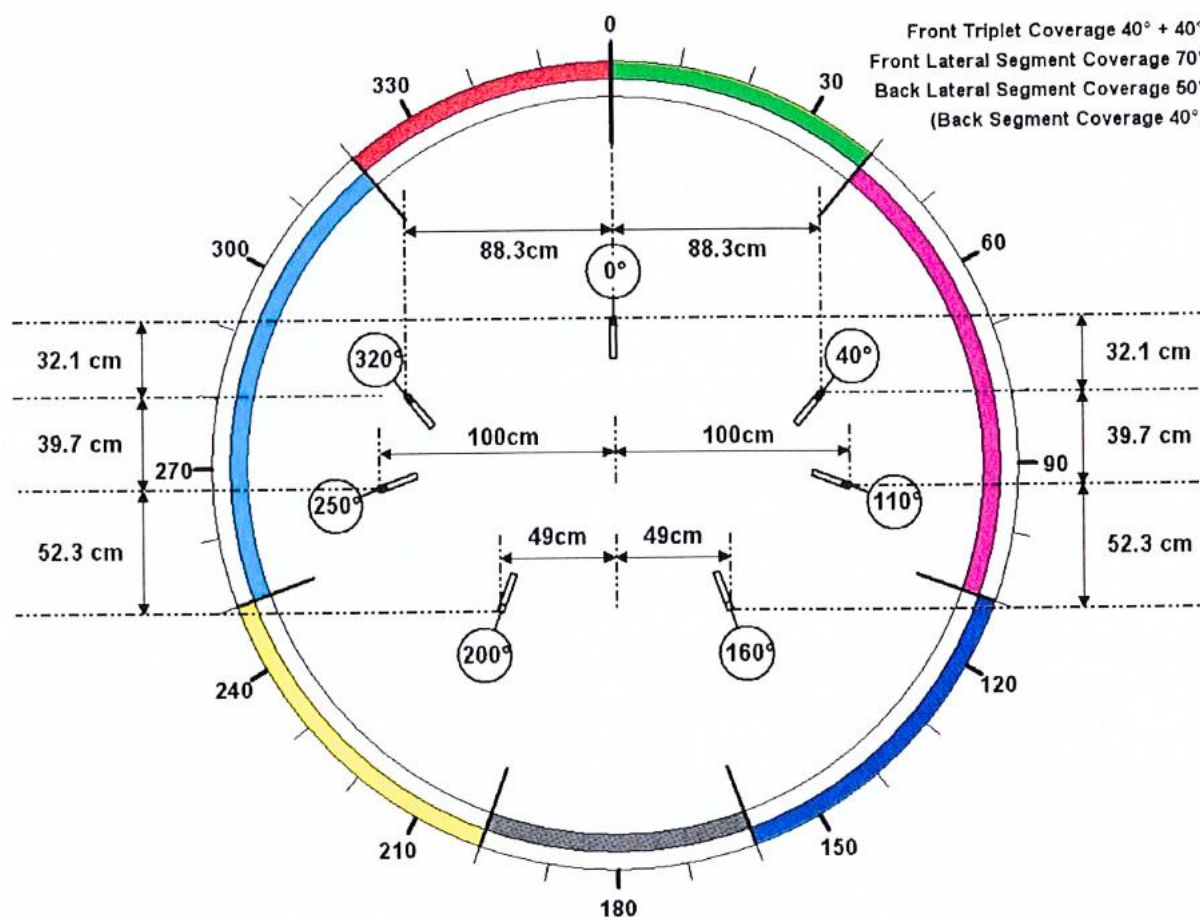


Figure 1.16 : Représentation d'un arbre multicanal à sept microphones cardioïde (WILLIAMS, 2013)

En 2008, lors de l'expansion du format 7.0, Williams propose des nouvelles méthodes afin de fabriquer un arbre multicanal à six, sept ou huit microphones. Son approche *"back to back array design"* consiste à former à un arbre à partir de deux autres arbres complémentaires (WILLIAMS, 2008). En d'autres termes, sa méthode consiste à placer dos à dos deux arbres de formats 4.0 et 5.0 pour obtenir soit un arbre 6.0 à l'aide de deux arbres 4.0, soit un arbre 7.0 avec un arbre 5.0 et 4.0, soit un arbre 8.0 avec deux arbres 5.0, toujours placés en opposition.

Même si cette méthode nous permet d'obtenir plus de précision dans notre espace de 360°, il est nécessaire que les deux arbres dits complémentaires respectent certaines conditions : la somme des angles utiles globaux doit être égale à 360° et les distances entre les microphones placés dos à dos doivent être aussi proches que possible. À partir de la Figure 1.17, on peut ainsi obtenir un arbre multicanal composé de sept microphones hypocardioides, comme illustré en Figure 1.18.

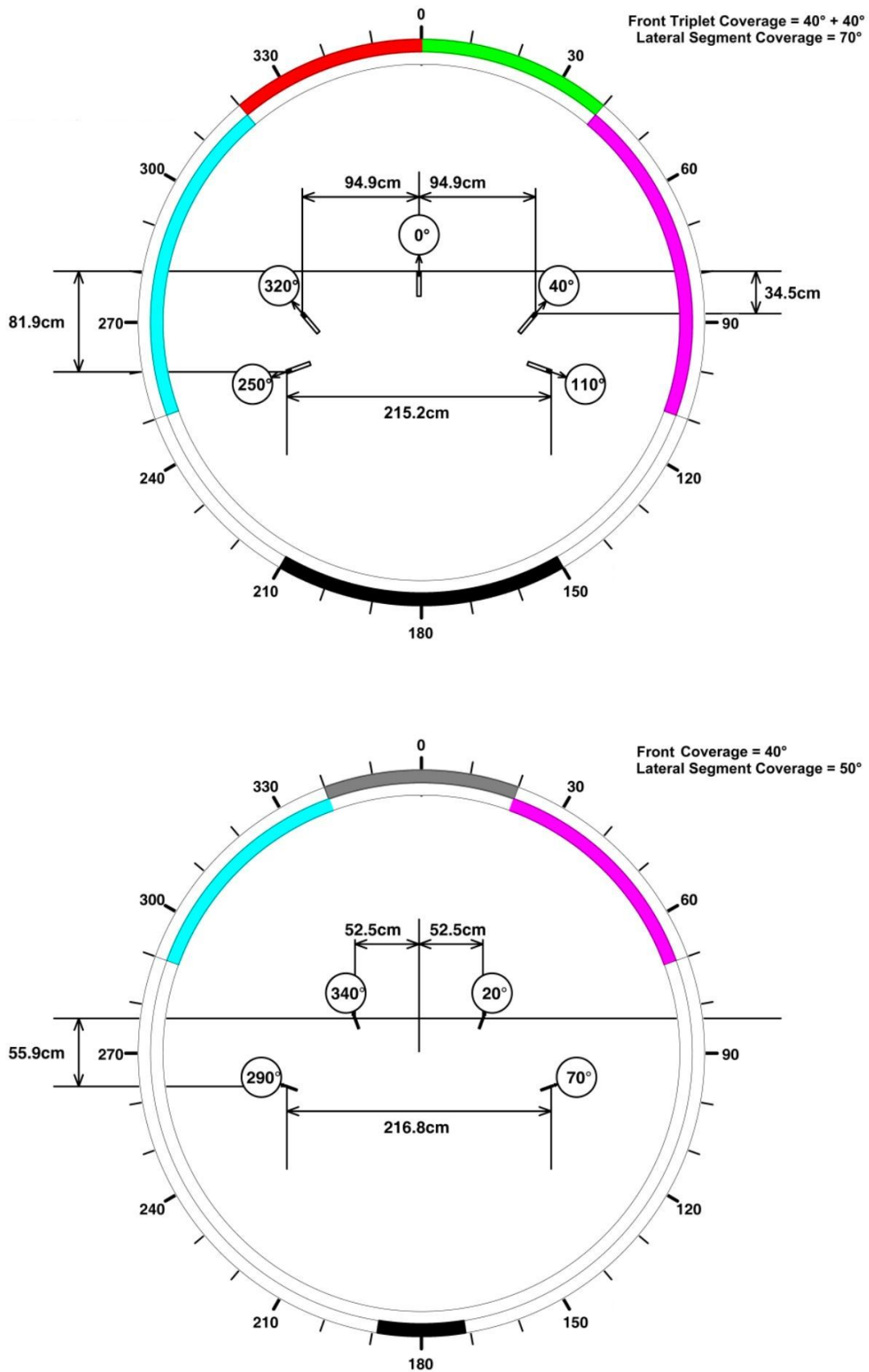


Figure 1.17 : Représentation de deux arbres multicanaux à cinq et quatre microphones hypocardioides (WILLIAMS, 2008)

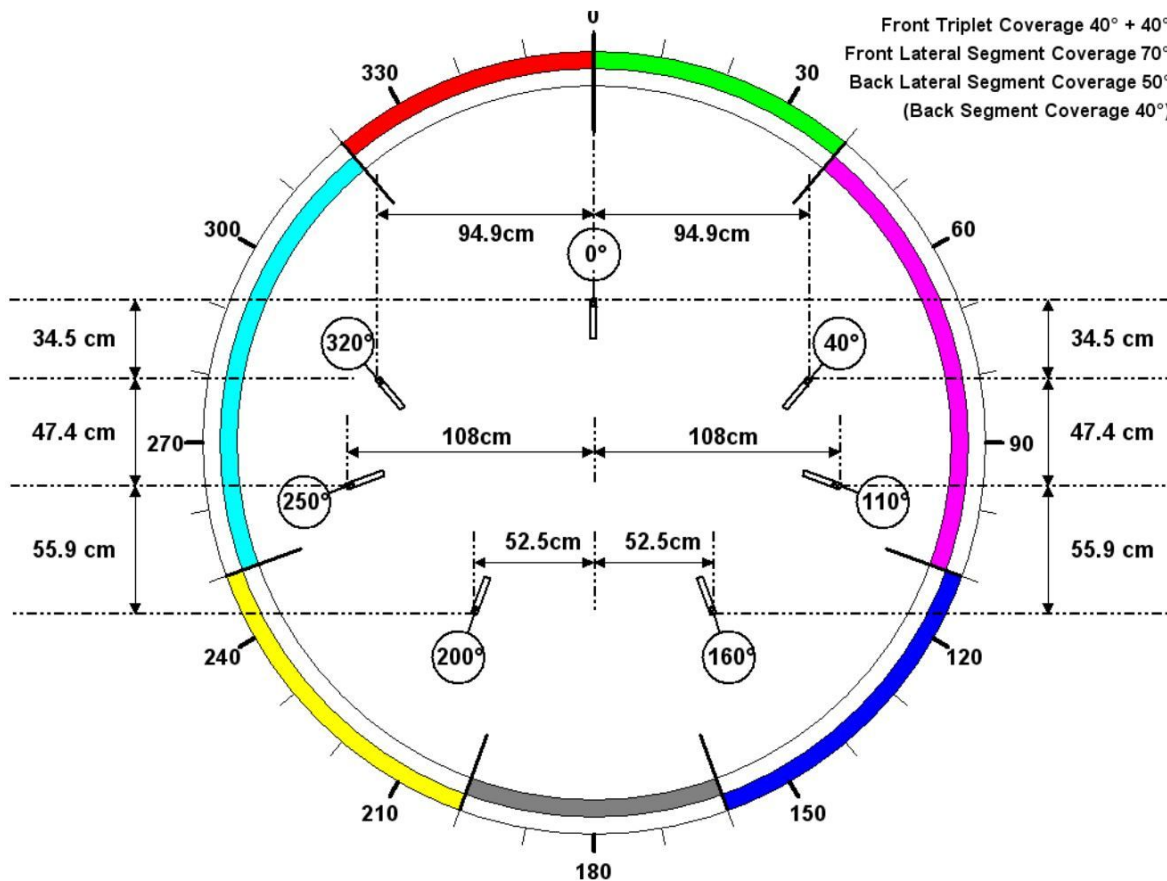


Figure 1.18 : Représentation d'un arbre hypocarديوide obtenue à partir de la méthode "back to back array design" (WILLIAMS, 2008)

1.3.2 Création d'un espace multicanal en postproduction

Il est également possible de recréer un espace sonore multicanal en postproduction sans passer par une captation sonore directe. Il y existe une multitude d'outils à notre disposition, nous n'en citerons que quelques-uns.

1.3.2.1 Techniques de spatialisation

D'après IRCAM (2021), il existe plusieurs façons d'aborder la spatialisation sonore, les technologies existantes sont les suivantes :

- le modèle **perceptif** : il comprend toutes les lois de panoramique allant de la stéréophonie au multicanal que nous aborderons brièvement à la suite en 1.3.2.2. Cette approche, selon Thibaut Carpentier, "essaie d'imiter les effets qui ont un impact sur le système auditif humain" comme l'ITD et l'IID. En d'autres mots, cette famille de techniques se base sur le système perceptif humain afin d'en reproduire les effets qui influencent la perception auditive humaine.

- le modèle **physique**, qui se base sur la synthèse d'un champ sonore. Cette famille comprend le format ambisonique qui attribue à chaque source sonore un encodage selon les axes x, y, z du repère cartésien, appelé format B, indépendamment du système. Ces données sont basées sur une décomposition mathématique de l'espace en harmoniques sphériques. Les signaux du format ambisonique ne sont pas envoyés directement aux enceintes, mais servent d'abord à recréer un champ sonore à partir des données recueillies de chaque source sonore, d'où la dénomination "orientée scène". Il est possible d'encoder selon un nombre défini d'harmoniques sphériques. Plus ce nombre est élevé, plus l'ordre ambisonique (HOA¹⁷) augmente, ce qui permet une restitution spatiale plus précise. Lorsque l'ordre ambisonique vaut 1, le format ambisonique contient les composantes x, y et z de chaque source, ainsi que l'information omnidirectionnelle W. La *Wave Field Synthesis*¹⁸ fait également partie de ce modèle, nous aborderons ce sujet en 1.4.1.
- le modèle **signal** : cette approche repose sur la restitution de deux signaux seulement, réglés en fonction des sons perçus au niveau du tympan après modification par le corps humain. Il existe des techniques de prise de son spécifiques à cette méthode, en utilisant des têtes artificielles ou naturelles, et des techniques de synthèse binaurale se reposant sur les fonctions HRTF. Cette synthèse consiste donc à échantillonner la sphère auditive (en mesurant les HRTF) et en la stockant dans une base de données. Pour spatialiser un son, il suffit de déclencher l'entrée correspondante dans cette base. Cette approche se base sur le traitement du signal.

1.3.2.2 Quelques lois de panoramique

Les premiers outils à disposition sont nos outils de spatialisation qui nous permettent de positionner une source sonore dans notre espace sonore. Plusieurs lois de panoramique existent afin de réaliser la spatialisation cherchée, Le VBAP (*Vector Based Amplitude Panning*) en fait partie. Cette technologie s'inspire des lois de panoramique stéréophonique à énergie constante (atténuation de trois décibels au centre) et manipule le gain des signaux vers deux(2D¹⁹) ou trois(3D²⁰) enceintes les plus proches de la source virtuelle (PULKKI *et al.*, 1999). Pour que ce soit possible, le VBAP a besoin de connaître précisément le positionnement des haut-parleurs du dispositif afin de réaliser une triangulation

¹⁷High Order ambisonic : se définit par un ordre n , la formule suivante permet d'en déduire le nombre d'enceintes (N) à implémenter dans notre système à partir de l'ordre n : $N = (n + 1)^2$

¹⁸Signifie "synthèse de front d'onde" en anglais

¹⁹À deux dimensions

²⁰À trois dimensions, on ajoute l'élévation

des vecteurs gains afin de restituer l'objet sonore dans le dispositif physique. Une des contraintes de ce système est qu'il requiert la même distance entre les haut-parleurs et la position d'écoute, ce qui résulte d'un *sweet-spot*²¹ assez restreint.

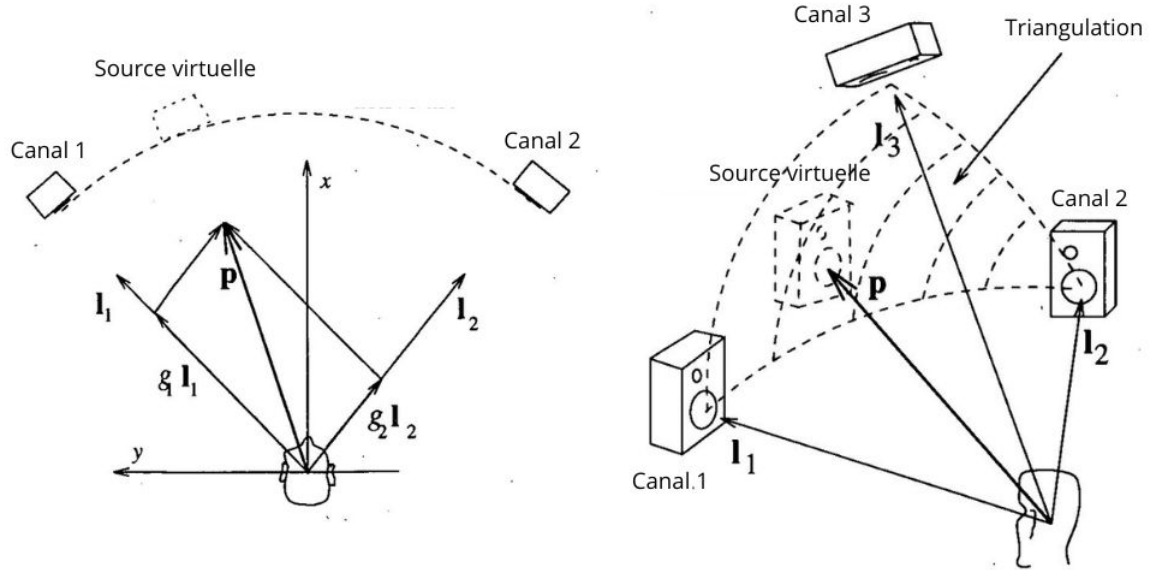


Figure 1.19 : Configuration d'un système à panoramique d'amplitude bidimensionnel et tridimensionnel (PULKKI et al., 1999)

D'autres technologies de panoramique ne dépendent pas de ce *sweet-spot*, comme le DBAP (*Distance Based Angular Panning*). Il a été réalisé avec deux hypothèses ("SPAT Revolution - Panning Algorithms", 2024) :

- "Tous les haut-parleurs sont toujours actifs, indépendamment de la position de la source.
- Le niveau résultant est indépendant de la position de la source."

Dans cette méthode, nous calculons les distances entre la source sonore et chaque haut-parleur afin d'imiter le comportement d'atténuation du gain en acoustique. Nous perdons alors six décibels à chaque doublement de distance. (PECQUAIS, 2019)

Nous avons inclus en Annexe D : Lois de Panoramiques, "SPAT Revolution - Panning Algorithms", 2024, d'autres lois de panoramiques.

²¹Zone d'écoute optimale ou "point de convergence du système d'écoute, en multicanal ITU le centre du cercle d'écoute" (PÉRIAUX et al., 2015)

1.3.2.3 Upmixage

L'upmixage est une méthode de fabrication d'un format sonore supérieur (cinq, sept, onze canaux) à partir d'un format inférieur (un seul canal ou deux canaux). À partir de la phase, du contenu fréquentiel et de l'écart temporel entre deux canaux, les algorithmes d'upmixage permettent de recréer une scène sonore à plus de deux canaux. Comme nous le décrivent PÉRIAUX *et al.* (2015) ; “*L’information hors phase est restituée dans les canaux arrière, l’information en phase, au centre, et le contenu restant alimente la stéréo du multicanal. C’est le principe du décodage Dolby Surround.*” Cette méthode n’en est que la base, en effet, de nombreux algorithmes d'upmixage intègrent d'autres paramètres comme : la réverbération, la séparation des champs direct et réverbéré, puis la spatialisation dans l'espace multicanal par différence d'intensité et de temps à partir des canaux initiaux.

1.4 Systèmes utilisant des différences temporelles

Depuis le début de cette étude, nous n'avons parlé que de systèmes qui utilisent des différences d'intensités afin de recréer un espace sonore. Cependant, d'autres systèmes qui utilisent les différences de temps comme la *Wave Field Synthesis* (WFS) ou le *Virtual Microphone Control* (ViMiC) existent.

1.4.1 *Wave Field Synthesis*

Le principe de la WFS est complètement différent de celui de la stéréophonie ou d'un système multicanal discret qui visent tous les deux à recréer des espaces sonores virtuels. Dans le cas de la WFS, il s'agit “*de recréer des champs acoustiques réels à partir de sources sonores qui peuvent être virtuelles selon le principe d'Huygens*” d'après PÉRIAUX *et al.* (2015) (Figure 1.20).

Les sources secondaires (Figure 1.20) correspondraient donc à notre dispositif de haut-parleurs. Dans notre cas, les enceintes d'un système WFS sont placées en rampe linéaire et non pas en rampe curviligne pour faciliter l'intégration et la gestion de l'installation (Figure 1.21), (CARREAU et MACQUART, 2015).

■ Principe d'Huygens

Huygens en 1678 a énoncé ce principe très intuitif en observant le comportement d'ondes à la surface de l'eau : les vibrations qui se propagent à l'extérieur d'une surface fermée So contenant la source sont identiques à celles qu'on obtiendrait en supprimant cette source et en la remplaçant par des sources convenablement réparties sur la surface So. ■

– PÉRIAUX et al. (2015).

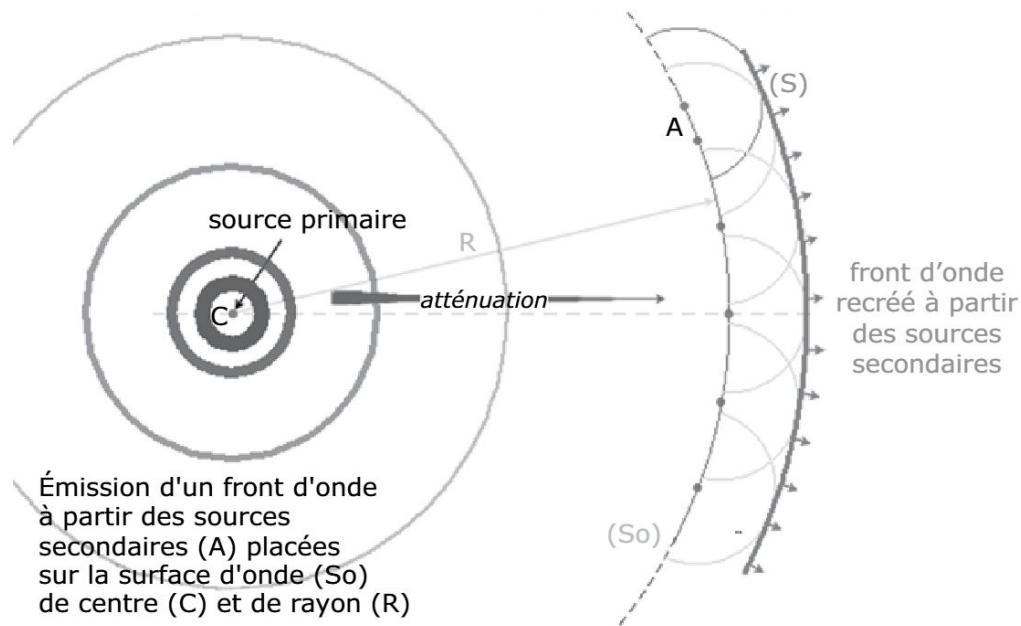


Figure 1.20 : Principe d'Huygens (PÉRIAUX et al., 2015)

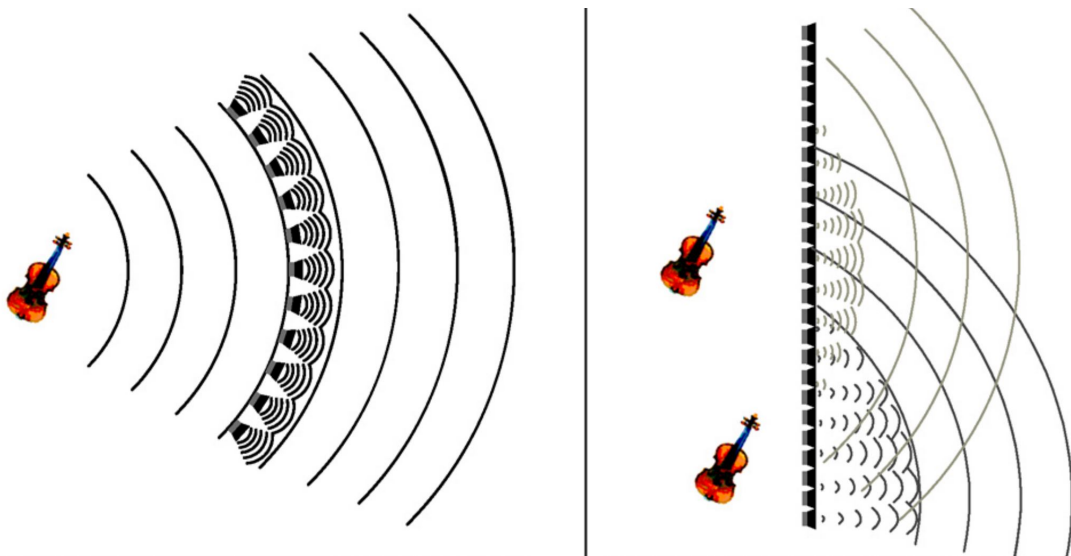


Figure 1.21 : Principe d'Huygens à gauche et exemple d'un dispositif WFS à droite (CORTEEL et CAULKINS, 2004)

On peut observer en Figure 1.21, qu'un champ sonore se déploie à partir de sources virtuelles à travers plusieurs points d'émissions qui contiennent chacun un niveau sonore et un retard temporel défini afin de simuler ce champ. Grâce à cette méthode, la WFS permet de recréer fidèlement un champ sonore, plus le système de reproduction est "défini", (plus il comprend de haut-parleurs). Il en découle ainsi une zone d'écoute élargie beaucoup plus large que des systèmes à canaux discrets (systèmes dérivés de la stéréophonie), où ce sweet-spot est restreint (BERKHOUT *et al.*, 1993).

1.4.2 ViMiC

Le ViMiC est un outil de spatialisation en temps réel développé sur le logiciel *Pure Data* de Miller Puckett puis sur *Max* de Cycling74. Cet outil est spécialement adapté pour des dispositions dans lesquelles le public est dispersé ou réparti sur une zone large (notamment lors de concerts ou dans le cadre d'installations sonores). Il se base sur la simulation d'un arbre microphonique dans un espace à trois dimensions et peut supporter des dispositifs comprenant jusqu'à vingt-quatre haut-parleurs, qui peuvent être placés au choix autour du public (PETERS *et al.*, 2008). Cet outil s'appuie sur la génération d'un espace virtuel dans lequel gains et retards temporels sont calculés à partir de la position des différents microphones et de celle de la source. Différents microphones à directivité variable peuvent être positionnés et inclinés au choix de l'utilisateur. Chaque signal fictif capté par l'un des microphones virtuels peut comporter des premières réflexions ou des informations de réverbération qui dépendent des paramètres de l'environnement virtuel (absorption des surfaces, taille de la pièce...).

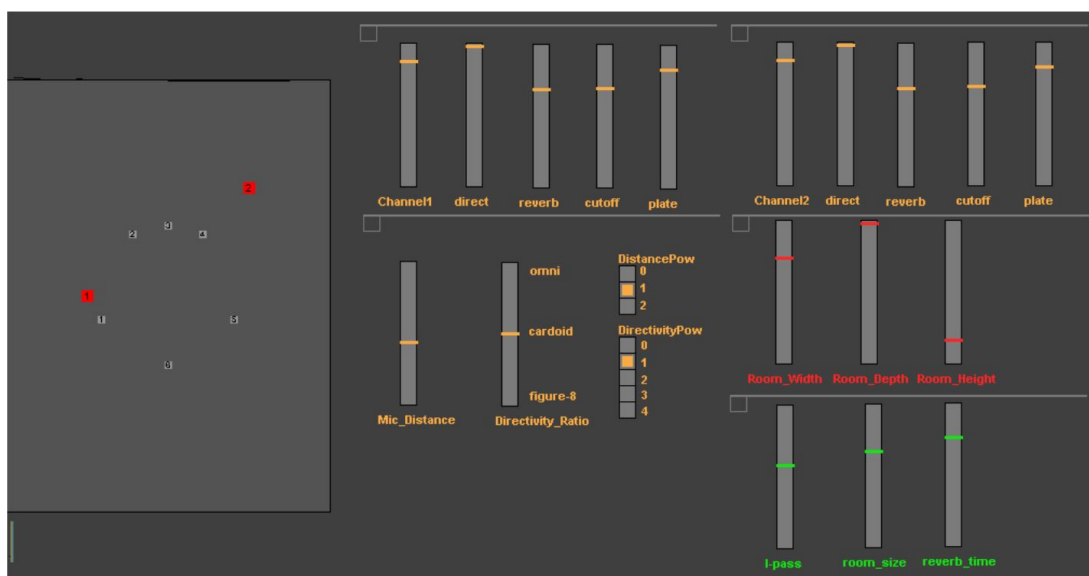


Figure 1.22 : Interface graphique du ViMiC (BRAASCH *et al.*, 2005)

Plusieurs paramètres se présentent à nous lorsque nous utilisons le ViMiC. Les microphones virtuels sont à directivité variable et comprennent six degrés de liberté : leurs positions dans l'espace cartésien (x, y, z) et leurs rotations selon ces mêmes axes x, y, z. Des paramètres liés à la salle virtuelle sont également modifiables : hauteur, longueur, largeur, ainsi que l'absorption et l'influence spectrale des parois à travers l'utilisation de filtres. Le ViMiC utilise un filtre passe-bas ajustable pour simuler l'absorption de l'air sur les premières réflexions, car plus le trajet de propagation du son est long, plus les fréquences aigües sont atténuées, ce qui rend la simulation acoustique plus réaliste. Cet outil s'appuie sur la formule (3) pour émuler différentes directivités et lui applique un exposant w (formule 5) afin d'obtenir un diagramme de directivité artificiellement plus serré (PETERS *et al.*, 2008).

$$\gamma = (a + b * \cos(\theta))^w \quad \text{avec } 0 \leq a, b \leq 1 \quad (5)$$

De plus, “contrairement aux caractéristiques réelles des microphones, qui varient en fonction de la fréquence, les microphones virtuels dans l'outil ViMiC sont conçus pour appliquer le concept de directivité microphonique sans simuler les dépendances fréquentielles indésirables.” (PETERS *et al.*, 2008)

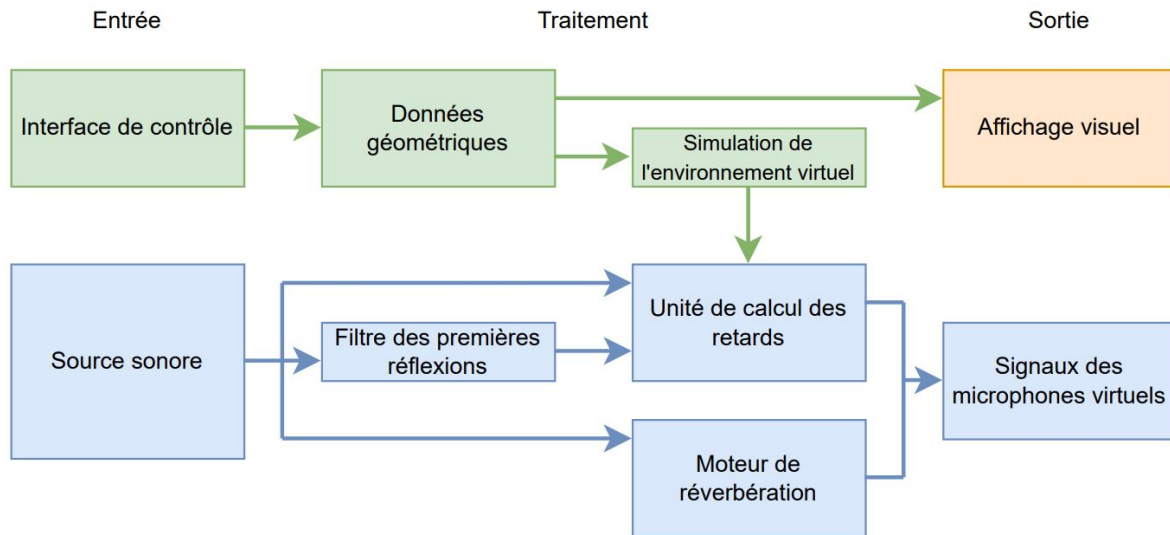


Figure 1.23 : Architecture de l'environnement virtuel ViMiC, d'après BRAASCH et al. (2007)

1.5 Perception spatiale en salle

1.5.1 Configuration de la salle et variabilité inter-salles

■ *En mettant en œuvre des délais entre les canaux, il est plus facile de combler les écarts relativement importants entre les haut-parleurs avant et les haut-parleurs arrière. Alors que la majorité des systèmes sonores reposant sur des lois de panoramique en amplitude sont optimisés pour des réseaux de haut-parleurs disposés à intervalles réguliers (comme huit haut-parleurs espacés de 45° dans le plan horizontal), notre approche [ViMiC] s’adapte facilement à diverses configurations surround, avec ou sans élévation. En effet, les techniques d’enregistrement actuelles, telles que l’omnidirectionnel espacé, peuvent être reproduites de manière simple au sein du système virtuel. Cette méthode présente donc l’avantage d’un contrôle spatial rapide et intuitif de l’emplacement de la source dans des environnements immersifs multicanaux, et la possibilité d’apprendre, dans l’espace virtuel, les performances de différents réseaux de microphones dans des espaces acoustiques.* ■

– BRAASCH *et al.* (2005).

Cette citation nous informe que les lois de panoramiques en amplitude, également utilisées au cinéma, sont efficaces pour des systèmes dont les haut-parleurs sont à équidistance les uns des autres, or au cinéma, c’est rarement le cas. L’Union Internationale des Télécommunications recommande la disposition présentée en Figure 1.24 en ce qui concerne le format 5.1. L’une des recommandations importantes concerne l’angle formé par les enceintes gauche (L) et droite (R) par rapport à la position d’écoute optimale (*sweet-spot*), qui est fixé à 60° (soit $\pm 30^\circ$ de part et d’autre de l’axe central).

D’après XIE (2022), les enceintes dites “*surround*” sont utilisées pour la reproduction de sons ambiants tels que des ambiances de ville ou fonds d’air. Le système “5.1” ne se définit pas comme un système qui permet de reproduire un espace sonore à 360° , mais à reproduire une scène frontale précise au détriment d’une mauvaise reproduction sonore à l’arrière.

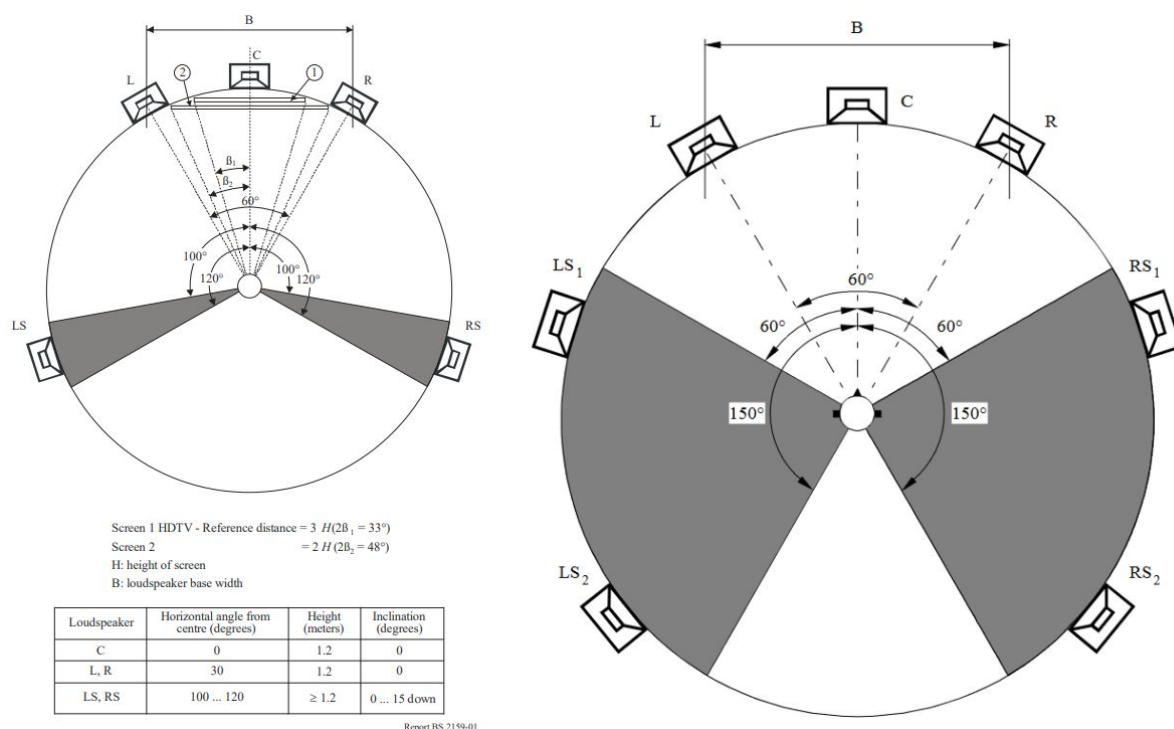


Figure 1.24 : Schéma de la disposition des enceintes pour le format de spatialisation 5.1 (gauche) et positionnement des enceintes “surround” en cas de multiples canaux arrière (droite) (RECOMMENDATION ITU-R BS.775-1 - Multichannel Stereophonic Sound System with and without Accompanying Picture, 1994)

Au cinéma, il existe le *Surround Array*²² qui permet de diffuser les canaux arrière dans plusieurs haut-parleurs (Figure 1.4). La recommandation quant à leur placement en salle est décrite par HOLMAN (2008), “il doit y avoir un nombre pair [de haut-parleurs] réparti dans les moitiés gauche et droite d’une salle qui occupe une région comprise entre $\pm 60^\circ$ et $\pm 150^\circ$, divisé de manière égale et symétrique.” (Figure 1.24) Cet ensemble d’enceintes permet donc de diffuser plus uniformément les canaux arrière à un large public dans de grandes salles, mais rendent la localisation de sources ponctuelles imprécise, comme des effets, impossible. En effet, cet effet réduit considérablement l’*exit effect*²³ au détriment de la précision de la localisation sonore. La panoramisation d’une source sonore des haut-parleurs frontaux vers les “surround” est également impactée : cette panoramisation semble se déplacer de l’avant vers les côtés plutôt que des côtés vers l’arrière (problème qui est moins important sur des systèmes discrets à plus de deux canaux arrière tels que le 7.1), d’après HOLMAN (2008). Un dernier désavantage concerne la restitution fréquentielle de ce dispositif. Chaque enceinte d’un même côté (gauche ou droit) qui

²²Ensemble de haut-parleurs “surround”

²³“Cet effet est dû au fait que lorsque notre attention est détournée de l’écran par un effet sonore, ce que nous regardons n’est plus l’écran où est projeté le film, mais plutôt le panneau de sortie.” d’après HOLMAN (2008)

diffuse un même canal audio, est perçue par le public avec des différences de temps dues aux variations de distance entre les haut-parleurs et l'auditeur ou l'auditrice, ce qui entraîne un effet de filtrage en peigne.

Quant aux autres canaux, le canal central doit être positionné à 0° afin d'obtenir un ancrage sonore sur cet axe pour des sources comme la voix. Les haut-parleurs à gauche et à droite sont positionnés selon la directive suivante : ces deux haut-parleurs doivent recouvrir l'entièreté de l'écran de projection à l'aide de l'enceinte centrale afin de pouvoir mettre en son chaque partie de l'écran (HOLMAN, 2008). Les enceintes à gauche et à droite se placent alors aux extrémités de l'écran de projection, leurs positions angulaires varient donc en fonction des salles.

WALKER (1998) préconise les paramètres liés à une salle de cinéma ou une salle munie d'un système multicanal, voici un tableau référençant ses préconisations :

Élément	Spécification
Taille de la pièce	215–645 ft ² (20–60 m ²)
Forme de la pièce	Rectangulaire ou trapézoïdale (pour que les effets des modes de la pièce soient calculables ; si une forme complexe est souhaitée, un autre test est possible)
Symétrie de la pièce	Symétrique par rapport à l'axe du haut-parleur central
Proportions de la pièce	$1.1 \times (w/h) \leq l/h \leq 4.5 \times (w/h) - 4$ et $l/h < 3$ et $w/h < 3$, où l est la longueur, w la largeur et h la hauteur. Les ratios entre la longueur (l), la largeur (w) et la hauteur (h) qui sont à $\pm 5\%$ des valeurs entières doivent être évités.
RT60 moyenne de 200 Hz à 4 kHz	$T_m = 0.3 \times (V/V_0)^{\frac{1}{3}}$ avec une tolérance de $+25\%$ à 0% , où V = volume de la pièce et V_0 = le volume de référence de 100 m ³ (3.528 ft ³).
Diffusion	Doit être élevée et appliquée de manière symétrique (1/2 réfléchissante, 1/2 absorbantes) ; aucune surface réfléchissante dure ne doit renvoyer le son vers les zones d'écoute critiques à un niveau supérieur de > 10 dB par rapport au niveau de réverbération (les ports ²⁴ et les portes sont une préoccupation particulière).
Contrôle des premières réflexions	Niveau spectral de -15 dB pour les premières 15 ms.
Bruit de fond	$< \text{NC-17}$ basé sur une enquête menée dans de nombreux salons aux États-Unis, NR15 équivalent européen.

Tableau 1.1 : *Caractéristiques acoustiques recommandées pour une salle munie d'un système multicanal, tiré de HOLMAN (2008)*

²⁴Traitements acoustiques : espace conçu pour permettre la circulation de l'air et donc du son dans une pièce.

Par l'utilisation de haut-parleurs aux directivités similaires dans des acoustiques similaires au cinéma, un accord fut adopté quant à la mise en place de la courbe "X" (accord ISO 2969). Cette courbe, autrement dit filtre, est appliquée sur chacune des enceintes de notre dispositif (mis à part le caisson de basses fréquences) et peut différer selon la taille de la salle. Depuis 1982, un tableau de correction de la courbe "X" en fonction du nombre de sièges en salle fut adopté (Tableau 1.2), de plus une seconde pente plus raide entre 10 et 16 kHz a été mise en place (Figure 1.25). (ALLEN, 2006)

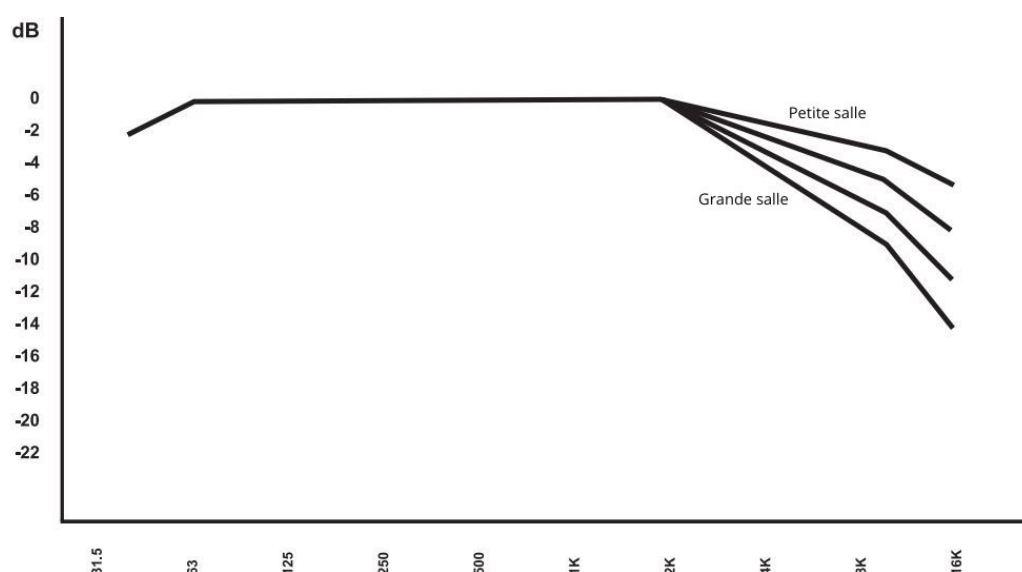


Figure 1.25 : Courbe "X" convention ISO 2969 (ALLEN, 2006)

Fréquence	Nombre de sièges					
kHz	30	150	500	1000	1500	2000
2.0	0	0	0	0	0	0
4.0	1.0	0.5	0	-0.5	-1.0	-1.5
8.0	2.0	1.0	0	-1.0	-2.0	-3.0

Tableau 1.2 : Tableau de correction en dB de la courbe "X" en fonction du nombre de sièges

Pour revenir à Braash cité en début de partie, une simulation d'arbre microphonique qui inclurait non seulement des différences d'intensité mais également des différences de temps, pourrait combler les écarts importants entre nos différents haut-parleurs (exemple de notre enceinte gauche et arrière gauche qui peuvent présenter un écart variant de 30° à 120°).

1.5.2 Le *sweet-spot*

■ *L'un des objectifs initiaux du développement du système multicanal était d'améliorer la zone d'écoute afin de permettre à un certain nombre de personnes d'écouter une image sonore satisfaisante en même temps. Malheureusement, la légère amélioration de la stabilité de la scène sonore frontale est compromise par l'ajout d'un segment arrière intrinsèquement instable. En d'autres termes, la légère amélioration de l'amplitude du "sweet-spot" créée par l'utilisation des trois haut-parleurs frontaux est contrecarrée par ce segment arrière très instable. Toutefois, les niveaux d'énergie dans ce segment arrière peuvent être si faibles que cet effet soit presque négligeable. C'est l'un des facteurs que l'ingénieur du son doit prendre en compte dans le processus de conception des réseaux de microphones multicanaux. L'amplitude du "sweet-spot" peut en effet être améliorée, mais seulement au prix de la perte d'une bonne localisation. L'utilisation des haut-parleurs frontaux comme sources individuelles, comme dans le mixage pour le cinéma, et bien sûr l'illusion créée par l'utilisation d'un traitement de localisation ambigu ou non spécifique, créeront évidemment une zone d'écoute plus large, mais nous perdrons toutes les qualités de reproduction d'une scène sonore virtuelle réaliste.* ■

– WILLIAMS (2004a).

Nous tenons à rappeler que le *sweet-spot* est la zone d'écoute optimale ou "*point de convergence du système d'écoute, en multicanal ITU le centre du cercle d'écoute*" (PÉRIAUX *et al.*, 2015). Cette citation de Williams est très précieuse dans notre travail de recherche, il nous précise qu'à partir d'un arbre microphonique multicanal, deux paramètres sont interdépendants : la localisation d'une source et la zone d'écoute. Selon Williams, l'amélioration d'un des paramètres entraînerait la dégradation de l'autre. Une amélioration de la stabilité d'une scène sonore frontale est compromise par l'instabilité de la scène sonore "*surround*". Cette instabilité arrière peut être minimisée par une diminution d'énergie des canaux arrière.

■ *Au fur et à mesure que l'angle entre les haut-parleurs augmente, le "sweet spot" devient de plus en plus critique. Que vous considériez cette zone comme un inconvénient ou non dépend simplement du nombre de personnes qui doivent écouter en même temps. Au cinéma ou dans tout rassemblement public, une zone d'écoute critique est évidemment prohibitive.* ■

– WILLIAMS (2004a).

Au cinéma, l'écart entre les différents haut-parleurs peut souvent être très important selon la taille de l'écran, ce qui en résulte est un *sweet-spot* potentiellement critique comme nous

l'indique Williams. Selon la position de l'auditeur ou de l'auditrice dans la salle, qu'il ou elle soit proche ou éloigné de l'écran, cet angle varie constamment. De plus, il existe très peu d'endroits où le public peut réellement bénéficier du *sweet spot*.

1.5.3 L'utilisation des différences de temps en complément de celles d'intensités

La différence de temps, en complément de celle d'intensité, pourrait être un atout majeur dans nos outils de spatialisation. Comme nous le décrivent ROGINSKA et GELUSO (2018), la combinaison de ces deux différences forme une meilleure localisation qui nous paraîtrait plus naturelle que d'utiliser l'une des deux individuellement, car dans un environnement naturel, notre système perceptif capte rarement des signaux acoustiques arrivant sans différence de temps. Helmut Haas fut l'un des premiers chercheurs à avoir travaillé sur le sujet et à conceptualiser l'effet de précedence. Cet effet, également appelé effet Haas nous explique que lorsqu'un même signal est retardé temporellement sur un des deux canaux, l'auditeur ou l'auditrice localise cette source sonore du côté du canal non retardé. La limite à laquelle nous percevons un écho est précisée à 5 ms pour des "clics" et 40 ms pour des sons complexes (WALLACH *et al.*, 1973).

L'effet Haas a souvent mal été interprété au fil des années : l'idée qu'un deuxième son retardé puisse être amplifié jusqu'à ce qu'il soit intelligible par rapport au premier son non retardé, sans induire de perturbation de perception, n'est valable que pour des délais très courts, inférieurs à 25 ms, la Figure 1.26 nous le confirme. De plus, ces analyses ont été réalisées dans des environnements contrôlés en laboratoire avec un temps de réverbération faible. La transposition de ces expériences dans des conditions de diffusion en salle ou en extérieur, qui sont des environnements qui possèdent des temps de réverbération plus importants, doit donc être considérée avec précaution.

Bien que l'effet Haas soit largement utilisé dans la conception de systèmes sonores et de diffusion publique, le gain de plus de 10 dB indiqué par la Figure 1.27 est difficilement atteignable sans perte de localisation ou d'audibilité. (MAPP, 2007)

Blauert, dans un environnement expérimental contrôlé, muni de haut-parleurs émettant des signaux cohérents non périodiques, indique qu'un retard τ compris entre $630\mu\text{s}$ et 1ms entraîne une localisation de la source du côté du haut-parleur émettant en premier le signal. Si τ est supérieur à 1 ms, on constate que la direction de la source sonore reste presque constante (BLAUERT, 1996).

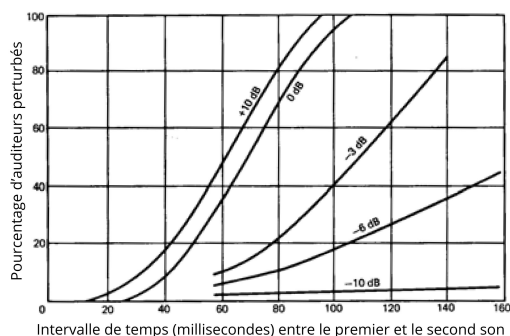


Figure 1.26 : *Courbes illustrant la gêne due à l'écho d'un premier son par rapport à un second son retardé et amplifié (MAPP, 2007)*

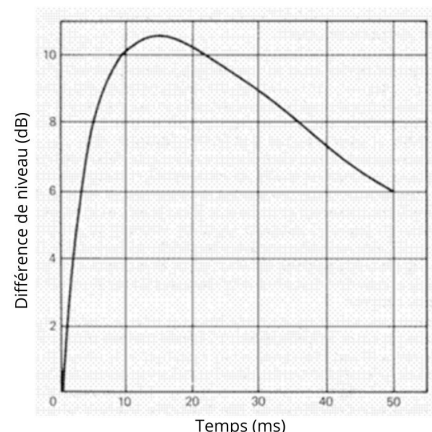


Figure 1.27 : *Courbe classique de l'effet Haas (MAPP, 2007)*

L'effet de précedence joue un rôle majeur dans la localisation sonore. Prenons, par exemple, une source sonore placée à mi-chemin à l'aide d'un pan-pot en amplitude pure entre l'enceinte de gauche et de droite. Nous sommes assis à gauche du centre, la localisation de cette source sera altérée et localisée plus vers le haut-parleur de gauche car l'auditeur ou l'auditrice percevra en premier le front d'onde provenant de l'enceinte de gauche. (XIE, 2022)

Nous nous demandons alors si la présence des différences de temps dans la conception d'un outil pourrait contrer cet effet.

Partie II

Conception de l'outil

2.1 Présentation générale de l'outil

2.1.1 Idée de l'outil

Ce mémoire se base sur la recherche d'un nouvel outil de spatialisation multicanal qui intègre plusieurs mécanismes perceptifs (vu en 1.2). L'objectif est de comprendre et de saisir la pertinence de ces mécanismes dans nos outils et leur utilité, tout en examinant si notre approche peut offrir une nouvelle façon de spatialiser une image sonore. Cet outil se fonde sur la simulation d'un arbre microphonique et a été réalisé sur le logiciel *Max*. Il permet de travailler avec les différents mécanismes perceptifs suivants : différence de temps et d'intensité, et ce, indépendamment ou simultanément.

Dans cette recherche, nous avons choisi de réaliser une simulation d'arbre multimicrophonique afin de recréer un espace sonore à l'aide d'un dispositif que nous connaissons tous bien, ou qui nous est familier en matière de prise de son. Cet outil est né du désir de recréer un environnement sonore plus cohérent à l'image, dû à l'absence de certains mécanismes perceptifs, mais aussi d'imaginer la spatialisation sonore d'une nouvelle manière. En effet, comme nous avons pu l'aborder en 1.1, la manière de penser la spatialisation au cinéma est devenue plus commerciale que rationnelle. De plus, les mécanismes perceptifs absents dans le monde cinématographique permettraient de penser la spatialisation d'une nouvelle façon et apporteraient de nombreux avantages (1.5.1 et 1.5.3).

Nous cherchons alors à créer un outil dont les paramètres et les configurations s'inspirent au mieux des concepts abordés en 1.3.1.2.

2.1.2 Cahier des charges

Cet outil s'articule autour de la simulation d'un arbre microphonique multicanal à sept capsules virtuelles. Nous devons répondre aux critères suivants :

Perceptif :

- Proposer une nouvelle expérience de spatialisation multicanale qui soit différente des outils actuellement disponibles

Technique :

- Réaliser une simulation d'environnement virtuelle, qui nous permettrait d'en déduire des différences d'intensités et de temps
- Réaliser une unité de calcul des retards (différences de temps)
- Proposer un panel de directivités microphoniques variables
- Permettre d'utiliser indépendamment ou simultanément les différences d'intensités et de temps

Interface :

- Avoir une interface ergonomique et facile d'utilisation
- Utiliser des termes liés à la prise de son concernant les différents paramètres à manipuler

2.1.3 Décomposition en blocs fonctionnels

Notre projet *Max* s'articulera en plusieurs blocs. Un premier s'occupera de récupérer les données paramétrées par l'utilisateur ou l'utilisatrice afin d'en déduire les coordonnées de chaque microphone virtuel et celle de notre source. Un second sera dédié aux différents calculs à effectuer et un dernier, à modélisation microphonique.

Lorsque l'utilisateur ou l'utilisatrice se sert de notre outil, il ou elle a plusieurs paramètres à sa disposition. Le premier qui paraît le plus évident est le positionnement de la source dans notre interface graphique à deux dimensions. Les autres paramètres dépendent tous de la simulation de notre arbre microphonique :

- Les écartements entre les différents microphones virtuels (en centimètres) ainsi que leur ouverture (en degrés)
- Le déplacement et la rotation de l'arbre multicanal (en degrés)

- Le choix de la directivité microphonique à partir d'un coefficient allant de zéro à un

Le premier bloc s'occupe d'obtenir les coordonnées cartésiennes de chaque microphone virtuel, qui s'articule autour d'un point central et nous permet de déplacer l'entièreté de l'arbre microphonique. Ce module détermine également l'angulation de chaque microphone à partir de l'ouverture et de la rotation choisie. Le second bloc récupère ces données ainsi que les coordonnées cartésiennes de la source afin de réaliser plusieurs opérations que nous détaillerons dans quelque temps. Enfin, notre dernier bloc nous permet de modéliser le comportement d'un microphone en simulant des directivités variables et en obtenant nos différences d'intensités et de temps à partir des résultats récoltés.

2.2 Calcul de la distance et de l'angulation source/microphone

Ce module nous permet d'obtenir la distance et l'angulation entre chaque microphone simulé et notre source. À partir des coordonnées cartésiennes de notre source et celles de nos microphones virtuels, nous pouvons en déduire leurs distances λ en mètre à l'aide de la formule suivante :

$$\lambda = \sqrt{(X_s - X_m)^2 + (Y_s - Y_m)^2} \quad (5)$$

Les variables X_s , Y_s , X_m et Y_m correspondent aux coordonnées cartésiennes de notre source et du microphone correspondant.

En effectuant la différence des coordonnées cartésiennes de notre source à celles de nos microphones virtuels (ΔX et ΔY), nous pouvons acquérir chaque coordonnée angulaire polaire relative en utilisant la fonction arc tangente.

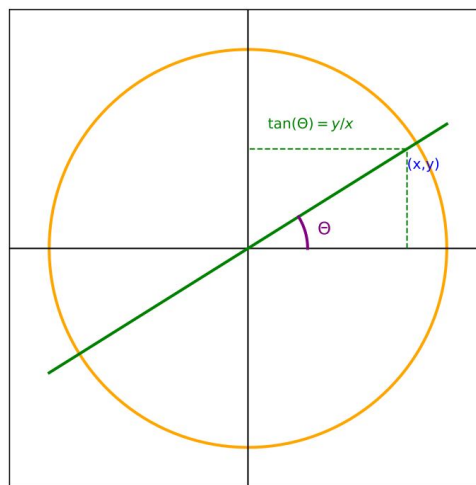


Figure 2.1 : Représentation de $\tan \theta$ à l'aide de Python

Afin d'obtenir l'angle du microphone vers la source, nous inversons le signe de notre fonction arc tangente. De plus, nous ajoutons $\frac{\pi}{2}$ afin d'ajuster le système de référence angulaire (correction de l'angle 0° situé sur l'axe des abscisses positif pour le ramener sur l'axe des ordonnées positif). Nous corrigeons ensuite l'angle en soustrayant à cette valeur l'angulation de notre microphone (α) dans l'environnement virtuel, modifié par l'ouverture d'une paire de microphones et de sa rotation, et nous appliquons un modulo 2π afin de contraindre notre angle θ dans le cercle trigonométrique, c'est-à-dire dans l'intervalle $[0, 2\pi[$.

$$\theta = -\arctan \frac{\Delta Y}{\Delta X} + \frac{\pi}{2} - \alpha \pmod{2\pi} \quad (6)$$

2.3 Conception du module de modélisation microphonique

À l'aide de notre module précédent, nous pouvons maintenant appliquer tous les traitements sur le signal audio de notre source.

Premièrement, nous pouvons affecter une intensité I en $\text{W} \cdot \text{m}^{-2}$ dépendante de la position de notre source à l'aide de notre paramètre de distance calculé auparavant. Nous utiliserons alors la loi en carré inverse qui définit qu'une quantité physique est inversement proportionnelle au carré de la distance :

$$I = \frac{1}{\lambda^2} \quad (7)$$

Cependant, suite à quelques essais, nous avons pu nous rendre compte que notre intensité tendait vers l'infini si λ était très petit, en d'autres termes, si notre source se rapprochait trop de l'un de nos microphones virtuels. Nous avons alors fixé une valeur minimum à notre paramètre λ à l'aide du bloc *clip* dans Max. Ce qui a pu résoudre notre problème.

Ajoutons maintenant un retard temporel t en millisecondes dépendant également de la position de notre source. En admettant que la célérité du son c dans l'air soit de $340 \text{ m} \cdot \text{s}^{-1}$ et en utilisant notre paramètre de distance, nous pouvons utiliser la formule suivante :

$$t = \frac{\lambda}{c} * 1000 \quad (8)$$

Nous n'avons plus qu'à convertir notre retard temporel en *samples*¹ à l'aide du bloc *mstosamps* afin d'utiliser le bloc *delay* dans Max que nous affectons au signal audio.

¹Échantillons audios

En récupérant l'angle θ calculé précédemment, nous pouvons en déduire une réponse dépendante de la directivité choisie. Comme mentionné antérieurement, à l'aide de la formule (4) nous pouvons en déduire une équation polaire contrôlée par un facteur compris entre 0 et 1 :

$$\gamma = a + (1 - a) \cos \theta \quad \text{avec } 0 \leq a \leq 1 \quad (9)$$

Nos microphones sont maintenant dépendants d'une directivité modulée par le positionnement de notre source et par le biais de l'angle θ .

Ce nouveau tableau référence les différentes directivités pour diverses valeurs de a :

Directivité	a
Omnidirectionnel	1.0
Subcardioïde	0.7
Cardioïde	0.5
Supercardioïde	0.33
Hypercardioïde	0.3
Bidirectionnel	0.0

2.4 Une palette variée de configurations

En fonction des paramètres à disposition sur notre outil (écartement, ouverture et rotation des microphones ainsi que le choix de directivités) nous pouvons ajuster notre arbre comme bon nous semble. Nous explorerons dans cette partie quelques configurations possibles, en sachant qu'un nombre important d'alternatives ne sera pas abordé dans ce mémoire.

2.4.1 Arbre multicanal à sept capsules omnidirectionnelles

Réalisons un arbre multicanal à sept microphones omnidirectionnels à l'aide de la méthode “*back to back array design*” de Williams. Pour cela, il nous faut deux plans de prise de son complémentaires ; l'arbre 5.0 de la Figure 2.2 couvre 220° à l'avant des 360° que doit remplir notre système de captation à sept microphones. Nous avons opté pour cet arbre 5.0 car nous souhaitons obtenir une couverture du triplet avant de 60° ou 80°, ainsi qu'une couverture des segments latéraux avant de 70°. Ce choix nous permet de respecter, à la fois, les recommandations de l'ITU pour les canaux frontaux, et de rester cohérents avec les pratiques en salle de cinéma, où, selon Holman, les *Surround Arrays* commencent typiquement à 60° et s'étendent jusqu'à 150° (Figure 1.24). Nous avons volontairement conservé un angle relativement modéré afin d'assurer une continuité spatiale acceptable entre les canaux frontaux et latéraux avant.

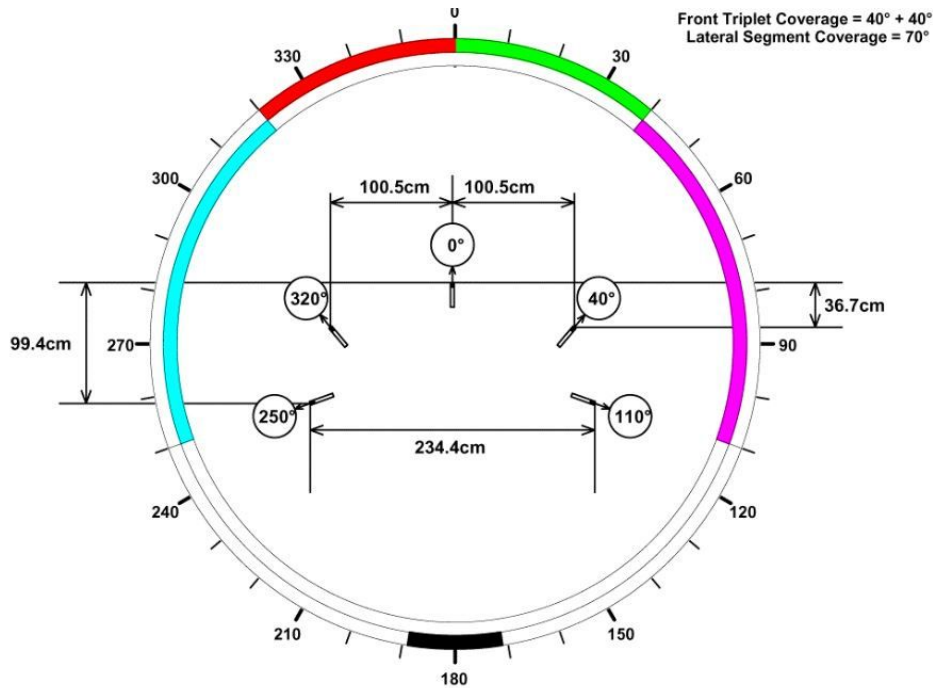


Figure 2.2 : Représentation d'un arbre multicanal à cinq microphones omnidirectionnel (WILLIAMS, 2004b)

Il manque alors 140° de couverture arrière que notre arbre 4.0 doit remplir afin que les deux plans soient complémentaires. La Figure 2.3 présente différentes configurations possibles pour un arbre 4.0, dans lesquelles le couple avant (*Left*, *Right*) couvre 40° . Afin d'assurer une couverture totale de 360° , chacun des deux segments latéraux arrière devra couvrir 50° . La distance entre les microphones Ls et Rs de l'arbre 4.0 est de $(120,8 \times 2 =) 241,6$ cm, ce qui est relativement proche de la distance Ls/Rs de notre arbre 5.0 (234,4 cm). Les deux plans sont donc compatibles et peuvent être superposés pour constituer l'arbre final illustré en Figure 2.4.

Front Segment Coverage	Lateral Segment Coverage	NO CENTRE MICROPHONE							
		LEFT SURROUND		LEFT		RIGHT		RIGHT SURROUND	
		X coord	Ycoord	X coord	Ycoord	X coord	Ycoord	X coord	Ycoord
40°	90°	-78,9cm	-48cm	-56,5cm	0 cm	56,5cm	0 cm	78,9cm	-48cm
	80°	-86cm	-51,1cm					86cm	-51,1cm
	70°	-94,4cm	-54,1cm					94,4cm	-54,1cm
	60°	-106cm	-59cm					106cm	-59cm
	50°	-120,8cm	-64,3cm					120,8cm	-64,3cm
	40°	-143,1cm	-72,6cm					143,1cm	-72,6cm
	30°	-181cm	-87,2cm					181cm	-87,2cm

Figure 2.3 : Représentation de deux arbres multicanaux à cinq et quatre microphones omnidirectionnels (WILLIAMS, 2004b)

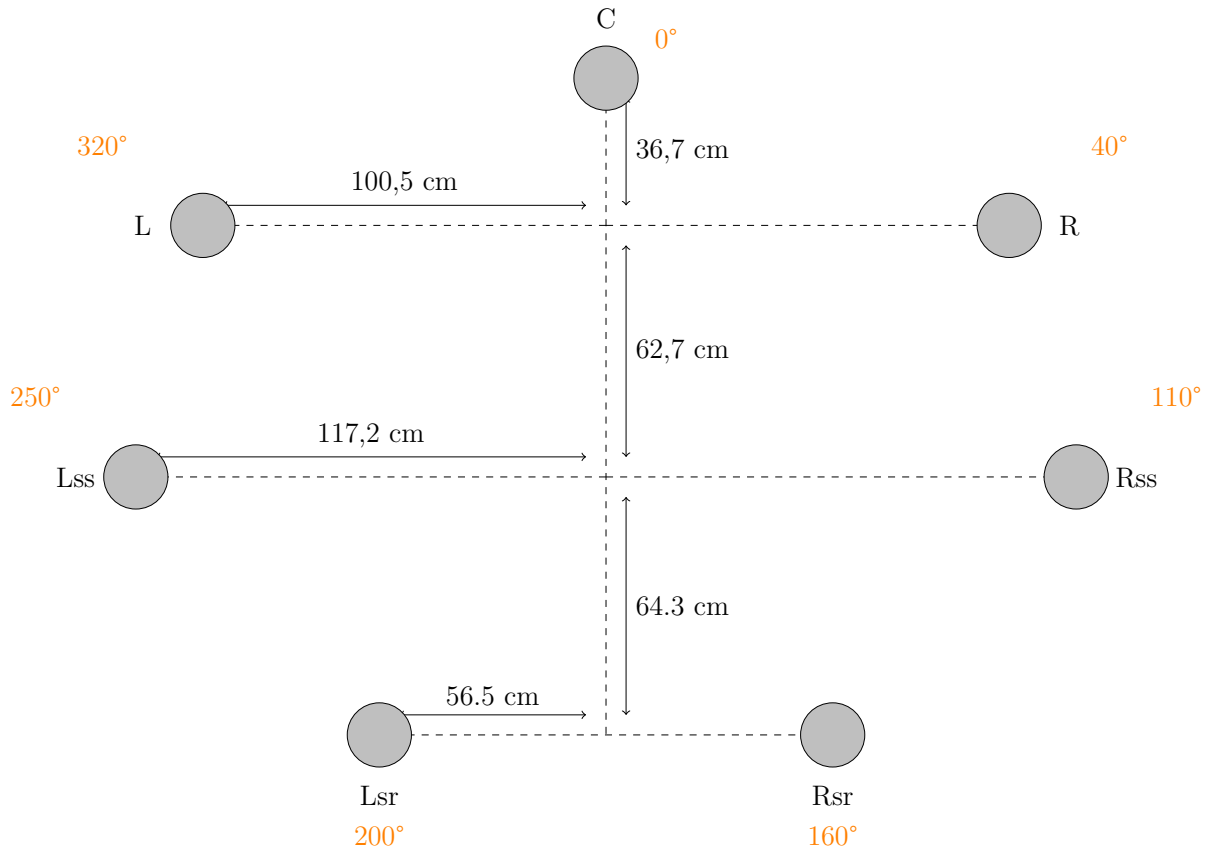


Figure 2.4 : *Arbre multicanal à sept microphones omnidirectionnels, créé à partir des plans complémentaires en Figure 2.2 et Figure 2.3*

2.4.2 Prédominance des différences de temps sur les enceintes frontales

Lors de nos recherches sur les arbres multi-microphoniques, nous avons pu observer qu'il était possible de mettre en avant une prédominance de différences de temps ou d'intensité entre chaque capsule. Nous avons envisagé la réalisation de deux arbres, chacun mettant en avant des prédominances de différences de temps sur différents segments de leur disposition : dans cette section, un arbre avec une prédominance de différence de temps entre les capsules avant, et un autre avec une prédominance de différence de temps entre les capsules latérales en 2.4.3. Ces deux arbres seront réalisés à l'aide de la méthode "*back to back array design*" de Williams.

Nous supposons qu'un écartement important entre nos capsules induit une prédominance de différences de temps grâce à la relation reliant temps et distance vu en 1.2.2. Nous avons donc, cette fois-ci, choisi nos deux plans d'arbres microphoniques complémentaires en nous basant sur la condition suivante : assurer un écart important entre les capsules frontales. En ce qui concerne l'arbre 5.0, nous avons donc opté pour

la configuration en Figure 2.5 qui comporte une distance de ($\sqrt{88.3^2 + 32.1^2} =$) 94 cm entre chaque capsule frontale, et un écart moins important entre nos capsules latérales et frontales de 41 cm. Avec cette disposition, il nous manque alors 140° de couverture arrière que doit remplir notre arbre quadriphonique.

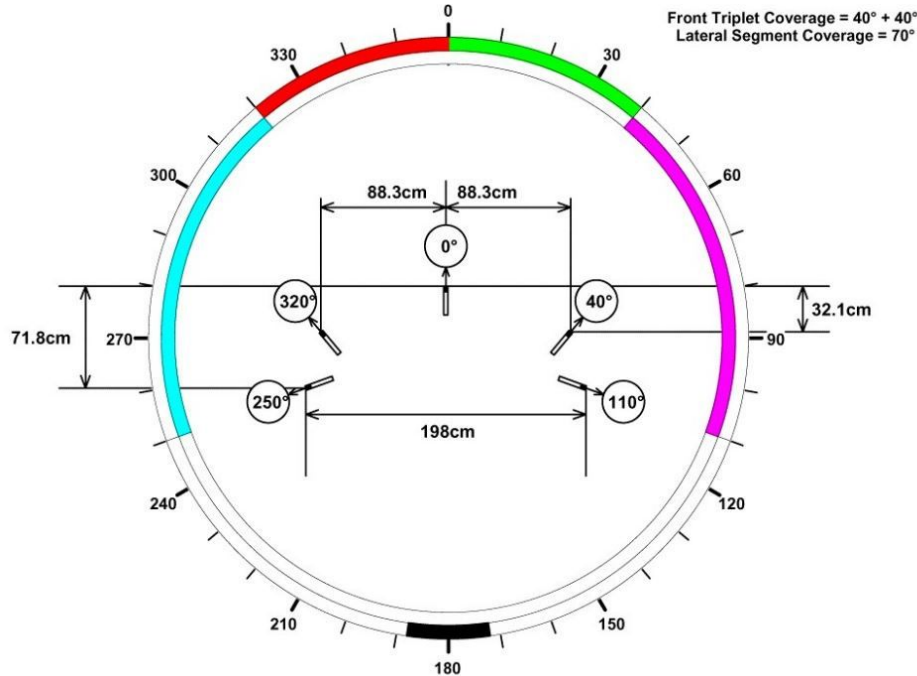


Figure 2.5 : Représentation d'un arbre multicanal à cinq microphones cardioïdes (WILLIAMS, 2004b)

Front Segment Coverage	Lateral Segment Coverage	LEFT SURROUND			LEFT			RIGHT			RIGHT SURROUND		
		Orient	X Coord	Y Coord	Orient	X Coord	Y Coord	Orient	X Coord	Y Coord	Orient	X Coord	Y Coord
40°	90°	250°	-59,6cm	-22,7cm	340°	-49cm	0 cm	20°	49cm	0 cm	110°	59,6cm	-22,7cm
	80°	260°	-65,5cm	-28,6cm							100°	65,5cm	-28,6cm
	70°	270°	-73,1cm	-34,4cm							90°	73,1cm	-34,4cm
	60°	280°	-85cm	-42,9cm							80°	85cm	-42,9cm
	50°	290°	-101,3cm	-52,3cm							70°	101,3cm	-52,3cm
	40°	300°	-124,1cm	-63cm							60°	124,1cm	-63cm
	30°	310°	-147,3cm	-68,8cm							50°	147,3cm	-68,8cm

Figure 2.6 : Différences intercanales de temps (ICTD) et d'intensité (ICLD) entre les différents segments d'un arbre microphonique

Pour remplir ces 140°, nous avons choisi un angle de prise de son de 40° pour notre couple arrière Lsr/Rsr. Nous recherchons alors une couverture de 50° pour nos deux segments latéraux arrières. La Figure 2.6 nous indique pour cette valeur que nos microphones latéraux Lss/Rss doivent être espacés de ($101.3 \times 2 =$) 202.3cm, ce qui est proche des 198 cm de l'écartement Ls/Rs de notre arbre 5.0 (Figure 2.5). Nos deux plans qui sont

complémentaires aboutissent finalement à la configuration présentée en Figure 2.7.

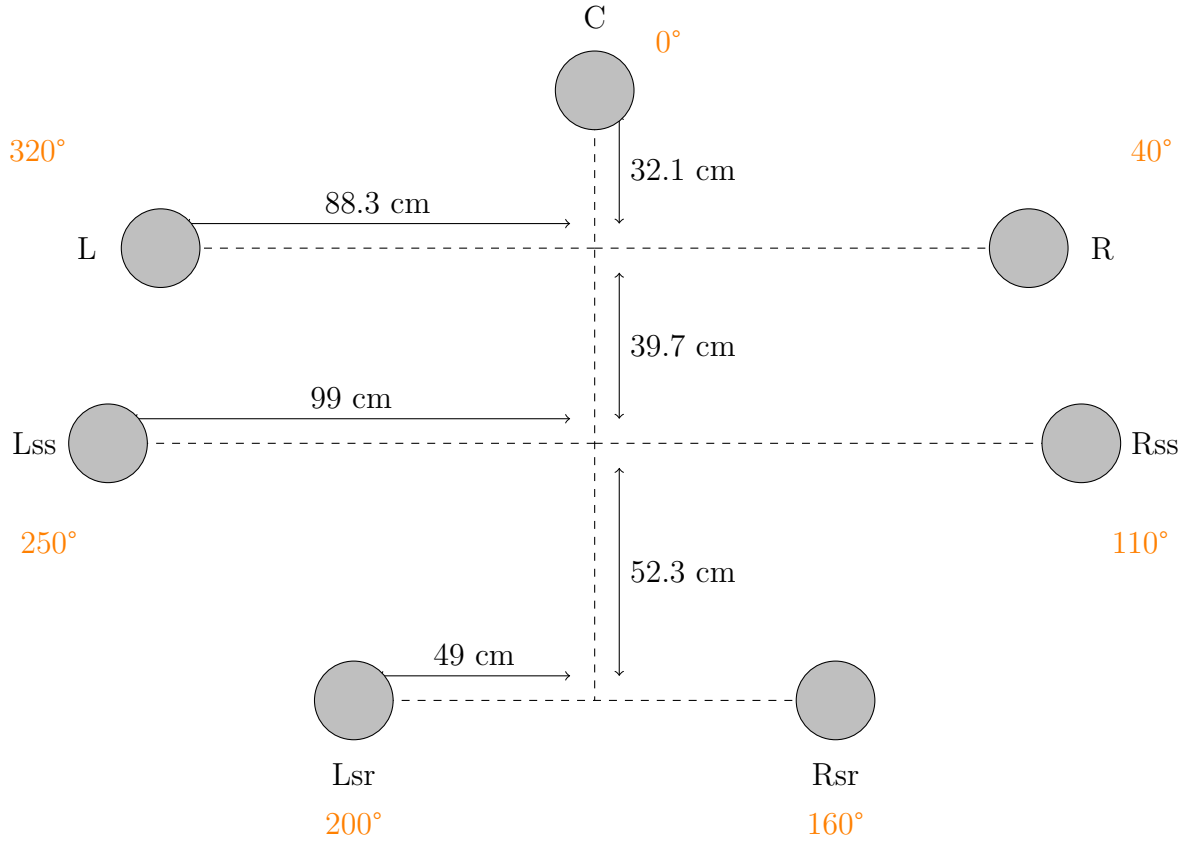


Figure 2.7 : *Arbre multicanal à sept microphones cardioïdes, créé à partir des plans complémentaires en Figure 2.5 et Figure 2.6*

Dans le but de valider nos hypothèses, sur la base des recherches présentées en 1.5.3, nous avons effectué quelques analyses en Figure 2.8 sur l'outil MARRS² développé par l'APL³ de l'université d'Huddersfield basé en Grande-Bretagne. Ces analyses ont été menées pour une source positionnée à un mètre du système de captation virtuel, sans inclinaison en élévation.

²*Microphone Array Recording and Reproduction Simulator*, se traduit en français par un simulateur d'enregistrement et de reproduction de système de captation microphonique

³*Applied Psychoacoustics Lab*

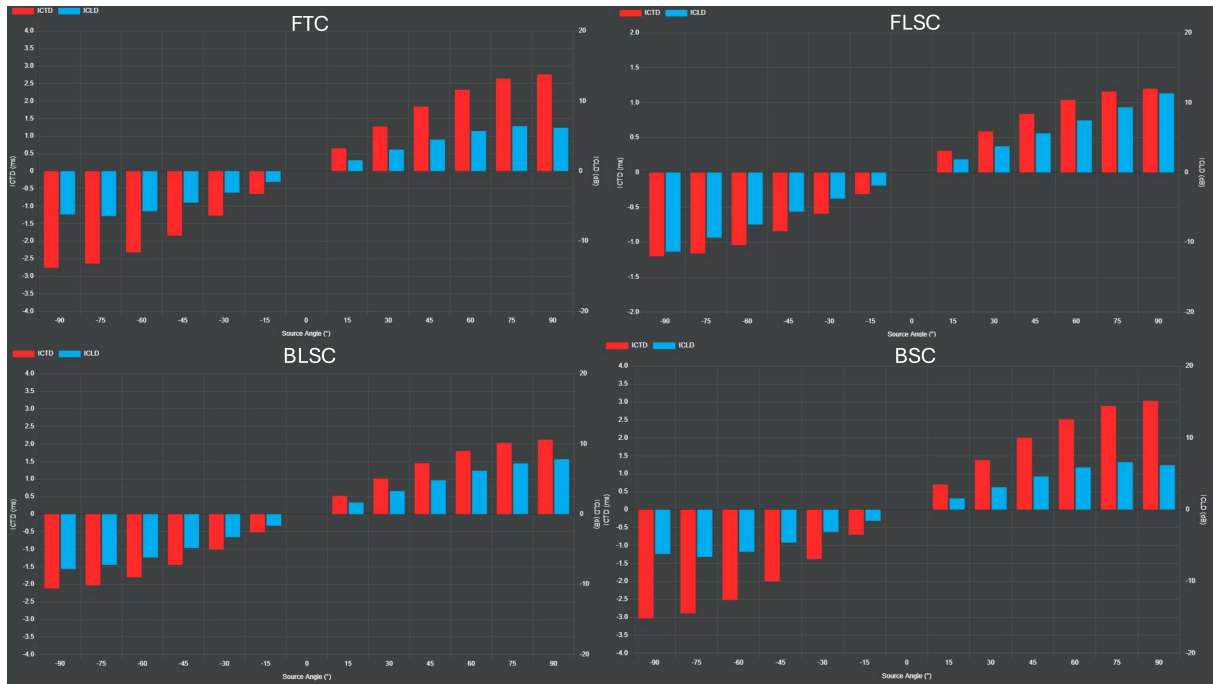


Figure 2.8 : Différences intercanales de temps en rouge (ICTD) et d'intensité en bleu (ICLD) entre les différents segments d'un arbre microphonique

Chacun des graphiques en Figure 2.8 montre les variations à la fois d'ICTD⁴ (en rouge) et de d'ICLD⁵ (en bleu) en fonction de l'angle de la source par rapport à un segment de l'arbre donné. Voici une analyse séparée des différents segments :

- **FTC** (*Front Triplet Coverage*) : L'ICTD atteint une valeur maximale de 2.76 ms, ce qui est assez élevé contrairement à l'ICLD qui n'atteint que 6.42 dB. Ce triplet repose donc bien sur une prédominance de différence de temps entre ces canaux.
- **FLSC** (*Front Lateral Segment Coverage*) : L'ICLD atteint cette fois-ci des valeurs plus élevées jusqu'à 11.34 dB pour une source positionnée à 90° de notre axe vertical (0°). L'ICTD est moins important pour ce segment (à un maximum de 1.2 ms). Ce segment se repose un peu plus sur les différences d'intensités entre ces canaux tout en ayant une présence de différences de temps.
- **BLSC** (*Back Lateral Segment Coverage*) : L'équilibre entre ICTD et ICLD est assez constant : au maximum, nous avons 2.12 ms de différences de temps intercanales et 7.83 dB de différences d'intensités intercanales.
- **BSC** (*Back Segment Coverage*) : Ce dernier segment repose comme pour le cas FTC

⁴Interchannel Time Differences, se traduisant par différences intercanales de temps

⁵Interchannel Level Differences, se traduisant par différences intercanales d'intensités

sur une prédominance de l'ICTD avec un maximum de 2.88 ms et 6.5 dB pour l'ICLD.

Le triplet microphonique frontal met clairement en avant des différences temporelles, et suggère que la localisation repose majoritairement sur les différences de temps. Cette configuration tend à équilibrer temps et intensité sur les segments latéraux et vise à se reposer sur une prédominance des différences de temps pour le segment arrière.

2.4.3 Prédominance des différences de temps sur les enceintes “*surround*”

Pour la réalisation de ce second arbre, nous supposons qu'un écartement important entre nos capsules frontales L/R et *surround* Lss/Rss nous permettrait d'obtenir la prédominance de différence de temps sur ce segment. Nous partons donc de la configuration en Figure 2.9 qui présente un écartement de $(\sqrt{((159.2/2) - 45.9)^2 + (152.1 - 26.5)^2}) = 130$ cm entre les capsules L/R et Ls/Rs et un écartement moins important entre nos capsules frontales de 53 cm. Cette disposition couvre 180° des 360° à remplir, il nous faut donc un arbre quadraphonique couvrant les 180° restant.

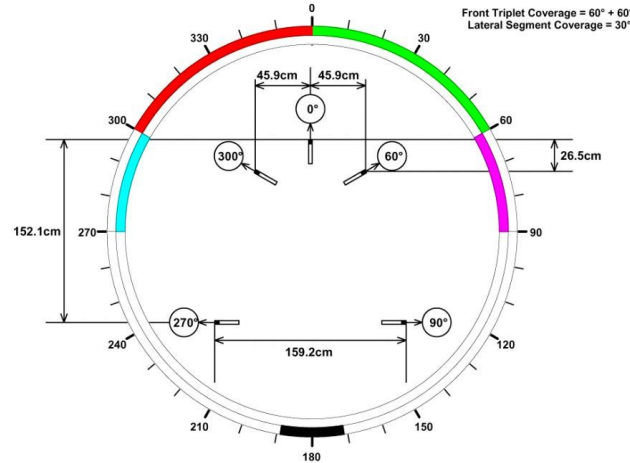


Figure 2.9 : Représentation d'un arbre multicanal à cinq microphones cardioïdes (WILLIAMS, 2004b)

Nous choisissons une couverture de 40° de notre couple Lsr et Rsr, nos segments latéraux arrières doivent donc former tous les deux une couverture de 70°. La Figure 2.6 nous indique pour ces valeurs que notre écartement Lss/Rss doit être de $(73.1 \times 2 =) 146.2$ cm, ce qui est relativement semblable au 159.2 cm de notre arbre 5.0 (Figure 2.9). L'arbre final est proposé en Figure 2.10.

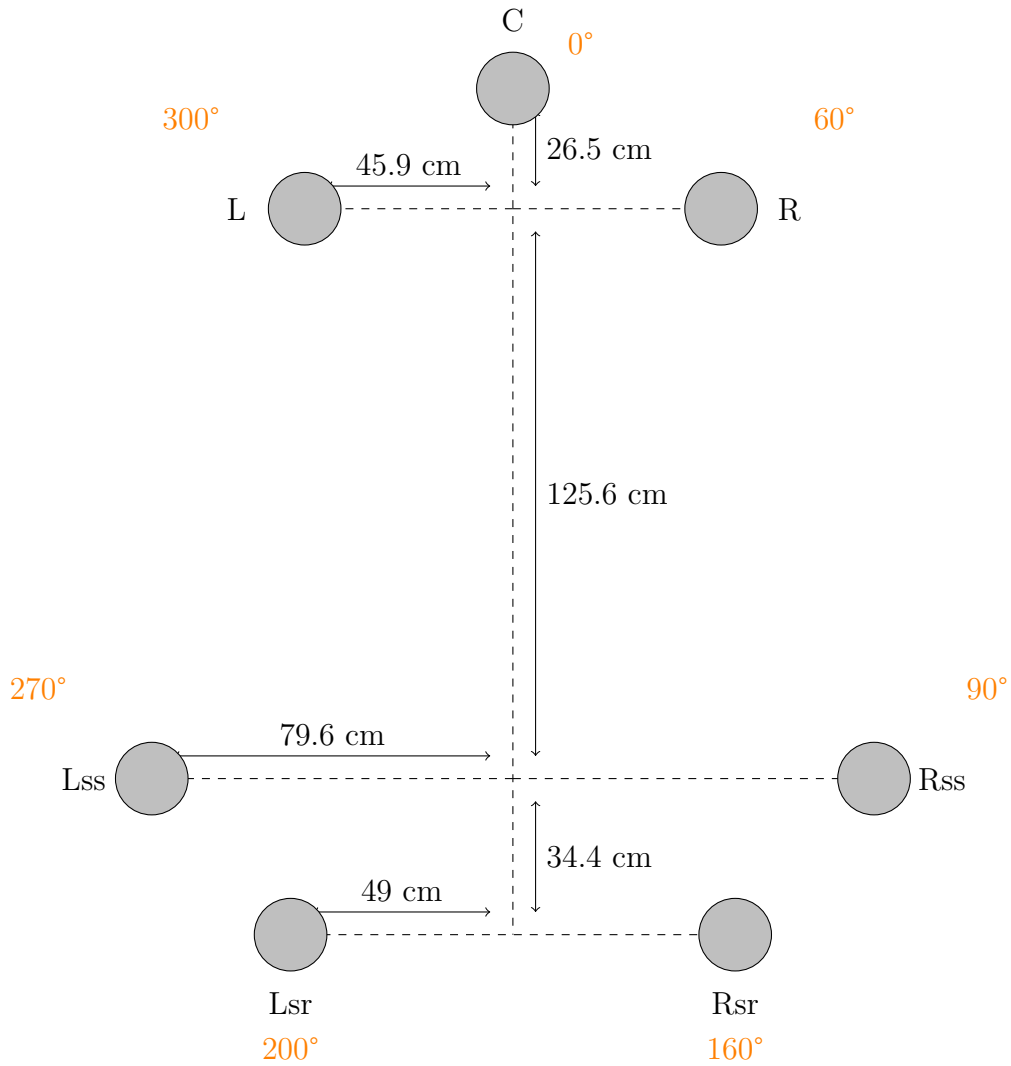


Figure 2.10 : *Arbre multicanal à sept microphones cardioïdes, créé à partir des plans complémentaires tiré de WILLIAMS, 2004b*

Comme précédemment, afin de confirmer nos hypothèses, voici en Figure 2.11 les graphiques correspondants aux variations à la fois d'ICTD (en rouge) et de ICLD (en bleu) en fonction de l'angle de la source par rapport à un segment de l'arbre donné.

Voici une analyse séparée des différents segments :

- **FTC (Front Triplet Coverage) :** L'ICTD atteint une valeur maximale de 1.56 ms, ce qui est assez faible contrairement à l'ICLD qui atteint 9.54 dB. Ce triplet repose donc sur une prédominance de différences d'intensités entre ces canaux.

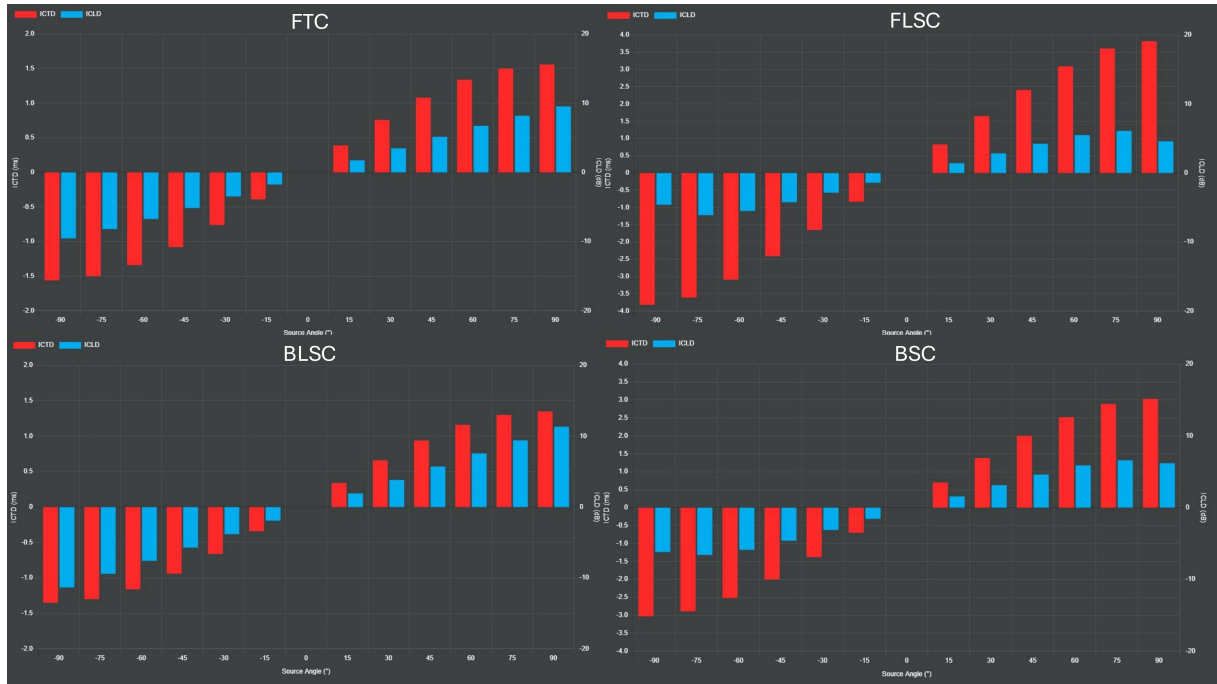


Figure 2.11 : Différences intercanales de temps en rouge (ICTD) et d'intensité en bleu (ICLD) entre les différents segments d'un arbre microphonique

- FLSC (*Front Lateral Segment Coverage*) : L'ICTD atteint cette fois-ci des valeurs plus élevées jusqu'à 3.82 ms pour une source positionnée à 90° de notre axe vertical (0°). L'ICLD est beaucoup moins important pour ce segment (à un maximum de 6.11 ms pour une source positionnée à 75° de notre axe vertical). Ce segment repose donc bien sur les différences de temps entre ces canaux avec des valeurs assez élevées.
- BLSC (*Back Lateral Segment Coverage*) : Pour ce segment, nous pouvons observer une prédominance de différences d'intensités où l'ICLD atteint des valeurs importantes jusqu'à 11.34 dB. En ce qui concerne l'ICTD, ses valeurs restent faibles et avoisinent généralement la milliseconde.
- BSC (*Back Segment Coverage*) : Ce dernier segment est le même que pour notre arbre précédent, il se repose donc sur une prédominance de l'ICTD avec un maximum de 2.88 ms et 6.5 dB pour l'ICLD.

Les segments latéraux avant présentent bel et bien une prédominance de différences de temps entre leurs canaux, contrairement aux segments latéraux arrière qui repose davantage sur l'ICLD. Lors de la conception de l'arbre, nous avons été bloqués par le nombre d'angles de prise de son à couvrir. Nous avons été amenés à faire des compromis sur la répartition de nos capsules. Ainsi, cela a induit des variations dans la prédominance de l'ICTD ou de l'ICLD.

2.5 Réflexion autour d'une interface graphique

La question de l'interface graphique a constitué un sujet d'intérêt dans notre travail de recherche, que nous essaierons de développer au mieux dans cette partie en nous appuyant sur le questionnaire du test pratique et des besoins que nous jugerons pertinents. Nous proposerons ainsi une esquisse de l'interface graphique car les possibilités en termes de personnalisation de l'interface sont limitées dans *Max*.

2.5.1 Quel contrôle souhaite-t-on accorder à l'utilisateur ?

Notre outil de spatialisation, comme tout autre outil, doit être intuitif dès sa première utilisation et répondre directement aux besoins de l'utilisateur ou l'utilisatrice. Pour cela, nous devons proposer une interface dans laquelle tous les paramètres seront à la fois clairs et faciles à appréhender.

Suite aux résultats des différents tests, nous avons pu recueillir que les différents mécanismes perceptifs étaient intéressants pour des effets esthétiques divers (4.1.2.4). Je trouverais dommage de ne pas permettre à l'utilisateur ou l'utilisatrice de choisir parmi la multitude de paramètres offerts par cet outil, afin de réaliser des effets esthétiques variés. Nous devons alors réfléchir à une méthode pour rendre facilement assimilable cette multitude de paramètres. La notion de *presets*⁶ nous intéresse beaucoup.

2.5.2 Présentation des réglages

2.5.2.1 Mécanisme perceptif

Afin de spatialiser une source sonore, notre outil proposera donc trois cas différents que nous avons déjà pu aborder précédemment. Un cas se basera uniquement sur des différences temporelles, un second sur des différences d'intensités, et un dernier combinera les deux.

En ce qui concerne l'unité de calcul des retards, un facteur multiplicateur a été mis en place afin d'accentuer plus ou moins les lignes de délais. Cette idée est née d'une simple curiosité et s'est révélée prometteuse lors de nos tests pratiques. Ce paramètre a été testé sur des sources musicales continues, comme la tenue d'une note d'un synthétiseur, et a plu à nos utilisateurs et utilisatrices en raison de la profondeur qu'il apportait et l'élargissement de la source.

⁶Ensemble de réglages prédéfinis et enregistrés

2.5.2.2 Configuration de l'arbre

La configuration de l'arbre serait libre avec la possibilité de choisir le nombre de capsules virtuelles afin de réaliser un couple stéréophonique, un arbre 5.0, 7.0 ou une configuration plus élevée si désirée. Nous laisserions également le choix de la directivité des microphones, où le symbole du diagramme polaire correspondant pourrait être affiché de manière graphique sur les capsules en fonction de leurs positions angulaires. Un retour précieux, d'un des sujets du test pratique, suggérerait d'appliquer une grille en fond avec une échelle en mètre afin de mieux se rendre compte de la taille du système de captation virtuelle.

Ce serait également intéressant de proposer des arbres prédéfinis dans notre base de données *presets*. Plusieurs configurations pourraient être créées, comme réalisées en 2.4. Nous pourrions proposer des arbres à prédominance de différences d'intensité sur les enceintes frontales, des arbres 5.0 spécifiques, un arbre Decca à trois ou cinq capsules, un double ORTF et bien d'autres encore abordés en 1.3.1.2.

2.5.3 Exploration de l'apparence de l'interface

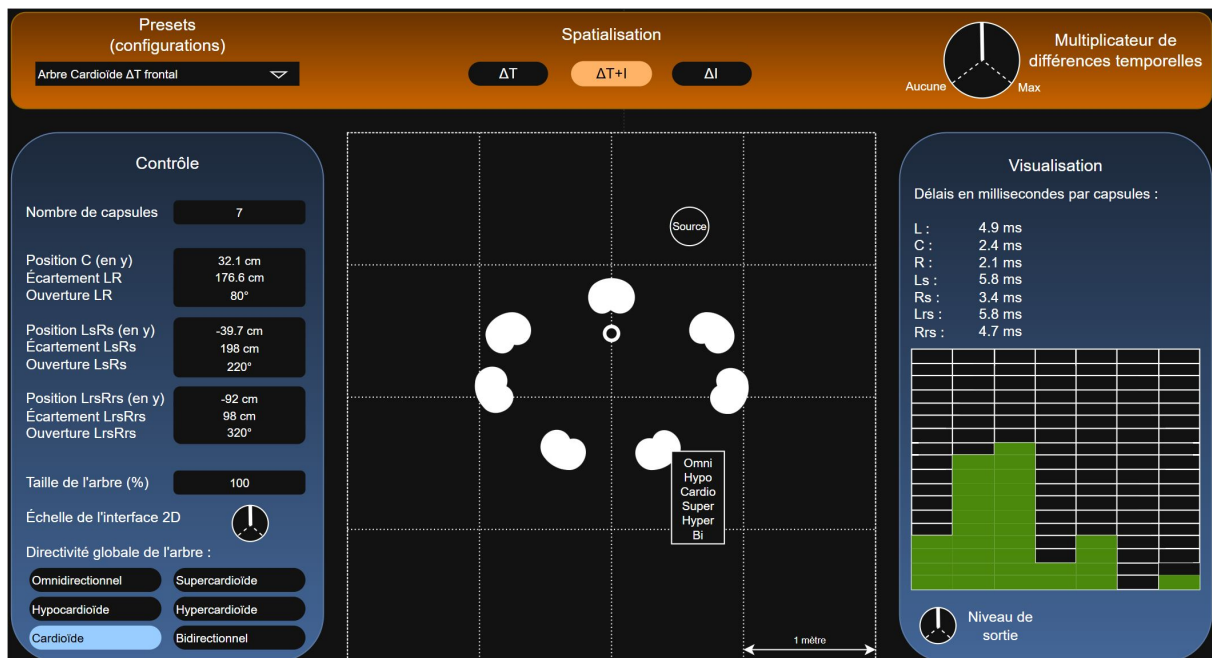


Figure 2.12 : Proposition d'une interface graphique

Suite à ces réflexions et aux résultats du test pratique (4.2), nous avons réalisé une esquisse graphique d'une interface utilisateur (Figure 2.12).

L'interface s'organise donc en quatre parties :

- La partie supérieure comprend trois contrôles qui sont primordiaux afin d'obtenir rapidement la spatialisation recherchée. L'onglet *Presets (configurations)*, représenté en Figure 2.13, comporte un panel de configurations d'arbres prédéfinies comme l'arbre Decca, la croix IRT, le double ORTF, le couple AB ou différentes configurations obtenues à l'aide des abaques de Williams et bien d'autres encore (1.3.1.2 et 2.4). Au centre, nous avons le choix de spatialiser de trois manières différentes, nous pouvons restituer spatialement notre source à l'aide de différences temporelles pures (ΔT), de différences d'intensités pures (ΔI) où les deux combinés ($\Delta T + I$). Il convient de préciser que le paramètre ΔI est relié à l'équation (7) et est indépendant de la directivité des microphones. Le dernier paramètre est un facteur multiplicateur directement relié à notre unité de calcul de retards.

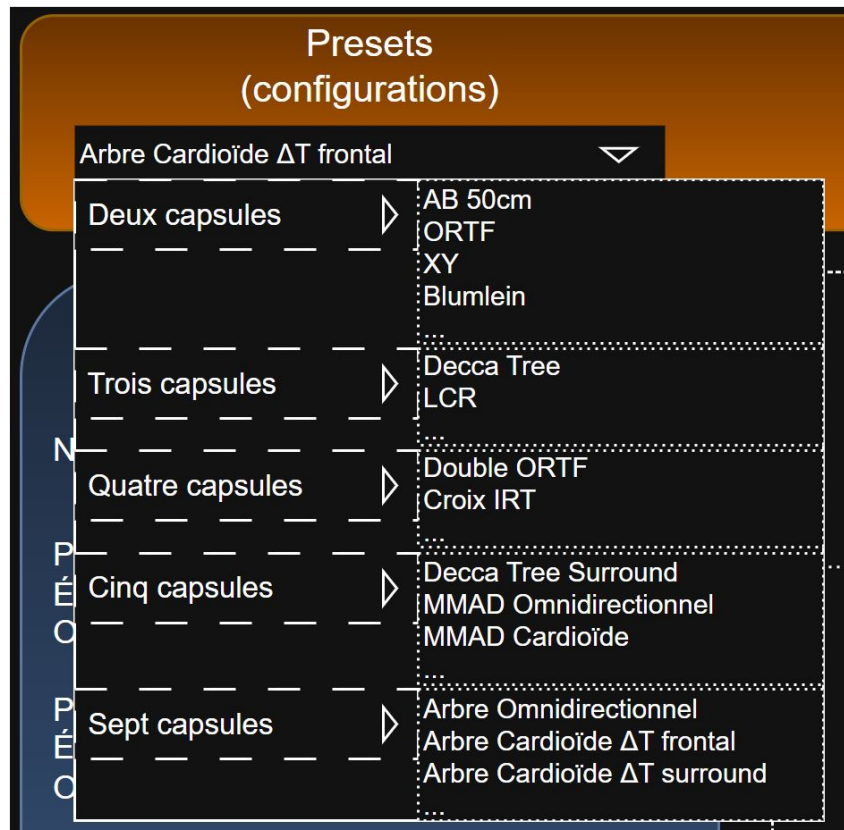


Figure 2.13 : Onglet “Presets (configurations)” de l’interface graphique

- La partie centrale représente une vue de dessus de notre arbre microphonique. Chaque capsule peut y être positionnée individuellement, et un clic droit permet d'attribuer une directivité spécifique à chacune. Une grille en arrière-plan, graduée en mètres, permet de mieux visualiser les dimensions de l'arbre microphonique.

- La partie située à gauche concerne la configuration personnalisée de notre arbre. Nous avons accès au nombre de capsules ainsi qu'à leurs positions et ouvertures. Il est possible d'agrandir la taille de notre arbre ainsi que de changer la directivité globale de notre arbre. Le paramètre *Échelle de l'interface 2D* permet d'ajuster la visualisation de la partie centrale.
- La dernière partie placée à droite est un onglet de visualisation où l'on observe les délais appliqués entre chaque microphone et les niveaux correspondants. Un potentiomètre de niveau est placé en sortie afin de permettre un réglage précis du volume.

Partie III

Méthodologie pour un protocole de tests expérimentaux

Nous avons réalisé deux tests, un premier purement perceptif qui nous a permis de déterminer l'intérêt des mécanismes perceptifs dans notre perception à partir d'un système multicanal ; et un second dans lequel nous avons invité des mixeurs, mixeuses, monteurs et monteuses son à utiliser notre outil dans des conditions de postproduction cinéma. Le premier test a permis d'évaluer l'utilité des différents mécanismes perceptifs du point de vue de la perception humaine et le second de déterminer la pertinence de l'outil.

3.1 Test perceptif : évaluation de l'impact des mécanismes perceptifs à travers un système multicanal

3.1.1 Paramètre et perception spatiale

3.1.1.1 Indices de localisation

Nous avons évoqué en première partie que l'utilisation des différences de temps en plus de celle d'intensité pourrait présenter une meilleure précision spatiale (1.5.3). L'objectif de ce test était de déterminer si cette combinaison était véritablement bénéfique à notre perception lorsque nous recréons un espace sonore multicanal à travers un système physique.

Lors de la réalisation de notre outil, nous avons veillé à ce qu'il permette d'utiliser indépendamment les différences de temps (1.2.2) et d'intensité (1.2.1) ou de les utiliser ensemble (1.5.3). Cela nous permettait alors de comparer la localisation d'une source sonore à partir des différentes combinaisons de mécanismes perceptifs suivants :

- Différence de temps seule
- Différence d'intensité seule
- Différence de temps et d'intensité combinés

Nous avons alors invité des personnes à réaliser un test perceptif. Ce test a été réalisé dans l'entreprise Polyson, dans l'une de leurs salles équipées d'un système multicanal à sept canaux discrets, car notre outil s'articule autour d'une simulation d'un arbre multi-microphonique à sept capsules. Pour ne pas trop fatiguer nos auditeurs et auditrices, nous avons essayé de réaliser ce test sur quelques sources choisies soigneusement.

3.1.1.2 *Localization et Locatedness*

Blauert a pu définir deux termes ; *Localization*¹ et *Locatedness*¹ dans la perception auditive, deux termes que je pense intéressants à prendre en compte dans la réalisation de nos différents tests (BLAUERT, 1996) :

<i>Localization</i>	<i>Locatedness</i>
Identification de la position d'une source	Jugement de la clarté et de la stabilité d'une image sonore
Direction et distance	Précision et stabilité de l'image sonore
Processus perceptif actif	Qualité subjective perceptive
Dépend des indices binauraux et spectraux	Dépend de l'environnement acoustique et des artefacts

Tableau 3.1 : *Tableau comparatif de deux termes qui caractérisent la perception sonore selon BLAUERT (1996)*

Pour clarifier ce principe, *Localization* est un processus qui définit la localisation d'un son, contrairement à *Locatedness* qui est un état et pourrait se définir par une sensation de localisation. Notre test a été en permanence tiraillé entre ces deux termes. En effet, le testeur ou la testeuse a fait appel à ses capacités perceptives mais a été également perturbé par l'environnement dans lequel il ou elle se trouvait, dans notre cas une salle équipée d'un système multicanal 7.0. N'oublions pas que ce premier test se basait principalement sur la *Localization* défini par Blauert plutôt que sur la stabilité de l'image sonore dans un environnement (*Locatedness*), car son objectif

¹Les deux termes utilisés par Blauert se traduisent tous deux en français par "localisation", bien qu'ils aient des nuances distinctes dans leur signification originale

était d'évaluer l'impact des différents mécanismes perceptifs mis en place dans notre outil.

L'objectif de ce test était de déterminer à la fois la précision de localisation d'une source sonore et sa largeur perçue, à partir de différents mécanismes perceptifs mis en place. Nous cherchions à comprendre comment chaque mécanisme, différence de temps, d'intensité ou la combinaison des deux, influait sur la perception de la position de la source et de son étendue à travers un système multicanal.

3.1.1.3 Choix de la configuration de l'outil

Après avoir installé *Max* dans l'auditorium de mixage de l'école, nous avons pu tester l'outil dans des conditions réelles de mixage. Ce qui nous a le plus marqués, c'est qu'à partir de plusieurs configurations de la simulation d'arbre, il était possible d'obtenir diverses sensations de spatialisation. Nous nous sommes demandés alors quelle configuration adopter pour ce test perceptif (nous avons détaillé quelques configurations possibles en 2.4).

Après réflexion, nous avons d'abord envisagé de couvrir l'ensemble de l'espace de prise de son sans laisser de zones non captées, en exploitant les différents angles de captation de notre arbre (voir 1.3.1.2). Lors de nos tests personnels, nous avons constaté la présence de creux dans l'image sonore qui correspondent aux zones non captées des angles de prises de son choisis, c'est un élément sur lequel nous devons porter une attention particulière. Ensuite, nous avons pu observer que la localisation était confuse entre les canaux L, R et Ls, Rs en utilisant la configuration initiale, nous avons donc fait attention à adopter une meilleure configuration pour ces canaux qui sont diffusés par les enceintes les plus éloignées de l'auditorium de l'école.

Lors de la réalisation de ce test, le choix de la directivité de nos capsules virtuelles nous a paru essentiel. Notre choix s'orientait entre deux types de capsules pour ce test : omnidirectionnelles ou cardioïdes, qui sont dans un contexte de prise de son un choix majeur, et ne permettent absolument pas de capter l'environnement sonore de la même manière. Une approche dite "cardioïde" nous permettrait de dissocier clairement les différentes sources de la scène sonore, alors qu'une approche "omnidirectionnelle" susciterait un sentiment de cohérence d'un espace sonore entre ces différentes sources, en d'autres termes une "fusion" au détriment de la clarté. Tout d'abord, il nous a semblé pertinent de présenter d'emblée le test avec des différences d'intensité appliquées à

l'ensemble des mécanismes à tester. Cette différence serait inévitablement induite par le choix d'une directivité directionnelle comme le cardioïde. Après réflexion, il nous est apparu nécessaire d'opter pour une directivité omnidirectionnelle, de manière à éliminer, en amont, toute différence d'intensité, et ainsi permettre une comparaison plus rigoureuse entre les différences de temps et d'intensité pures.

Cependant, nous nous sommes permis d'ajouter un quatrième mécanisme : il nous permettrait d'utiliser davantage les différences d'intensités et seraient la combinaison des différences de temps et d'intensité en simulant en plus un système de captation cardioïde.

Il ne nous reste plus qu'à choisir les écartements et l'ouverture entre nos capsules omnidirectionnelles virtuelles. Le choix de l'ouverture n'est utile que pour un mécanisme perceptif : lorsqu'on simule un système de captation cardioïde, il repose sur des différences de temps et d'intensité. Dans le cas d'un système qui utilise des microphones omnidirectionnels, l'ouverture n'a pas d'incidence. L'arbre que nous avons utilisé dans le cadre de ce test en choisissant des microphones omnidirectionnels est détaillé en 2.4.1.

3.1.1.4 Sources sonores

J'ai eu l'occasion de pouvoir discuter du choix des sources sonores à utiliser dans le cadre de ce test avec plusieurs personnes, notamment Cyril Holtz et Michael Williams, et la question de la voix est ressortie plusieurs fois dans nos conversations. En effet, la voix est une source familière à l'oreille humaine dont les mécanismes perceptifs sont exploités de manière optimale (BOTTE *et al.*, 1988).

■ *La perception de la parole met en oeuvre des mécanismes qui lui assurent une grande robustesse face à un environnement acoustique hostile. Une part importante de ces mécanismes semblent être de nature auditive (ils sont également utilisés dans la perception de signaux sonores non linguistiques). Le signal de parole, lui-même redondant, semble être particulièrement bien adapté à l'exploitation maximale de ces mécanismes auditifs.* ■

– Christel Sorin, d'après (BOTTE *et al.*, 1988).

L'un des mécanismes de compensation et de robustesse perceptive auxquels fait référence cette citation est la redondance de la parole. Celle-ci présente de nombreux indices acoustiques redondants, comme la prosodie qui correspond aux fluctuations rythmiques, dynamiques et mélodiques d'un son, en d'autres termes concernant la voix, à son rythme,

son accentuation et son intonation. Nous avons donc choisi d'utiliser de la voix parlée captée en proximité, due à son importance dans notre système perceptif et à sa place centrale dans une bande sonore au cinéma.

■ *On peut donc supposer que les signaux d'entrée de l'oreille à large bande peuvent contenir davantage d'informations sur la position de la source sonore que les signaux à bande étroite.* ■

– BLAUERT, 1996.

Concernant le choix des autres sources, nous voulions sélectionner de courtes sources riches en harmoniques, à large bande, qui sont plus localisables que des sources continues et pauvres en harmonique. Des sources comme une guitare, un clavecin, un marimba ou une darbouka¹ étaient à prendre en compte par leur richesse en harmoniques et leur attaque franche. Notre choix s'est donc porté sur la guitare rythmique qui est une source riche en harmonique et mélodique, la darbouka qui nous permettra d'évaluer nos mécanismes sur une source percussive et la voix parlée pour la réalisation de notre test.

Chaque stimulus a été entendu d'une manière statique par nos quatre combinaisons de mécanismes perceptifs traitées auparavant en 3.1.1.1 et 3.1.1.3. Ces stimuli ont été présentés deux fois, de façon non consécutive. Les stimuli ont été diffusés pendant 30 secondes et une dernière écoute a été proposée où les quatre cas s'enchaînaient sur une durée de 30 secondes. L'ensemble du test s'étendait donc sur une vingtaine de minutes, une durée qui restait confortable auditivement pour nos auditeurs et auditrices.

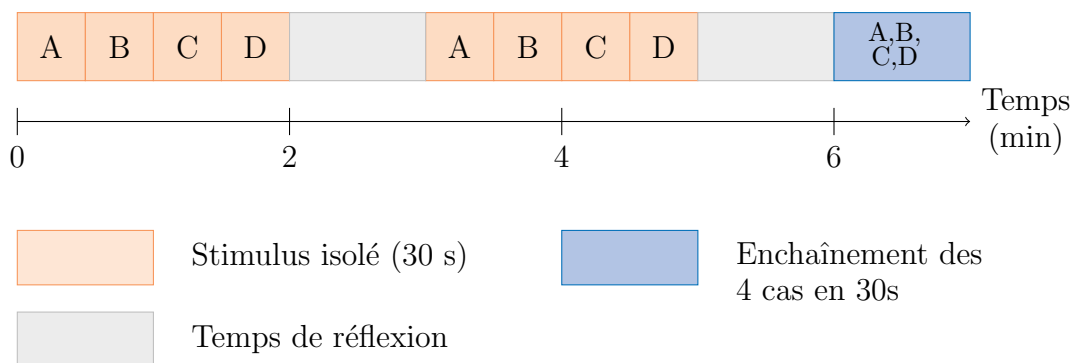


Figure 3.1 : Représentation temporelle du déroulé du test perceptif pour une source sonore

Les extraits sonores ont été volontairement limités à 30 secondes pour plusieurs raisons. Nous pensions que ce format était adéquat pour que les participantes et participants

¹Percussion traditionnelle du Moyen-Orient

puissent évaluer les critères perceptifs de manière efficace. De plus, ce test nécessitait une attention particulière qui pouvait causer de la fatigue auditive, ainsi 30 secondes représentaient une durée qui reste agréable.

3.1.2 Mise en place du test

3.1.2.1 Environnement de l'expérience

Ce test a été réalisé le jeudi 17 et vendredi 18 avril dans la société Polyson Post Production dans l'une de leur salle équipée en 7.0.2 (Figure 3.2). Cette salle est équipée d'un système composé de trois enceintes frontales *Genelec Model S30C*, de quatre enceintes *Tannoy Precision 8P* et de deux enceintes en élévation que nous n'utiliserons pas pour ce test. Notre logiciel Max communiquait avec ce système par l'intermédiaire d'une interface audio *Scarlett 18i20 4th Generation*.



Figure 3.2 : *Salle de réalisation du test*

Nos sept capsules virtuelles ont donc été assignées aux sept canaux correspondants : la capsule virtuelle gauche a été assignée à l'enceinte frontale gauche, la capsule virtuelle centre à l'enceinte centrale avant, et ainsi de suite pour les autres capsules et leurs enceintes respectives.

3.1.2.2 Répartition des sources et mécanismes perceptifs par session

Chaque participante et participant ne prenait part qu'à une seule session. Nous avons donc fait le choix de mélanger l'ordre des sources et des mécanismes perceptifs par session

afin d'éviter d'une part, que les sujets ne fassent des associations systématiques entre une source particulière et un mécanisme donné, et d'autre part, que les résultats ne soient pas influencés par les caractéristiques propres à une source. Ce mélange contribuait à réduire l'impact de l'ordre de présentation et garantit une évaluation plus neutre et objective.

Nous avons donc dressé un tableau (3.2) référençant l'ordre de présentation des sources sonores et des mécanismes associés selon les conditions suivantes :

- Répartir les sources sonores de manière différente pour chacune des trois sessions
- Répartir les mécanismes perceptifs différemment pour chaque source au sein de chaque session

Les mécanismes sont représentés de la manière suivante :

- “**T**” désigne la spatialisation d'une source sonore utilisant uniquement des différences de temps (ΔT)
- “**I**” désigne la spatialisation d'une source sonore utilisant uniquement des différences d'intensité (ΔI)
- “**TI**” désigne la spatialisation d'une source sonore combinant différences de temps et d'intensité ($\Delta T + \Delta I$)
- “**TIC**” désigne la spatialisation d'une source sonore combinant différences de temps et d'intensité ($\Delta T + \Delta I$) en simulant un système de captation cardioïde

Session	Instrument	1	2	3	4
A	Voix parlée	T	I	TI	TIC
	Guitare	I	TI	T	TIC
	Darbouka	TI	T	TIC	I
B	Darbouka	T	TIC	I	TI
	Voix parlée	I	TI	TIC	T
	Guitare	TI	T	I	TIC
C	Guitare	T	I	TIC	TI
	Darbouka	I	TI	T	TIC
	Voix parlée	TIC	T	I	TI

Tableau 3.2 : Tableau de répartition des sources sonores et des mécanismes perceptifs par session

3.1.2.3 Positionnement des sources dans l'outil

L'objectif de ce test était également d'analyser la capacité à localiser une source sonore à travers un système "surround", qui présente des canaux arrière. Nous avons ainsi choisi de répartir spatialement nos trois sources sonores de manière différente. Idéalement, chaque source aurait dû être testée dans trois positions distinctes afin de pouvoir comparer rigoureusement la source à travers une spatialisation donnée. Toutefois, afin de limiter la durée du test et d'éviter une fatigue auditive à nos participants et participantes, nous avons opté pour une spatialisation fixe, en associant chaque source à une position spécifique.

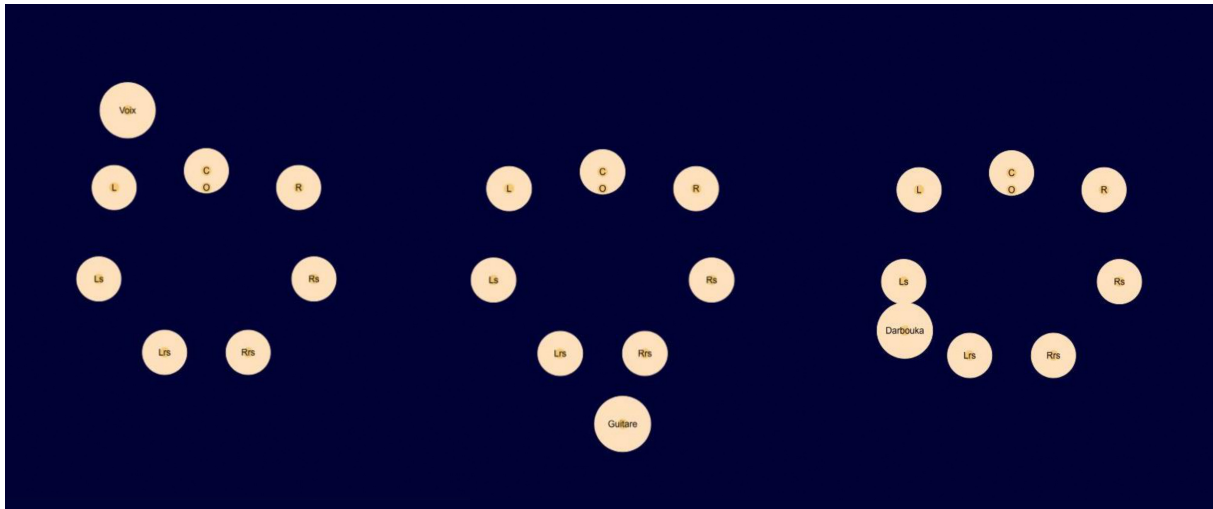


Figure 3.3 : Valeurs de panoramisation appliquées à nos différentes sources dans l'ordre suivant : Voix, Guitare et Darbouka à travers l'interface graphique de l'outil

Voici en Figure 3.3 les différentes positions de nos sources placées dans l'interface graphique de notre outil représentant un espace virtuel à deux dimensions où notre arbre multi-microphonique est simulé. Nous avons donc choisi de placer nos sources de la manière suivante :

- La voix est placée entre les capsules L et C afin de recueillir des données concernant la localisation frontale.
- La guitare est placée à l'arrière de notre simulation d'arbre axé davantage vers la droite.
- Pour finir, la darbouka est placée dans un segment latéral de notre arbre afin de pouvoir rassembler des données sur l'entièreté de la couverture spatiale de notre outil.

3.1.3 Protocole de test

3.1.3.1 Choix des participants, participantes et répartition par groupe

Les tests perceptifs sollicitent une écoute particulière qui demande à un sujet d'être analytique sur ce qu'il perçoit auditivement. Cette écoute critique est davantage développée chez les personnes ayant une expérience plus ou moins importante dans le domaine du son. Afin d'obtenir des résultats rigoureux, nous avons donc restreint ce test à des personnes qui sont familières au domaine du son (étudiants, amateurs et professionnels du son).

Nous avons tout d'abord sollicité des personnes potentielles à l'établissement ENS Louis-Lumière, des étudiants et des enseignants qui pourraient être intéressés par ce test. Nous avons ensuite ouvert la participation au test à des anciens de l'école, à des étudiants d'autres écoles (La Fémis, la FSMSI²) ou professionnels du son.

L'expérience s'est déroulée sur deux jours et dix-sept sessions d'une heure chacune ont été réalisées. Par la taille de la salle et selon les disponibilités des volontaires, j'ai formé des groupes d'une à quatre personnes, afin de les répartir dans la salle et d'obtenir un maximum de résultats en fonction de leur emplacement. Chaque personne était assigné à l'une des trois sessions (A, B ou C), en veillant à répartir équitablement le nombre de participantes et participants entre eux.

3.1.3.2 Réalisation du questionnaire

Les participants et participantes étaient invités à remplir le questionnaire d'une certaine manière (Annexe E : Questionnaires des différents tests) pendant ou après l'écoute de notre source sonore spatialisée. Au final, la personne aura rempli le questionnaire quatre fois pour chaque source, soit douze fois au total, ce qui correspond à l'écoute des quatre restitutions spatiales pour chacune des trois sources sonores.

Le questionnaire comprend trois questions, interrogeant chacune un critère spécifique. Les questions Q1 et Q3 utilisent une échelle de Likert allant de 1 à 5. Cette méthode permet de convertir une réponse relative en nombre afin d'en déduire des statistiques lorsque nous exploiterons nos résultats. Les réponses de la question Q2 s'obtiendront de façon graphique.

²Formation supérieure aux métiers du son et de l'image du Conservatoire National Supérieur de Musique et de Danse de Paris

La question Q1 : “Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?” vise à évaluer la capacité des sujets à localiser la source sonore au sein d’un système sonore physique. La question Q2 : “Dans quelle direction avez-vous perçu la source ?”, se repose sur un schéma que les participantes et participants doivent annoter en coloriant la zone perçue comme origine du son, ce qui permet de déterminer la direction estimée de la source, qu’elle soit ponctuelle ou étendue. Enfin, la question Q3 : “Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?” a pour objectif d’évaluer la perception de l’étalement spatial de la source.

3.2 Test pratique : Utilisation de l’outil par des professionnels du son

Un second test a été réalisé avec des professionnels du son qui ont essayé l’outil. Mixeurs, mixeuses, monteurs et monteuses son ont donc été invités à manipuler l’outil afin de latéraliser les sources de leur choix. Nous avons laissé à chaque personne le choix des sources sonores et lui avons demandé pourquoi elle avait fait ce choix afin de comprendre ce qu’elle a cherché à tester, comparer ou trouver à l’aide de notre outil.

Ce test se base davantage sur la *Locatedness* puisqu’il repose sur les perceptions subjectives de nos invités. Nous avons cherché à juger de la clarté, de la stabilité de l’image sonore et de la pertinence de l’outil à travers ce test.

3.2.1 Mise en place du test

3.2.1.1 Environnements de l’expérience

Nous sommes allés à la rencontre de nos participants et participantes sur leurs lieux de travail, afin qu’ils et elles restent familiarisés avec leurs systèmes d’écoute, et par soucis de praticité. Nous sommes donc allés les rencontrer dans des salles de montage ou auditorium de mixage équipés d’un système multicanal 5.0 ou 7.0.

3.2.1.2 Préparation de notre interface graphique

Afin que nos testeurs et testeuses ne se perdent pas dans notre patch *Max*, nous avons réalisé une première interface graphique à l’aide du mode *Presentation* avec les paramètres suivants :

- écartements, rotation et choix des directivités de nos microphones virtuels

- possibilité de moduler les mécanismes perceptifs sollicités : ΔT pur, ΔI pur ou $\Delta T+I$
- un facteur multiplicateur lié à l'unité de calcul des retards

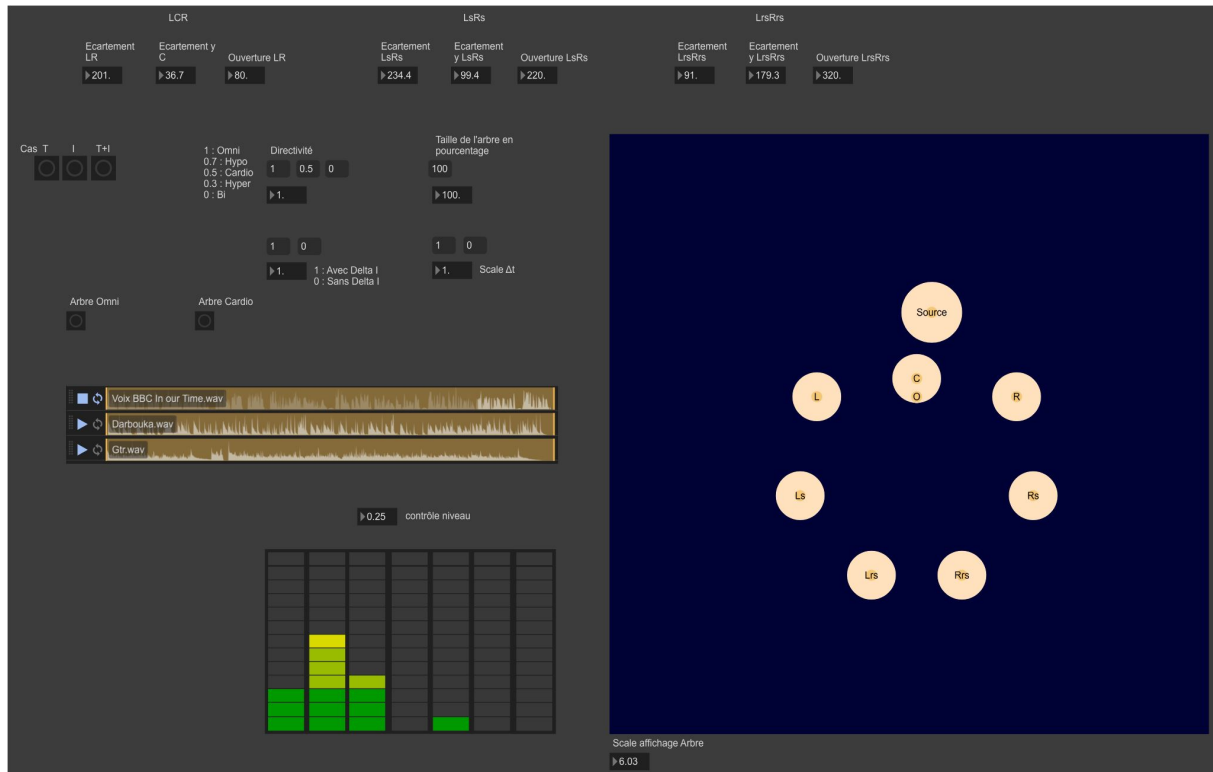


Figure 3.4 : Interface réalisée à l'aide du mode *Presentation* de *Max* en vue de réaliser un test

Nous avons laissé à disposition les sources sonores utilisées lors du test précédent à nos participants et participantes s'ils souhaitaient s'y référer ou les utiliser pendant l'expérience.

3.2.2 Protocole de test

3.2.2.1 Présentation de nos participants et participantes

Nous avons choisi un corpus de professionnels qui comprend quatre mixeurs et deux monteurs son. Voici la liste de nos participants et participantes :

- Olivier Guillaume : mixeur (OG)
- Alexandre Hecker : monteur son (AH)
- Félix Davin : monteur son (FD)

- Cyril Holtz : mixeur (CH)
- Lucas Le Néouanic : mixeur (LLN)
- Nathalie Vidal : mixeuse (NV)

3.2.2.2 Réalisation du questionnaire

Afin d'évaluer notre outil et la spatialisation engendrée par celui-ci, un second questionnaire (Annexe E : Questionnaires des différents tests) a été conçu afin de recueillir les retours qualitatifs de professionnels. Les objectifs principaux étaient de mesurer l'intuitivité et la pertinence perceptive de l'outil dans un contexte de montage et mixage multicanal et de mesurer son ergonomie. Ce questionnaire a donc été axé sur ces objectifs : ergonomie de l'interface, usage des paramètres et comparaison perceptive avec des outils professionnels existants. La dernière partie était consacrée aux remarques générales et suggestions d'amélioration.

Questions sur l'ergonomie :

- Est-ce que les paramètres de l'outil (déplacement de la source sonore, ajustement des paramètres, etc.) sont facilement accessibles ? Sont-ils suffisants, insuffisants ou trop nombreux ?
- Avez-vous essayé de changer les paramètres de l'outil (positionnement, orientation, directivité) ? Si oui, avez-vous trouvé cela intuitif ?

Questions perceptives :

- Avez-vous entendu des différences perceptives lorsque vous changiez les paramètres de l'arbre microphonique ? Si oui, quelles différences avez-vous perçues ?
- En comparant l'outil proposé à ceux que vous pourriez utiliser au quotidien, avez-vous perçu des différences perceptives notables ? Si oui, trouvez-vous que ces choix pourraient impacter significativement le résultat final d'un mixage ?

Et cette dernière question : “Quelles fonctionnalités ou améliorations souhaiteriez-vous voir ajoutées à l'outil ?”

Les questions ont été formulées de manière ouverte afin de permettre une expression libre des ressentis de nos participants et participantes. Ce choix visait à obtenir des résultats subjectifs afin de saisir la pertinence de l'outil d'un point de vue perceptif et pratique.

Partie IV

Exploitation des résultats

4.1 Test perceptif

4.1.1 Statistiques de l'échantillon

Nous avons pu recueillir les données de trente-six personnes réparties sur trois sessions différentes pour ce test (A, B et C). Deux critères ont été retenus pour évaluer la comparabilité des groupes : l'âge des participants et leur placement en salle (gauche, droite, arrière, centre) (Figure 4.1).

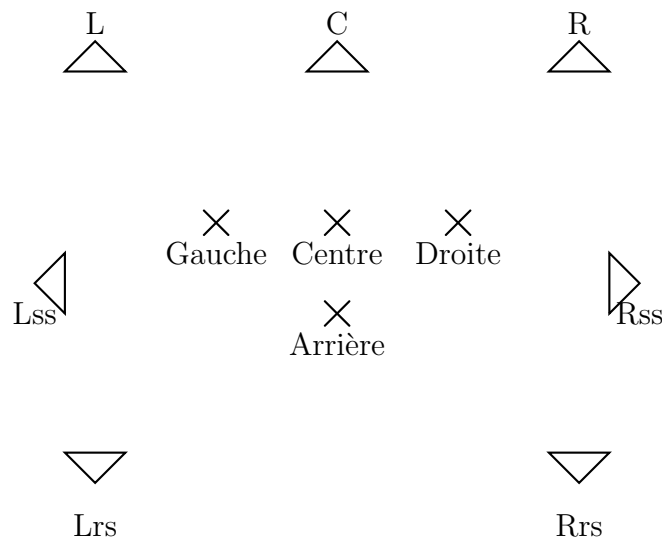


Figure 4.1 : *Représentation des différentes positions occupées par les participants dans la salle de test*

Toutes ces informations sont synthétisées ci-dessous (Figure 4.2). Les sujets ont en moyenne 24 ans, avec des âges compris entre 20 et 32 ans pour l'ensemble des participants et participantes. Les participants et participantes sont quasiment répartis de façon

homogène dans la salle, même s'il existe de petites différences entre les différentes sessions et qu'au total, plus de personnes ont passé ce test placées à gauche ou à droite.

Malgré ces variations, nous considérons les groupes suffisamment homogènes pour réaliser un ensemble de statistiques descriptives des réponses au questionnaire. Il convient toutefois de rester prudent sur l'interprétation de ces résultats compte tenu du nombre restreint de participants et participantes.

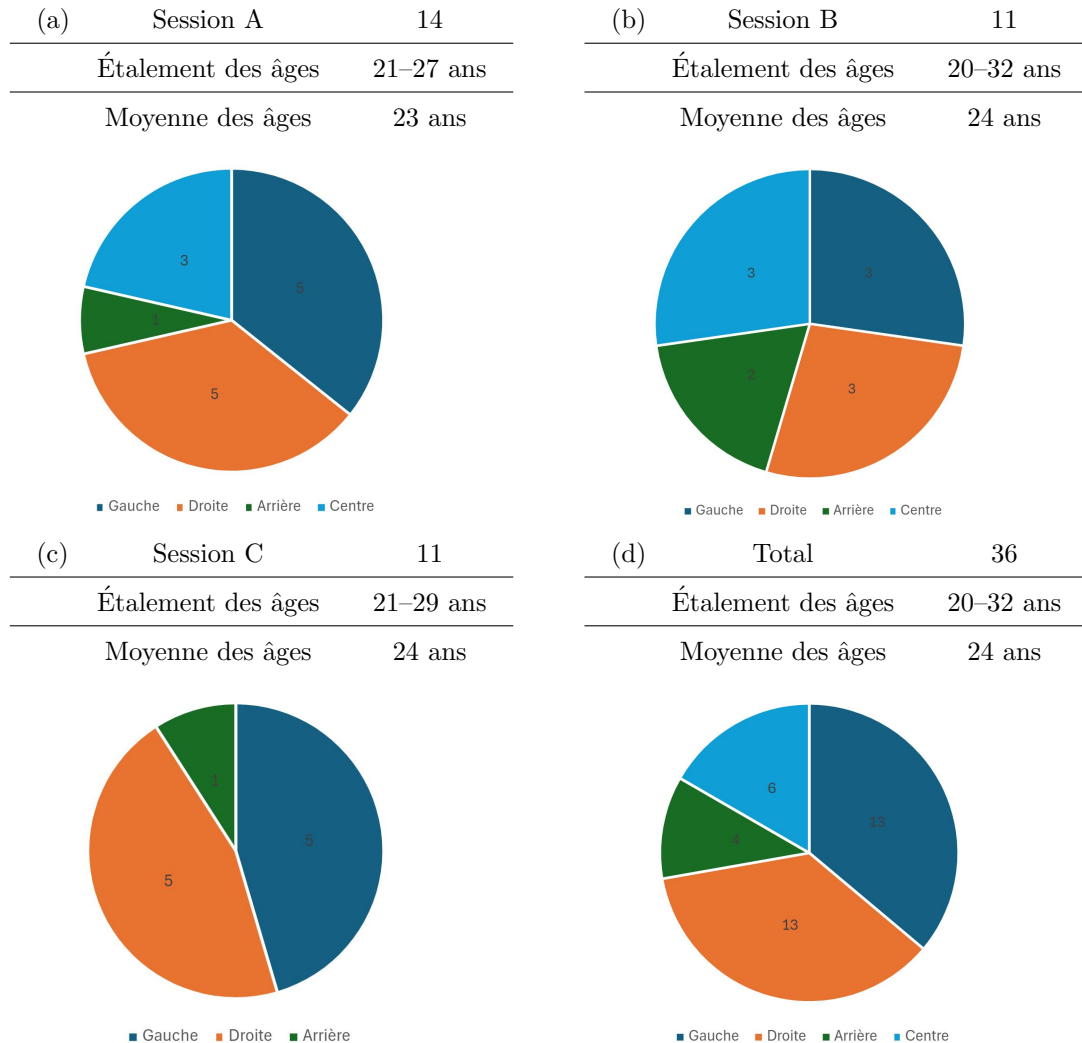


Figure 4.2 : Répartition et position des participantes et participants par session

Nous avons choisi d'analyser nos résultats par question. Pour les questions Q1 et Q3, nous représenterons nos résultats de manière graphique à l'aide du *Boxplot*, autrement dit à l'aide d'une “boîte à moustaches”. Cette représentation graphique permet de recueillir nos résultats sous une échelle de Likert allant de 1 à 5 et permet de mettre en valeur sous une même plage de données la répartition des valeurs, ainsi que de potentielles valeurs aberrantes (Figure 4.3).

Le *Boxplot* permet de mettre en forme un ensemble de données sous la forme d'une boîte encadrée par deux moustaches. Elle permet de mettre en évidence la valeur moyenne par une croix, la valeur médiane par un trait noirci présent dans la boîte délimitée par le premier et troisième quartile. La différence de ces deux valeurs (formant la boîte) représente la valeur interquartile (IQR) où sont contenues 50% des données recueillies.

Le minimum et maximum sont représentés par les moustaches de notre graphique et peuvent s'étendre jusqu'à $1.5 \times \text{IQR}$ où sont délimitées les valeurs extrêmes de la série. Toutes valeurs qui se trouvent au-delà de l'étendue attendue ($1.5 \times \text{IQR}$) sont considérées comme valeurs aberrantes et méritent d'être examinées afin de les considérer comme aberrantes. Elles sont représentées par des points dans le *Boxplot*.

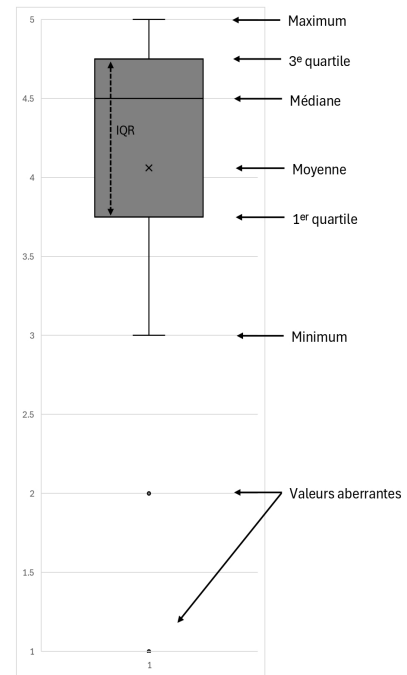


Figure 4.3 :
*Représentation d'un
Boxplot*

L'analyse de la question Q2 est particulière de par son caractère graphique. Nous avons laissé le choix aux personnes qui ont passé le test de dessiner des parts autour d'une tête vue de dessus. Nous avons volontairement laissé le choix quant aux nombres de parts à dessiner pour caractériser la direction perçue de chacun et chacune. Cependant, la densité du trait n'avait aucune importance dans la mise en forme des résultats.

Afin d'examiner cette question, nous avons opté pour une visualisation ludique sous la forme d'une *heatmap*¹ circulaire. Ce format organisé en anneau permet de représenter visuellement en mettant en évidence les zones les plus sollicitées par nos participantes et participants. L'anneau d'origine comprenait 18 parts. Cependant, pendant la récolte des résultats, j'ai pu observer qu'un nombre important de personnes avait choisi de colorier des demi-parts, j'ai donc opté pour un anneau de 36 parts pour la mise en forme des résultats de cette question. Nous avons ensuite reporté informatiquement chaque part dessinée par l'un ou l'une de nos participants et participantes et avons appliqué une échelle de couleur dont les valeurs sont présentées en Figure 4.4.

¹Carte thermique



Figure 4.4 : Nuancier de couleur utilisé pour la question Q2, une case non dessinée sera représentée en jaune alors qu’une case dessinée le plus de fois sera rouge

4.1.2 Analyse par questions

4.1.2.1 Q1 : Localisation

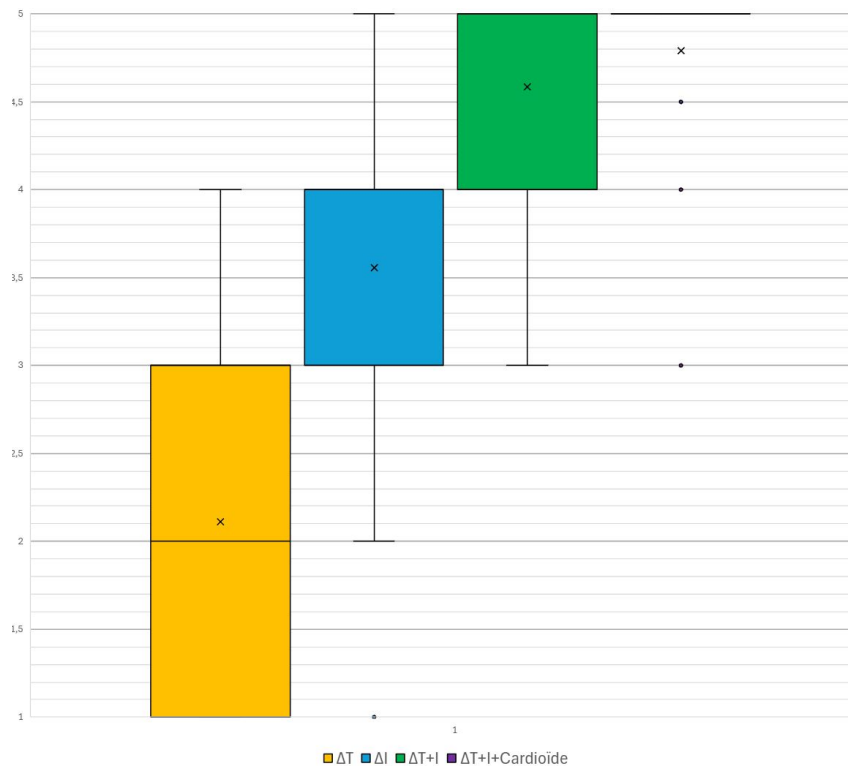


Figure 4.5 : Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser?” pour la source “Voix parlée”

Pour la voix parlée en Figure 4.5, on observe une plus grande facilité de localisation dès qu’on combine les différences de temps avec les différences d’intensité. Le ΔT pur présente une médiane de 2 et une moyenne 2.1 contrairement au ΔI pur qui présente une médiane de 4 et une moyenne de 3.6. Donc, nous arrivons mieux à localiser en ΔI pur plutôt qu’en ΔT pur. Cependant, les résultats sont très bons lorsque ΔI et ΔT se joignent avec une médiane de 5 et une moyenne de 4.6. Lorsque nous précisons les différences d’intensité, en simulant des capsules cardioïdes, les résultats sont encore meilleurs avec une médiane de 5 et une moyenne de 4.8. Nous pouvons observer des valeurs aberrantes pour ce cas (une réponse de 3; cinq réponses de 4; une réponse de 4.5).

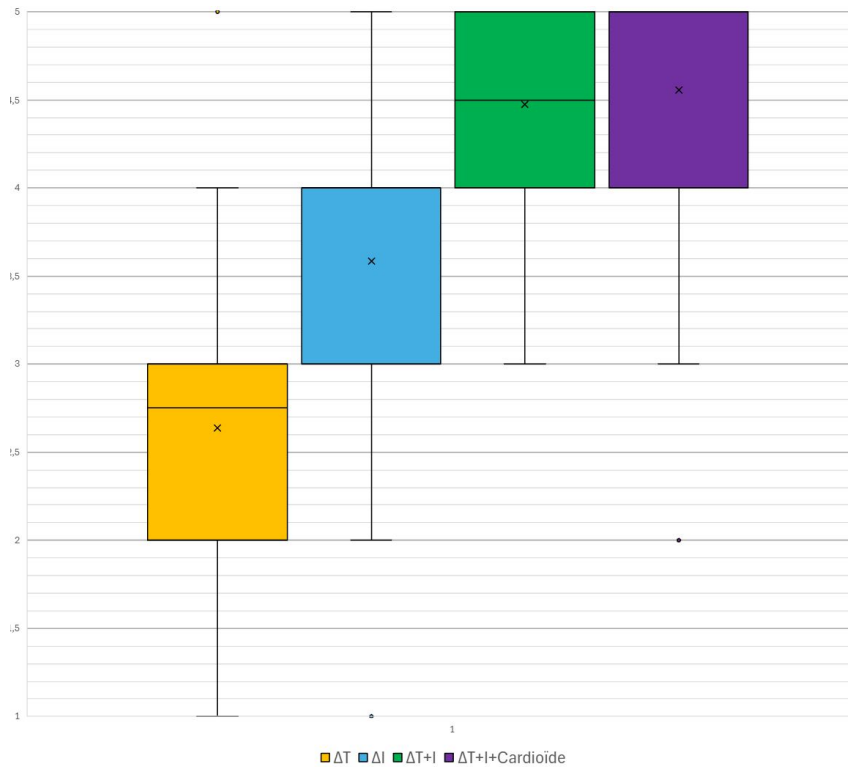


Figure 4.6 : *Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?” pour la source “Guitare”*

Pour la guitare, la répartition des réponses est assez similaire à celles de la voix même si nous pouvons observer quelques différences (Figure 4.6). En effet, 50% des réponses se trouvent entre 2 et 3 avec une valeur moyenne de 2.6 et valeur médiane de 2.75 en ΔT pur contrairement à la voix parlée. Nous avons également recueilli une valeur aberrante de 5. Pour les ΔI pur et $\Delta T+I$ les réponses sont assez similaires, comme pour le cas précédent. En ce qui concerne les cas $\Delta T+I$ et $\Delta T+I+“Cardioïde”$, nous pouvons observer que la répartition des réponses est équivalente pour les deux cas. Seul la valeur médiane est différente, respectivement 4.5 et 5, car les valeurs moyennes sont très proches, respectivement 4.47 et 4.56. Pour la guitare, la précision amenée en ΔI à l’aide de la simulation de capsules cardioïdes n’apporte pas de bénéfices.

Les réponses qui concernent la darbouka, en Figure 4.7, sont intéressantes et différentes de ce que nous avons pu analyser auparavant. Pour les cas ΔT pur et ΔI pur, les réponses sont relativement similaires, car 50% de nos valeurs se répartissent entre 3 et 4 avec comme seules différences les valeurs médianes et moyennes qui sont respectivement pour la médiane de 3 et 4 et pour la moyenne de 3.25 et 3.64. Cette similitude pourrait être expliquée par la présence de composantes aiguës et de transitoires grâce à la nature

percussive de cet instrument. Comme la guitare, les cas $\Delta T+I$ pur et $\Delta T+I$ +“Cardioïde” sont équivalents et ont pour seule différence la valeur moyenne qui est respectivement de 4.36 et 4.58 car les valeurs médianes sont égales pour une valeur de 5.

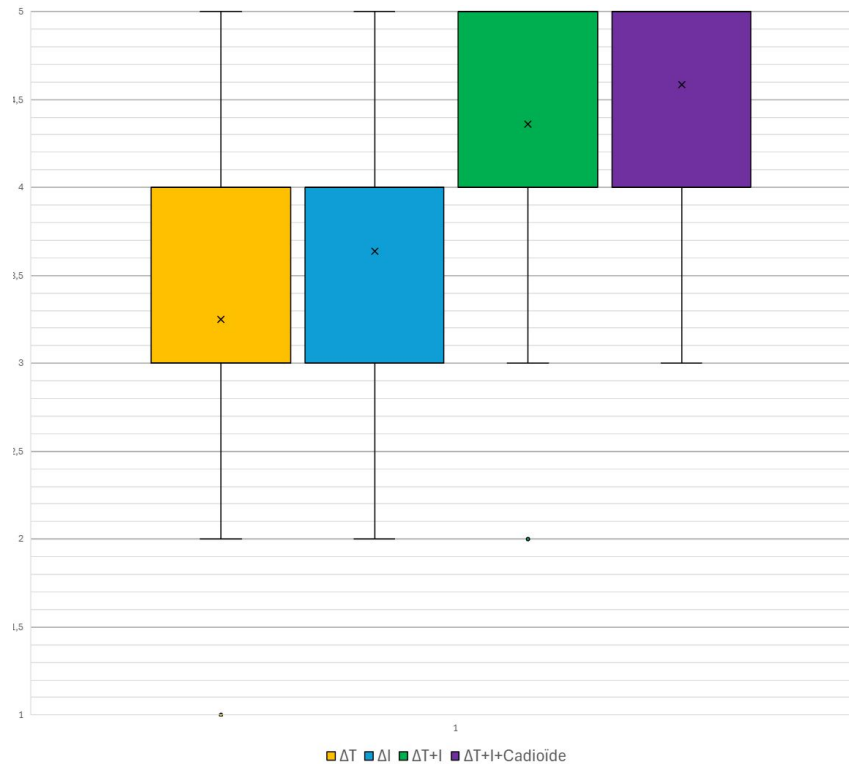


Figure 4.7 : *Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?” pour la source “Darbouka”*

En évaluant les analyses de cette question pour les trois sources testées, nous pouvons déduire avec certitude que la combinaison du ΔT et ΔI est plus que bénéfique dans notre capacité à localiser une source sonore. L’ajout de la simulation de capsules cardioïdes amène une précision qui est à prendre en compte, mais sans être très importante.

Nous pouvons également ajouter que cette combinaison reste performante quel que soit le positionnement des participantes et participants dans la salle. En effet, en regroupant les réponses des trois sources sonores tout en distinguant les différentes positions dans la salle, nous pouvons constater que les cas $\Delta T+I$ et $\Delta T+I$ +“Cardioïde” offrent une excellente précision de localisation de la source sonore.

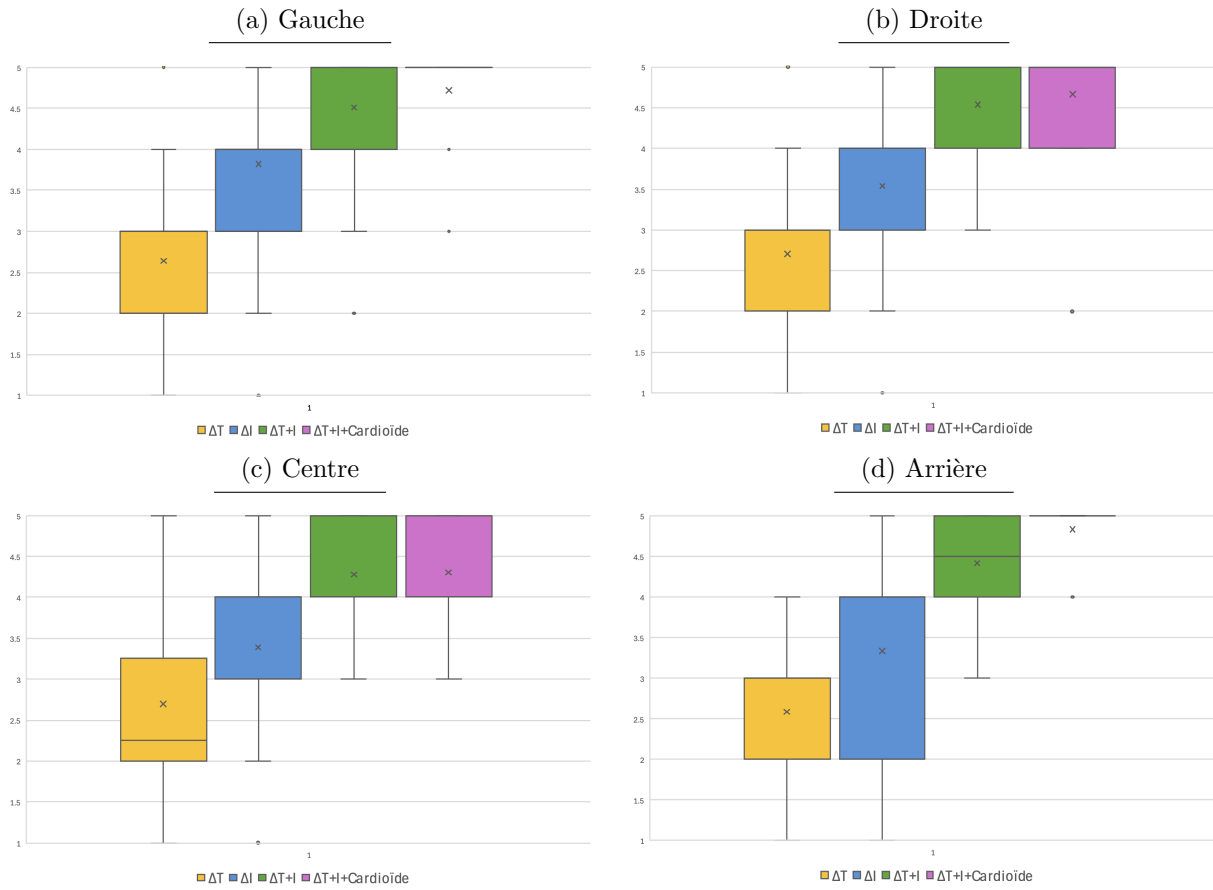


Figure 4.8 : *Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser?” pour différentes positions dans la salle, en regroupant les réponses des trois sources sonores*

4.1.2.2 Q2 : Direction

Nous pouvons observer nos différentes *heatmap* circulaires en fonction des différents mécanismes perceptifs en Figure 4.9 pour la voix parlée, Figure 4.10 pour la guitare et Figure 4.11 pour la darbouka.

Pour la voix parlée en Figure 4.9, on observe que les représentations des participants suggèrent une source sonore diffuse, pour le cas en ΔT pur, elle semble provenir de plusieurs directions à la fois. Cependant, on note une concentration marquée des réponses autour d'une zone relativement étendue. Cette source est moins diffuse et sa zone directive perçue se déplace légèrement en ΔI pur. Dès lors que ces deux mécanismes sont combinés, la zone diffuse rétrécit considérablement, ne laissant apparaître aucune réponse provenant des côtés. Cette zone se rétrécit légèrement à l'aide de la simulation de nos capsules en cardioïdes.

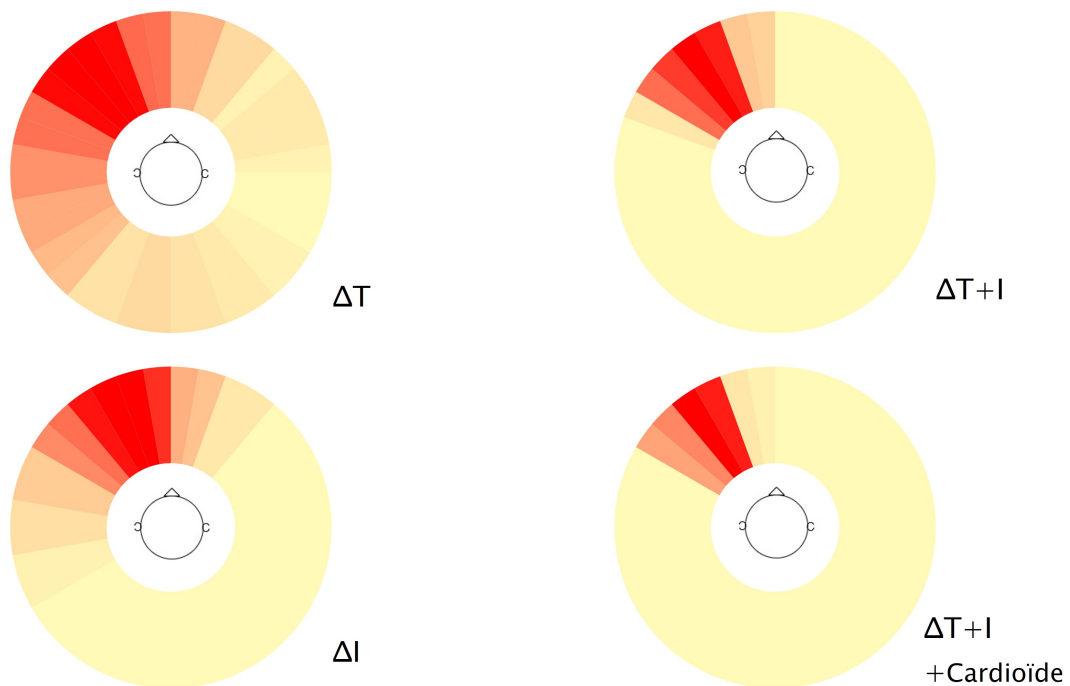


Figure 4.9 : *Heatmap circulaire de la question “Dans quelle direction avez-vous perçu la source ?” pour la source “Voix parlée”*

En ce qui concerne la guitare en Figure 4.10, l'analyse est similaire à celle de la voix parlée. On peut observer qu'en ΔT pur la source semble provenir de plusieurs directions à la fois tout en ayant une vague idée de la provenance de la source. En ΔI pur, la direction perçue se précise modérément, puis de façon plus marquée lorsque les deux mécanismes sont combinés, et encore davantage avec une simulation cardioïde.

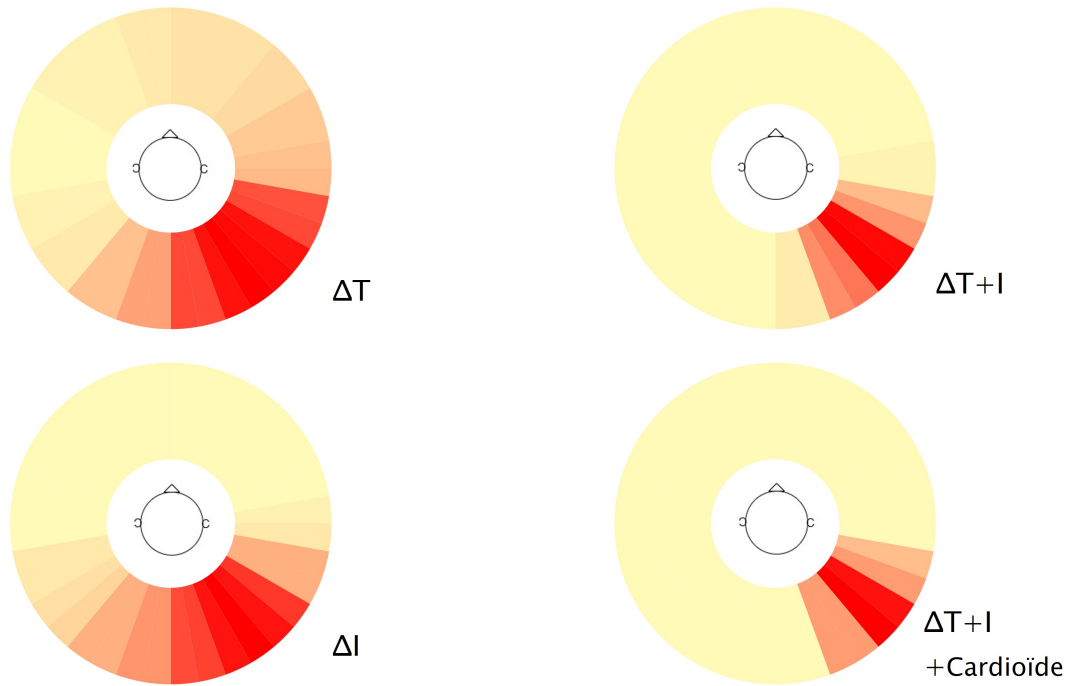


Figure 4.10 : Heatmap circulaire de la question “Dans quelle direction avez-vous perçu la source ?” pour la source “Guitare”

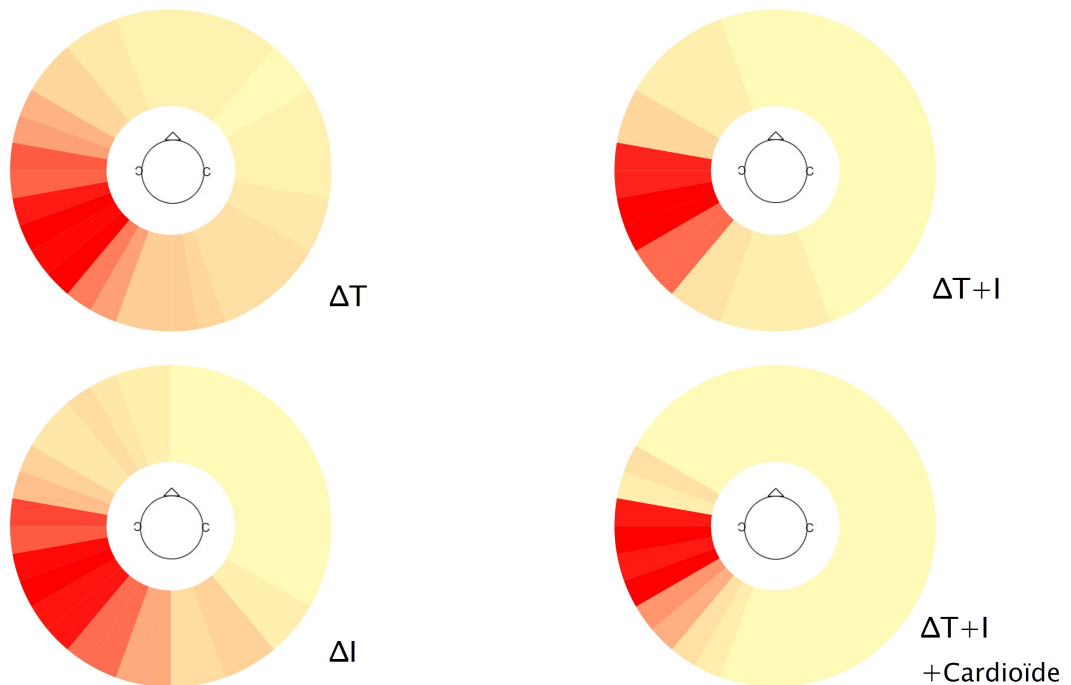


Figure 4.11 : Heatmap circulaire de la question “Dans quelle direction avez-vous perçu la source ?” pour la source “Darbouka”

De nouveau, le cas de la darbouka est particulier, car les résultats, en Figure 4.11, pour les cas ΔT pur et ΔI pur, sont assez similaires même si le ΔI pur précise légèrement la direction perçue de la source. Cette remarque est également valable pour les cas $\Delta T+I$ et $\Delta T+I$ +“Cardioïde” où la simulation en cardioïde précise légèrement. Cependant, la combinaison de ces deux mécanismes contribue malgré tout à une clarification de la direction perçue par les participants, bien que dans une moindre mesure comparée à nos deux sources précédentes.

Lorsque nous mettons en commun les analyses de nos trois sources sonores, on peut en déduire que la combinaison en $\Delta T+I$ permet de préciser la direction de notre source à travers un système physique. La simulation de nos capsules virtuelles en cardioïde améliore légèrement la précision, sans que cet effet soit réellement significatif.

4.1.2.3 Q3 : Étalement

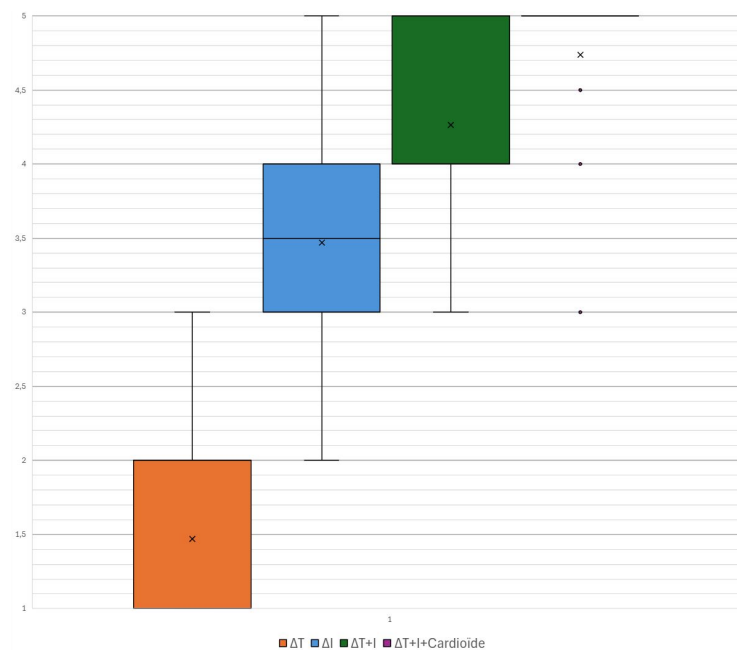


Figure 4.12 : *Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise?” pour la source “Voix parlée”*

Pour la voix parlée, on peut observer en Figure 4.12 qu’en ΔT pur, la source sonore est perçue comme très large avec une valeur médiane de 1 et moyenne de 1.47. Dès notre passage en ΔI pur (médiane de 3.5 et moyenne de 3.47), la source tend à se ponctualiser progressivement, phénomène qui s’accroît dans les deux derniers cas $\Delta T+I$ et $\Delta T+I$ +“Cardioïde” et qui présente des valeurs médianes respectives de 4 et 5 et des valeurs moyennes respectives de 4.26 et 4.74.

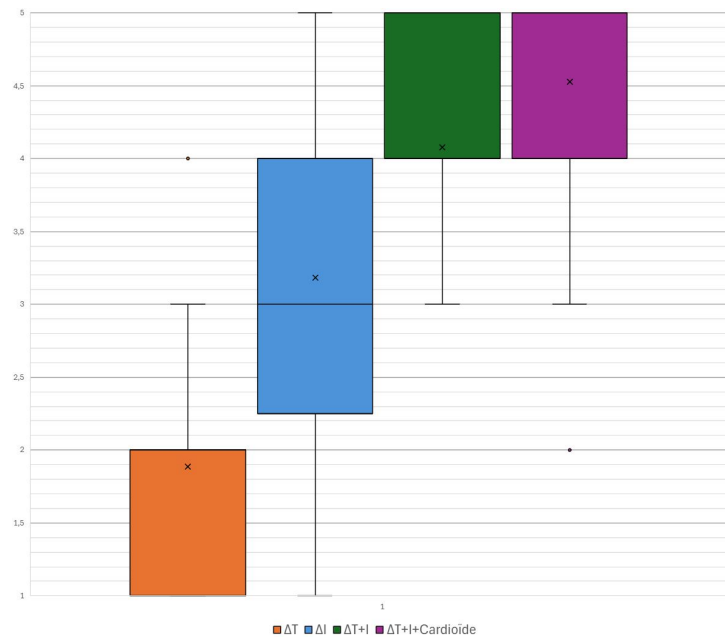


Figure 4.13 : *Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?” pour la source “Guitare”*

Les résultats en ΔT pur pour la guitare sont similaires à l’analyse de la source précédente. Cependant, nous pouvons noter que les résultats en ΔI pur sont dispersés, avec un IQR de 1,75, une étendue allant de 1 à 5, et une médiane située à 3. Ce résultat pourrait s’expliquer par le positionnement de notre source à l’arrière de notre dispositif sonore. Notre capacité à localiser une source sonore située à l’arrière est nettement moins précise que pour une source située à l’avant (le pouvoir séparateur azimuthal d’une source sonore pour un être humain est de $\pm 3.6^\circ$ à l’avant, $\pm 5.5^\circ$ à l’arrière et de $\pm 10^\circ$ sur les côtés. Les graphiques sont référencés en Annexe F : Pouvoir séparateur selon différents plans, BLAUERT, 1996). Les cas $\Delta T+I$ et $\Delta T+I+“Cardioïde”$ sont assez similaires même si le $\Delta T+I+“Cardioïde”$ présentent une source légèrement moins étalée.

Pour finir, nos sujets ont perçu en ΔT pur une darbouka assez large, médiane de 2 et moyenne de 2.54. Étonnement, les résultats en ΔI pur sont équivalents à ceux de la guitare avec une dispersion de nos résultats avec une médiane de 3, une étendue de 1 à 5, un IQR de 2 et une moyenne de 2.92. En ce qui concerne le cas $\Delta T+I$, c’est la seule source présentant un IQR aussi élevé pour cette source de 1.75. La médiane reste tout de même élevée avec une valeur de 4, mais les résultats s’étendent de 2 à 5. Ces résultats peuvent également s’expliquer par le pouvoir séparateur azimuthal qui est moins performant sur les côtes ($\pm 10^\circ$). En $\Delta T+I+“Cardioïde”$, la source se précise davantage avec une valeur médiane de 4, une moyenne de 4.3 et une répartition de 50% des réponses entre 4 et 5.

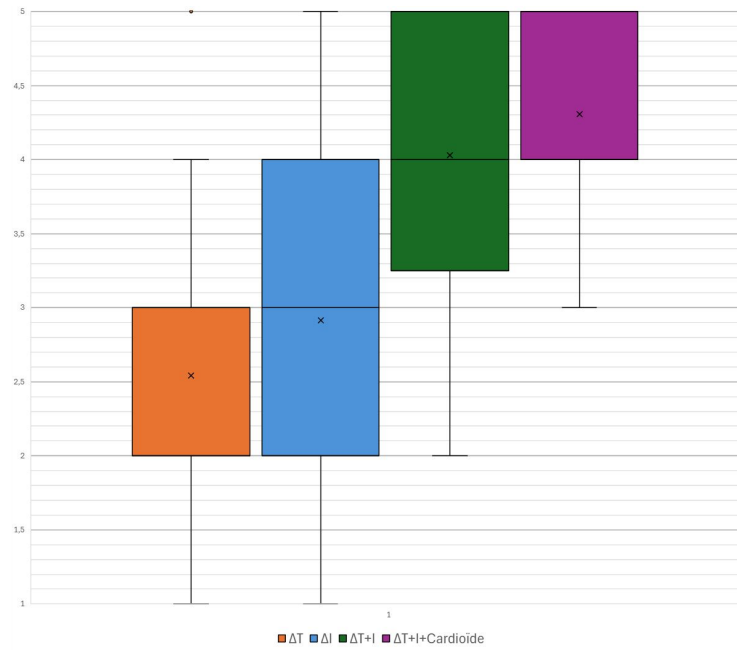


Figure 4.14 : *Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?” pour la source “Darbouka”*

On obtient ainsi une source davantage ponctuelle dès que ΔT et ΔI s’assemblent et davantage dès que nous simulons un système de captation cardioïde.

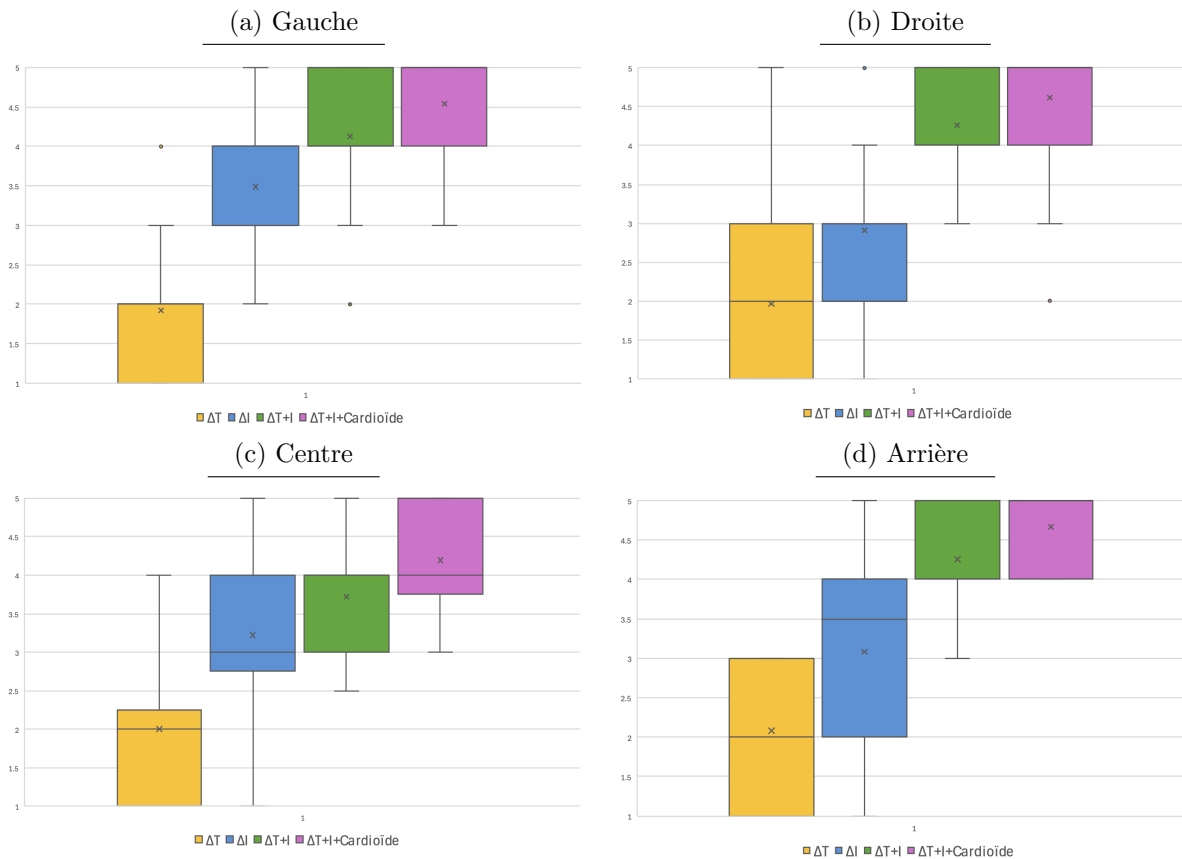


Figure 4.15 : *Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?” pour différentes positions dans la salle, en regroupant les réponses des trois sources sonores*

Pour conclure, les analyses de la Figure 4.15 montrent, en tenant compte du positionnement de nos participantes et participants et en regroupant les résultats de nos trois sources sonores, que la source apparaît plus ponctuelle lorsque ΔT et ΔI se combinent, et encore davantage avec une simulation de captation cardioïde.

4.1.2.4 Q4 : Préférence

Bien que cette question soit de nature plutôt empirique, elle a néanmoins permis de faire émerger quelques observations intéressantes. Lors d'échanges suite à l'expérience, nous avons pu discuter de cette question ainsi que des raisons ayant justifié les choix de réponse de nos sujets. Ces raisons étaient pour la plupart subjectives et dépendaient du goût et de l'expérience de chacun et chacune.

Par exemple, certains participants m'ont dit préférer la voix (Figure 4.16a) en ΔT pur pour des effets esthétiques qui pourraient être intéressants au cinéma afin de jouer une voix omnisciente ou intérieure grâce à la largeur de la source dans ces conditions. D'autres m'ont dit préférer le cas $\Delta T+I$ parce que la voix leur paraissait claire et plus intelligible.

Concernant la guitare et la darbouka (Figure 4.16b et Figure 4.16c), les remarques sont assez équivalentes. Certaines personnes ont particulièrement apprécié la restitution en ΔT pur, en raison de la sensation de largeur apportée par ce paramètre, et d'autres les cas $\Delta T+I$ et $\Delta T+I$ +“Cardioïde” pour la précision de la localisation.

Nous avons laissé la possibilité aux participantes et participants de sélectionner plusieurs réponses s'ils le souhaitaient, les résultats sont présentés ci-dessous (Figure 4.16). On remarque néanmoins que, pour le cas de la voix parlée, les cas $\Delta T+I$ et $\Delta T+I$ +“Cardioïde” ont été très fréquemment choisis.

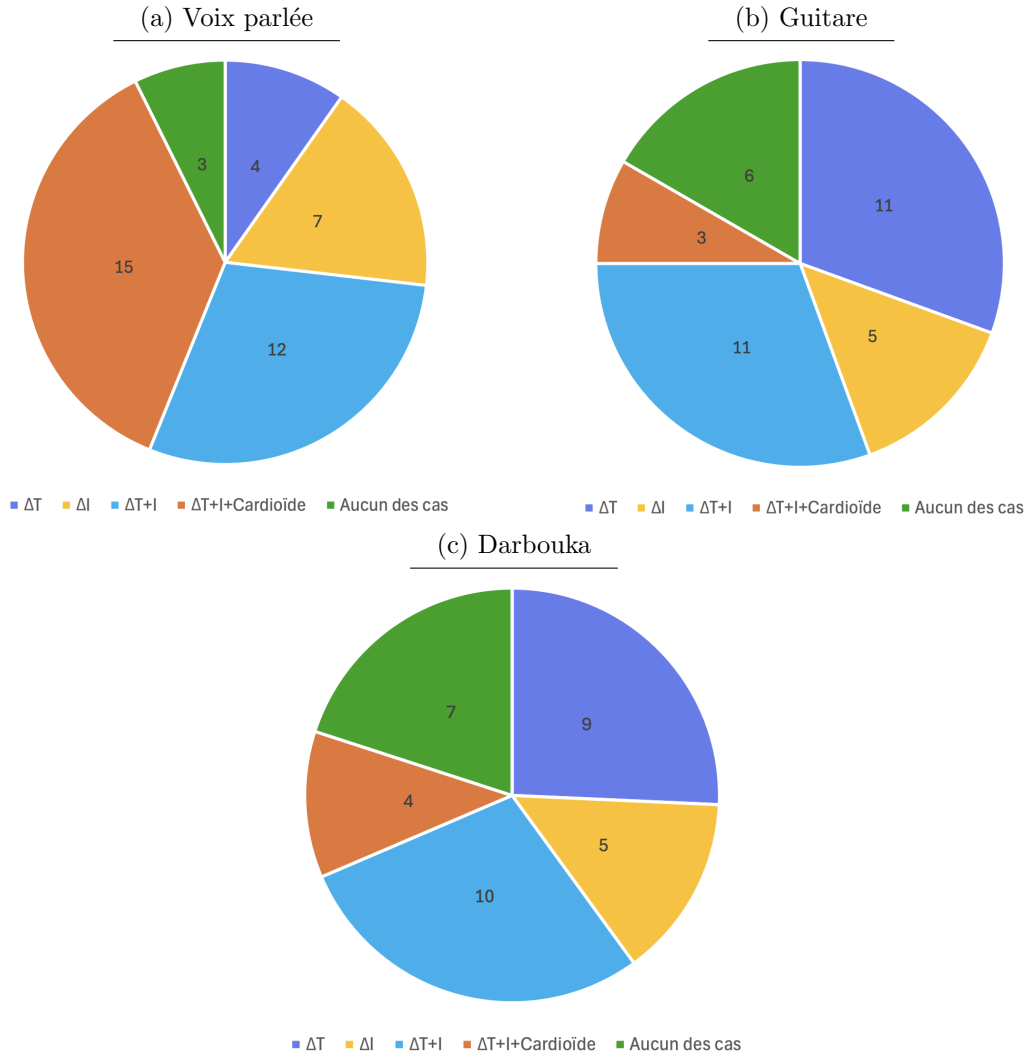


Figure 4.16 : Répartition des réponses des participantes et participants par sources sonores

4.1.3 Discussion autour des analyses

Cette analyse a permis de mettre en évidence comment chaque mécanisme seul ou combiné influe sur la perception de la position de la source et de son étendue à travers un système multicanal. Ces résultats nous permettent de confirmer l'hypothèse menée en 1.5.3 et appuyée par ROGINSKA et GELUSO (2018), que les différences de temps en complément de celles d'intensité forment une meilleure localisation qui nous paraît plus naturelle. Cette combinaison permet également un ancrage plus robuste de la source dans l'espace sonore en fonction de la position de nos participants et participantes.

Lorsque nous regroupons l'ensemble des analyses issues des différentes questions, nous pouvons observer clairement que la combinaison des mécanismes perceptifs offre une amélioration notable sur nos critères d'évaluation (localisation, direction, étalement). Bien que la précision apportée par la simulation d'un système de captation cardioïde à travers

les différences d'intensité soit pertinente, elle ne se démarque pas de façon significative par rapport à la performance obtenue avec la combinaison des deux mécanismes en configuration omnidirectionnelle.

Plusieurs remarques de nos participants sont remontées à travers ce test. La notion de profondeur est revenue plusieurs fois, selon les cas, nos participantes et participants n'ont pas senti la même sensation de profondeur de source. Lors des échanges après le test, nous avons pu déterminer que cette sensation était plus ou moins présente dès qu'un mécanisme présentait des différences temporelles. Ensuite, plusieurs remarques récurrentes ont été recueillies, notamment une sensation de séparation du spectre perçue sur la darbouka dans le cas de la spatialisation en ΔT pur. Certains participants ont également signalé une instabilité ou un déplacement de certaines composantes spectrales dans l'espace sonore.

Pour finir, nous avons pu identifier plusieurs aspects qui auraient pu influencer nos analyses dans notre protocole de test.

Premièrement, la répartition de nos participants et participantes en session était nécessaire, afin que l'ordre de passage des sources sonores et des cas perceptifs n'influent pas nos résultats, cependant cette contrainte nécessitait de trouver un nombre suffisant de personnes dans chaque session. Nous avons accueilli entre 11 et 14 personnes sur les trois différentes sessions, ce qui est à prendre en compte, car ce sont de petits effectifs qui peuvent introduire de grandes variations dans les résultats et rendent ainsi l'analyse compliquée.

Deuxièmement, l'expérience s'est déroulée dans une petite salle où nos participants et participantes se retrouvaient rapidement proches du dispositif sonore. Nous pensons que cette contrainte a pu induire des différences quant aux résultats de la question, sur la provenance et direction de la source qui est en lien direct avec l'ancrage de la source sonore.

Enfin, certains participants ont souligné l'importance d'une écoute continue de 30 secondes où les différents cas étaient enchaînés. Cette écoute leur permettait de se rendre compte aisément des différences entre chaque restitution, facilitant une comparaison perceptive.

4.2 Test pratique

4.2.1 Analyse par critères

4.2.1.1 Ergonomie

De manière générale, nos participants et participantes ont trouvé que l’interface graphique était complète et intuitive. Le nombre de paramètres a paru suffisant, cependant plusieurs remarques récurrentes concernant l’ajout de fonctionnalités sont revenues :

- la possibilité de spatialiser des sources stéréophoniques ou à trois canaux
- le fait de pouvoir créer son arbre “à la main” directement sur la représentation 2D et non à partir des boîtes *float* de *Max*
- permettre de simuler des directivités différentes sur notre arbre en même temps
- afficher plus en évidence le cas de spatialisation appliqué

Les participants et participantes m’ont également transmis des remarques sur l’affichage de certains paramètres. Les utilisateurs et utilisatrices auraient aimé voir affichées les directivités microphoniques afin de mieux visualiser l’ouverture de nos microphones virtuels. Une grille à l’échelle (en mètre) m’a été suggérée afin de mieux réaliser la taille réelle de notre arbre et l’affichage des délais temporels par capsule aurait également apparaître. Un participant m’a soumis cette idée : proposer des *presets* ou donner la possibilité à l’utilisateur de dessiner une forme dans l’espace 2D que suivrait notre source sonore comme trajectoire.

4.2.1.2 Perception

La plupart de nos participants et participantes ont trouvé que la spatialisation induite par le cas $\Delta T+I$, était très intéressante grâce à la précision angulaire et l’ancrage/stabilité de la source dans le système de diffusion. La restitution spatiale en ΔI pur a pour résultat une source sonore plus large qui s’étale davantage. La précision spatiale de l’outil “*apparaît de manière flagrante avec la localisation $\Delta T+I$. C’est un vrai apport par rapport à un pan² classique*” (CH). Avec CH, LLN et NV, nous avons comparé la voix parlée en ΔI pur, $\Delta T+I$ et la différence de précision était notoire. “*Les différences sont notables, notamment sur une perception plus définie, moins “floue”, plus stable*” (NV) pour une spatialisation en

²Correspond à une abréviation du mot panoramique

$\Delta T+I$. Nous avons pu réaliser le même type de test avec OG et AH en générant un signal impulsionnel sur une station de travail. Ce signal était constitué d'un échantillon non nul précédé et suivi de silence. Bien qu'il s'agisse, en temps discret, de l'impulsion unité, ce signal constitue l'équivalent discret de l'impulsion de Dirac³ et joue le rôle d'élément neutre de la convolution dans les systèmes linéaires discrets. Nos participants ont utilisé ce type de signal afin de vérifier si les retards temporels induits par notre outil étaient gênants (perception de différentes sources due aux délais temporels) ou permettaient une meilleure localisation. Le résultat obtenu était concluant et similaire au test sur la voie parlée. Nos participants ont également essayé plusieurs configurations d'arbres, tout en se déplaçant dans la salle pour évaluer la "zone d'action" de l'outil et ce qui en découle fréquemment est (en $\Delta T+I$) :

- une image sonore "flatteuse" obtenue à l'aide d'un arbre omnidirectionnel
- la précision de localisation pour un arbre cardioïde ou hypercardioïde, qui demeure notable dans plusieurs endroits de la salle

Le cas en ΔT pur a également plu afin de créer des effets sonores. Avec OG et AH, nous avons pu tester cette spatialisation sur une source musicale continue en jouant sur le facteur multiplicateur des lignes de délai, le résultat obtenu était intrigant et permettait de créer des effets attrayants. Avec FD, nous avons pu tester ce même type de spatialisation avec des sons ponctuels d'impact dans les basses fréquences. L'effet résultant était également captivant grâce à l'attaque qui se délayait à travers les canaux et la partie basse fréquence qui liait le tout.

Plusieurs personnes semblent intéressées par les possibilités de l'outil pour réaliser de l'upmixage (1.3.2.3). Le manque de prises de son d'ambiances ou d'effets captés directement en multicanal, souvent lié à la difficulté de mettre en place un système de captation conséquent, tel qu'un arbre à sept capsules ou un arbre Atmos, laisse les monteuses et monteurs avec une insuffisance de sons pour des formats supérieurs. Ils manquent de matière sonore multicanale et sont ainsi régulièrement amenés à recourir à l'upmixage pour répondre aux exigences des projets qui utilisent la technologie Atmos. De fait, notre outil a suscité de l'intérêt en raison de la nouveauté perceptive qu'il apporte, et pourrait être utilisé afin de recréer des environnements sonores riches à des

³L'impulsion de Dirac, notée $\delta(t)$, est considéré par un signal où "la quasi-totalité de son énergie, unitaire, est concentrée au voisinage de l'origine des temps. [...] Cet outil théorique, donc impossible à réaliser dans la pratique, correspondrait à la localisation complète de l'information à l'instant $t = 0$." (MILLOT, 2008) Cette impulsion se définit par $\int_{-\infty}^{+\infty} \delta(t)dt = 1$

formats supérieurs à partir d'enregistrements monophoniques ou stéréophoniques. Avec CH et OG, nous avons expérimenté l'upmixage d'un canal issu d'un titre de musique actuelle en utilisant une simulation d'arbre microphonique omnidirectionnel. Le résultat s'est révélé prometteur, ouvrant ainsi une nouvelle voie pour aborder l'upmixage.

4.2.1.3 Remarques générales et suggestions d'améliorations

Une remarque récurrente était de rendre possible l'intégration de plusieurs sources (monophoniques, stéréophoniques...) dans la même instance afin de recréer des ambiances sonores multicanales complexes. Cette idée est très intéressante, car elle offrirait une flexibilité considérable dans la création d'ambiances sonores multicanales : l'utilisatrice ou l'utilisateur pourrait créer son ambiance à partir d'une multitude de sons simulée par un système de captation qui apporterait une certaine crédibilité à l'image sonore.

Une remarque pertinente d'OG portait sur l'absence d'acoustique dans notre outil. Actuellement, nos sources sont spatialisées à partir de la simulation d'un système de captation fictif placé dans "une chambre anéchoïque". Si aucune acoustique n'est produite, la notion de distance entre la source et le système est bel et bien présente. Nous pourrions ainsi introduire un moteur de réverbération ainsi qu'un filtre de premières réflexions, comme pour le ViMiC (1.4.2), et ainsi proposer plusieurs acoustiques différentes : intérieur salon, église...

La proposition d'un panel de *presets* est une remarque qui est revenue fréquemment. Ces configurations prédéfinies pourraient offrir à l'utilisatrice ou à l'utilisateur une solution rapide pour la spatialisation d'une source. On pourrait alors penser à plusieurs configurations pour des spatialisations voulues : arbre multicanal à sept capsules à prédominance de différences temporelles sur les enceintes frontales, arbre à effets (avec un facteur multiplicateur des différences de temps élevé) et plein d'autres encore.

4.2.2 Discussions autour des analyses

Nous pouvons constater que les résultats de ce dernier test sont en accord avec ceux du test perceptif (4.1.2). La spatialisation en $\Delta T+I$ augmente, pour les deux tests, la précision de la localisation d'une source sonore. Lorsque nous simulons des capsules cardioïdes, la précision s'affine légèrement. Le cas en ΔT a pu susciter l'intérêt de nos participants dans les deux expériences, notamment pour des raisons esthétiques.

De manière générale, nous avons pu observer lors de ce dernier test un attrait pour les arbres omnidirectionnels ou cardioïdes. Ces choix ne sont pas surprenants, compte tenu de l'expérience de chacune et chacun dans le domaine de la prise de son. Ces deux types d'arbres sont en effet couramment utilisés, bien que pour des raisons différentes. L'arbre multicanal omnidirectionnel était davantage utilisé, dans le cadre du dernier test, afin de recréer des ambiances ou d'upmixer, alors que l'arbre multicanal cardioïde était utilisé afin d'évaluer la précision de localisation de l'outil. Dans le domaine de la prise de son cinématographique, les arbres cardioïdes sont privilégiés afin de décorréler au maximum les différents canaux de diffusion, ce qui est le cas de l'arbre Atmos de HAL (HAL, s. d.).

Nous ne nous attendions pas à ce que notre outil suscite d'autres intérêts pratiques que celui d'un simple spatialisateur. Les participants et participantes ont trouvé l'outil prometteur et riche en ressources et possibilités, ce qui en fait un outil complexe mais intrigant. Il y a autant de possibilités que de configurations possibles avec cet outil, ce qui rend la mise en place de *presets* essentielle dans sa conception.

Pour conclure cette discussion, l'objectif de ce test était d'évaluer la pertinence d'un tel outil. Après nos rencontres avec les différents professionnels, nous pouvons dire avec conviction qu'il est pertinent et qu'il pourrait apporter de nouvelles manières de penser la spatialisation.

Conclusion

Cette recherche avait pour but de répondre à un questionnement sur l'instabilité spatiale d'une scène sonore au cinéma en envisageant la conception d'un outil susceptible d'y remédier. L'outil s'est avéré pertinent pour compenser cette instabilité, car il s'appuie sur une architecture fondée sur la combinaison des différences de temps et d'intensité ; cette combinaison permet effectivement de contrecarrer les défauts de spatialisation des systèmes actuels comme l'ont démontré nos tests. L'ajout de différences temporelles permet une meilleure précision de localisation, un ancrage de la source sonore dans la position choisie, tandis que les différences d'intensité seules offrent une bonne localisation qui reste cependant moins précise et qui peut varier selon la position de la spectatrice ou du spectateur. L'outil proposé pourrait ainsi constituer un atout majeur pour la spatialisation sonore au cinéma et permettre la fabrication de scènes sonores bien plus stables. Cet outil pourrait permettre aux utilisatrices et utilisateurs de jouer sur l'importance des différents mécanismes perceptifs puisqu'il contribuerait à renforcer la stabilité des scènes sonores et à favoriser une meilleure cohérence de l'espace sonore perçu par le public, indépendamment de la variabilité des configurations des systèmes de diffusion. La perception de l'espace sonore perçu par le public est donc affectée et cette recherche démontre qu'il est possible de changer la façon dont nous percevons certaines sources sonores en salle.

Ces tests nous ont également montré qu'aucune préférence particulière ne se dégageait pour une restitution spatiale de l'outil. Au contraire, c'est la diversité des restitutions proposées qui a plu à nos auditrices et auditeurs, et celles-ci semblent permettre de répondre à des idées spécifiques. Nous ne pensions pas que l'outil allait être à la fois un spatialisateur, un upmixeur et un moteur d'effets. Selon les mécanismes choisis, les participants y voyaient des utilisations esthétiques variables. Par exemple, les différences temporelles seules permettaient de créer des effets assez poignants pour divers objectifs : omniprésence d'une voix "intérieure", effet de largeur et upmixage sur des sources musicales. Alors que la combinaison des deux différences était utilisée pour

préciser et ancrer la localisation d'une source dans l'espace sonore.

Au fil de notre démarche, de nouvelles pistes de réflexion sont apparues, mais n'ont pu être explorées davantage par soucis techniques ou de temps. Tout d'abord, nous aurions voulu réaliser un dernier test afin de mixer un court projet cinématographique à l'aide de notre outil. Ce test aurait permis de prendre en compte l'intégration de l'image, élément qui joue un rôle majeur dans la façon dont nous percevons la scène sonore en salle de cinéma. Cependant, notre outil développé sous *Max* ne permet pas une intégration simple et fluide de notre outil dans l'univers Avid d'une session de mixage sous Protools. Nous aurions néanmoins pu réaliser un test à l'aide d'une autre station de travail audionumérique telle que Reaper ou Nuendo, mais cela aurait nécessité une adaptation de l'implémentation de l'outil.

Lorsqu'un mixage pour un projet cinématographique est terminé, l'étape suivante consiste fréquemment à effectuer le downmixage. Son principe se rapproche énormément de celle de l'upmixage mais en sens inverse. Le downmixage consiste à adapter notre fichier audio multicanal à un format inférieur, par exemple un fichier audio à sept canaux vers deux canaux. De fait, nous sommons différents canaux en un seul ou plusieurs, et nous créons ainsi de nouveaux canaux audio. Cependant, nous savons que la sommation de canaux audio sur lesquels des retards temporels sont appliqués peut induire des modifications spectrales. Il aurait été intéressant d'évaluer cette problématique à l'aide d'une session de mixage utilisant l'outil.

Lors de notre rencontre avec Michael Williams, une dernière problématique est apparue. Il est vrai que la simulation d'un système de captation virtuel permet une excellente retranscription d'une image sonore, mais uniquement depuis une position d'écoute précise. La conception d'un tel système est conçue pour une écoute unique qui se trouve généralement au centre du système de reproduction. C'est pourquoi, notre outil nous a paru contradictoire dans sa conception. Cependant, le dernier test a mis en évidence une bonne stabilité d'une source sonore à une position choisie, même lorsque nous nous déplaçons dans la salle avec nos participants et participantes.

Cet outil ne se destine donc pas seulement au monde cinématographique, mais pourrait trouver des applications dans d'autres domaines tels que la création sonore pour le théâtre, les expositions muséales ou les installations sonores. Nous pourrions imaginer un dispositif où plusieurs capsules virtuelles seraient distribuées au nombre de canaux correspondants. Les conceptrices et concepteurs sonores auraient alors à leur disposition "une boîte à outils" qui leur permettrait de spatialiser de multiples manières et de créer des effets sonores élaborés.

Nous sommes conscients que l'outil issu de cette recherche n'est pas un outil de spatialisation classique, comme ceux que l'on pourrait trouver dans les stations de travail audionumériques actuelles. En effet, c'est un outil bien plus complexe aux multiples utilisations qui pourrait s'adapter à des contextes de création divers. Dans une prochaine recherche, il serait donc pertinent de se demander quelle serait la perception engendrée par un panoramique classique si l'on y appliquait des différences temporelles.

Bibliographie

- ALLEN, I. (2006). The X-Curve : Its Origins and History : Electro-Acoustic Characteristics in the Cinema and the Mix-Room, the Large Room and the Small. *SMPTE Motion Imaging Journal* (cf. p. 40).
- BERKHOUT, A. J., DE VRIES, D., & VOGEL, P. (1993). Acoustic control by wave field synthesis. *The Journal of the Acoustical Society of America*, 93(5), 2764-2778 (cf. p. 35).
- BLAUERT, J. (1996). *Spatial Hearing : The Psychophysics of Human Sound Localization*. MIT Press. (Cf. p. 21, 42, 62, 65, 83, 102, 128).
- BOTTE, M.-C., CANÉVET, G., DEMANY, L., & SORIN, C. (1988). *Psychoacoustique et perception auditive*. INSERM. (Cf. p. 18, 19, 64).
- BRAASCH, J., PETERS, N., & VALENTE, D. L. (2007). Sharing Acoustic Spaces Over Telepresence Using Virtual Microphone Control. *123rd Convention of the Audio Engineering Society, Convention Paper 7209* (cf. p. 36).
- BRAASCH, J., RYAN, T. J., & WOSZCZYK, W. (2005). An Immersive Audio Environment with Source Positioning Based on Virtual Microphone Control. *119th Convention of the Audio Engineering Society, Convention Paper 6546* (cf. p. 35, 37).
- CARREAU, R., & MACQUART, T. (2015). *Utilisation de la technologie WFS dans la création sonore cinématographique : possibilités et limites*. [Mémoire de master, ENS Louis-Lumière]. (Cf. p. 33).
- CORTEEL, E., & CAULKINS, T. (2004). *Sound Scene Creation and Manipulation using Wave Field Synthesis* (rapp. tech.). IRCAM. (Cf. p. 34).
- DOLBY LABORATORIES, I. (2012). Dolby® Atmos™ Un son de prochaine génération pour le cinéma (cf. p. 16).
- HAL. (s. d.). Urban Atmos - 7.0.2 Immersive City Ambiences. <https://www.hal-audio.com/product/urban-atmos/> (cf. p. 91).
- HOLMAN, T. (2008). *Surround Sound : Up and Running*. Focal Press. (Cf. p. 8, 38, 39).

- IRCAM. (2021, mars). Thibaut Carpentier (Ircam) : Workshop on Spatialization. [Vidéo]. <https://www.youtube.com/watch?v=EYMNBNz5JHQ> (cf. p. 30).
- KERINS, M. (2011). *Beyond Dolby (Stereo) : Cinema in the Digital Sound Age*. Indiana University Press. (Cf. p. 10-14).
- LETOWSKI, T. R., & LETOWSKI, S. T. (2012, mai). *Auditory Spatial Perception : Auditory Localization* (rapp. tech.). Defense Technical Information Center. Fort Belvoir, VA. (Cf. p. 17-19).
- MAGNIER, V. (2019). *Guide de la prise de son pour l'image* (4^e éd.). Dunod. (Cf. p. 23, 25).
- MAPP, P. (2007). Psychoacoustics in Sound Reinforcement and PA Design. *22nd Audio Engineering Society UK Conference, Illusions in Sound* (cf. p. 42, 43).
- MILLOT, L. (2008). *Traitement du signal audiovisuel - Applications avec Pure Data et Matlab*. Dunod. (Cf. p. 89).
- OZCAN, K., BUSBRIDGE, S. C., FRYER, P. A., GEAVES, G. P., & MOORE, J. P. (2003). The Significance of Phase as an Auditory Cue. *114th Convention of the Audio Engineering Society, Convention Paper 5740* (cf. p. 20).
- PECQUAIS, J.-L. (s. d.). Guide pratique des techniques du son. [https : / / doc . jeanlouppecquais.com/spatialisation/spat-historique.html](https://doc.jeanlouppecquais.com/spatialisation/spat-historique.html) (cf. p. 12).
- PECQUAIS, J.-L. (2019). *Proposition d'une alternative au potentiomètre de panoramique*. [Mémoire de master, ENS Louis-Lumière]. (Cf. p. 32).
- PÉRIAUX, B., OHL, J.-L., & THÉVENOT, P. (2015). *Le son multicanal : De la production à la diffusion du son 5.1, 3D et binaural*. Dunod. (Cf. p. 11, 14-16, 21, 24, 25, 32-34, 41, 102, 103, 110-113).
- PETERS, N., MATTHEWS, T., BRAASCH, J., & MCADAMS, S. (2008). Spatial sound rendering in Max/msp with ViMiC [CIRMMT - Centre for Interdisciplinary Research in Music Media and Technology]. *Proceedings of the International Computer Music Conference (ICMC)* (cf. p. 22, 35, 36).
- PULKKI, V., KARJALAINEN, M., & VÄLIMÄKI, V. (1999). Localization, coloration, and enhancement of amplitude-panned virtual sources [Affiliation des auteurs : "Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Espoo, Finland"]. *16th Audio Engineering Society International Conference* (cf. p. 17, 31, 32).
- RECOMMENDATION ITU-R BS.775-1 - *Multichannel Stereophonic Sound System with and without Accompanying Picture* (Recommandation N° ITU-R BS.775-1). (1994) (Adopted : 1992, revised : 1994. Defines standard 5.1 multichannel configuration).

- International Telecommunication Union – Radiocommunication Sector (ITU-R). (Cf. p. 38).
- ROGINSKA, A., & GELUSO, P. (2018). *Immersive Sound - The art and science of binaural and multi-channel audio*. Routledge. (Cf. p. 42, 86).
- SPAT Revolution - Panning Algorithms. (2024, mars). https://doc.flux.audio/spat-revolution/Spatialisation_Technology_Panning_Algorithms.html#dual-band-vector-based-panning-vbp-dual-band (cf. p. 32, 102, 115).
- THEILE, G. (2001). Multichannel Natural Music Recording Based on Psychoacoustic Principles. *Technical report, IRT* (cf. p. 108-111, 113, 114).
- WALKER, R. (1998). A Controlled-Reflection Listening Room for Multichannel Sound. *Reproduced Sound 1998* (cf. p. 39).
- WALLACH, H., NEWMAN, E. B., & ROSENZWEIG, M. R. (1973). The Precedence Effect in Sound Localization. *The Journal of the Acoustical Society of America*, 21(10) (cf. p. 42).
- WILLIAMS, M. (2004a). *Microphone Arrays for Stereo and Multichannel Sound Recording* (T. 1). Editrice II Rostro. (Cf. p. 22, 25, 41).
- WILLIAMS, M. (2004b). Multichannel Sound Recording using 3,4 and 5 channel arrays for Front Sound Stage Coverage. *117th Convention of the Audio Engineering Society, Convention Paper 6230* (cf. p. 49, 51, 54, 55).
- WILLIAMS, M. (2008). Migration of 5.0 Multichannel Microphone Array Design to Higher Order MMAD (6.0, 7.0 & 8.0) with or without the Inter-format Compatibility Criteria. *124th Convention of the Audio Engineering Society, Convention Paper 7480* (cf. p. 28-30).
- WILLIAMS, M. (2013). *Microphone Arrays for Stereo and Multichannel Sound Recording* (T. 2). Editrice II Rostro. (Cf. p. 28).
- WILLIAMS, M., & LE DÛ, G. (1999). Microphone Array Analysis for Multichannel Sound Recording. *107th Convention of the Audio Engineering Society* (cf. p. 27).
- WILLIAMS, M., & LE DÛ, G. (2000). Multichannel Microphone Array Design. *108th Convention of the Audio Engineering Society* (cf. p. 26).
- WILLIAMS, M., & LE DÛ, G. (2004). The Quick Reference Guide to Multichannel Microphone Arrays Design - Part II : Using Supercardioid and Hypocardioid Microphones. *116th Convention of the Audio Engineering Society, Convention Paper 6059* (cf. p. 27).
- XIE, B. (2022, septembre). *Spatial Sound : Principles and Applications* (1^{re} éd.). CRC Press. (Cf. p. 37, 43).

Table des figures

1.1	<i>Photos du système Fantasound installé en salle, www.pmamagazine.org/fr/comment-disneys-fantasound-a-introduit-le-son-surround-a-hollywood-en-1940/</i>	11
1.2	<i>Une photo agrandie d'une pellicule 35 mm, quatre formats sont représentés respectivement de gauche à droite : le SDDS de Sony, le Dolby Digital, le Dolby Stéréo et les informations de <i>timecode</i> pour DTS. (PECQUAIS, s. d.)</i>	12
1.3	<i>Format 7.1 SDDS cinéma (PÉRIAUX et al., 2015)</i>	14
1.4	<i>Format 7.1 ITU⁴ cinéma (PÉRIAUX et al., 2015)</i>	15
1.5	<i>Logiciel “Dolby renderer”, https://professional.dolby.com/product/dolby-atmos-content-creation/dolby-atmos-renderer/</i>	15
1.6	<i>Représentation d'une salle équipée en Dolby Atmos (DOLBY LABORATORIES, 2012)</i>	16
1.7	<i>Schéma d'une tête vue de dessus dans le plan azimutal avec la représentation d'un cône de confusion (figure de l'auteur)</i>	17
1.8	<i>Évaluation de la différence de marche Δd aux oreilles pour une source à l'infini d'azimut θ, Woodworth, 1938 (BOTTE et al., 1988)</i>	18
1.9	<i>Limites de la précision d'identification en azimut (Stevens et Newman, 1936), d'après BOTTE et al. (1988)</i> <i>a - Erreurs de la localisation en fonction de la fréquence,</i> <i>b - Pourcentage des confusions avant-arrière en fonction de la fréquence. .</i>	19
1.10	<i>Effet de l'ITD et l'IID en fonction de la fréquence (LETOWSKI et LETOWSKI, 2012)</i>	19
1.11	<i>Représentation de la phase</i>	20
1.12	<i>Diagrammes polaires des différents modèles théoriques de directivité microphonique obtenues à l'aide de la librairie matplotlib de Python</i>	23
1.13	<i>Configuration du triangle Decca (PÉRIAUX et al., 2015)</i>	24
1.14	<i>Photo de la croix IRT et d'un double ORTF (MAGNIER, 2019)</i>	25
1.15	<i>Représentation d'un arbre multicanal à cinq microphones hypocardiöide (WILLIAMS et LE DÛ, 2004)</i>	27

1.16	<i>Représentation d'un arbre multicanal à sept microphones cardioïde (WILLIAMS, 2013)</i>	28
1.17	<i>Représentation de deux arbres multicanaux à cinq et quatre microphones hypocardiïdes (WILLIAMS, 2008)</i>	29
1.18	<i>Représentation d'un arbre hypocardiïde obtenue à partir de la méthode "back to back array design" (WILLIAMS, 2008)</i>	30
1.19	<i>Configuration d'un système à panoramique d'amplitude bidimensionnel et tridimensionnel (PULKKI et al., 1999)</i>	32
1.20	<i>Principe d'Huygens (PÉRIAUX et al., 2015)</i>	34
1.21	<i>Principe d'Huygens à gauche et exemple d'un dispositif WFS à droite (CORTEEL et CAULKINS, 2004)</i>	34
1.22	<i>Interface graphique du ViMiC (BRAASCH et al., 2005)</i>	35
1.23	<i>Architecture de l'environnement virtuel ViMiC, d'après BRAASCH et al. (2007)</i>	36
1.24	<i>Schéma de la disposition des enceintes pour le format de spatialisation 5.1 (gauche) et positionnement des enceintes "surround" en cas de multiples canaux arrière (droite) (RECOMMENDATION ITU-R BS.775-1 - Multi-channel Stereophonic Sound System with and without Accompanying Picture, 1994)</i>	38
1.25	<i>Courbe "X" convention ISO 2969 (ALLEN, 2006)</i>	40
1.26	<i>Courbes illustrant la gêne due à l'écho d'un premier son par rapport à un second son retardé et amplifié (MAPP, 2007)</i>	43
1.27	<i>Courbe classique de l'effet Haas (MAPP, 2007)</i>	43
2.1	<i>Représentation de $\tan \theta$ à l'aide de Python</i>	46
2.2	<i>Représentation d'un arbre multicanal à cinq microphones omnidirectionnel (WILLIAMS, 2004b)</i>	49
2.3	<i>Représentation de deux arbres multicanaux à cinq et quatre microphones omnidirectionnels (WILLIAMS, 2004b)</i>	49
2.4	<i>Arbre multicanal à sept microphones omnidirectionnels, créé à partir des plans complémentaires en Figure 2.2 et Figure 2.3</i>	50
2.5	<i>Représentation d'un arbre multicanal à cinq microphones cardioïdes (WILLIAMS, 2004b)</i>	51
2.6	<i>Différences intercanales de temps (ICTD) et d'intensité (ICLD) entre les différents segments d'un arbre microphonique</i>	51

2.7	<i>Arbre multicanal à sept microphones cardioïdes, créé à partir des plans complémentaires en Figure 2.5 et Figure 2.6</i>	52
2.8	<i>Différences intercanales de temps en rouge (ICTD) et d'intensité en bleu (ICLD) entre les différents segments d'un arbre microphonique</i>	53
2.9	<i>Représentation d'un arbre multicanal à cinq microphones cardioïdes (WILLIAMS, 2004b)</i>	54
2.10	<i>Arbre multicanal à sept microphones cardioïdes, créé à partir des plans complémentaires tiré de WILLIAMS, 2004b</i>	55
2.11	<i>Différences intercanales de temps en rouge (ICTD) et d'intensité en bleu (ICLD) entre les différents segments d'un arbre microphonique</i>	56
2.12	<i>Proposition d'une interface graphique</i>	58
2.13	<i>Onglet "Presets (configurations)" de l'interface graphique</i>	59
3.1	<i>Représentation temporelle du déroulé du test perceptif pour une source sonore</i>	65
3.2	<i>Salle de réalisation du test</i>	66
3.3	<i>Valeurs de panoramisation appliquées à nos différentes sources dans l'ordre suivant : Voix, Guitare et Darbouka à travers l'interface graphique de l'outil</i>	68
3.4	<i>Interface réalisée à l'aide du mode Presentation de Max en vue de réaliser un test</i>	71
4.1	<i>Représentation des différentes positions occupées par les participants dans la salle de test</i>	73
4.2	<i>Répartition et position des participantes et participants par session</i>	74
4.3	<i>Représentation d'un Boxplot</i>	75
4.4	<i>Nuancier de couleur utilisé pour la question Q2, une case non dessinée sera représentée en jaune alors qu'une case dessinée le plus de fois sera rouge .</i>	76
4.5	<i>Boxplot de la question "Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser?" pour la source "Voix parlée"</i>	76
4.6	<i>Boxplot de la question "Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser?" pour la source "Guitare"</i>	77
4.7	<i>Boxplot de la question "Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser?" pour la source "Darbouka"</i>	78
4.8	<i>Boxplot de la question "Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser?" pour différentes positions dans la salle, en regroupant les réponses des trois sources sonores</i>	79
4.9	<i>Heatmap circulaire de la question "Dans quelle direction avez-vous perçu la source?" pour la source "Voix parlée"</i>	80

4.10	<i>Heatmap circulaire de la question “Dans quelle direction avez-vous perçu la source ?” pour la source “Guitare”</i>	81
4.11	<i>Heatmap circulaire de la question “Dans quelle direction avez-vous perçu la source ?” pour la source “Darbouka”</i>	81
4.12	<i>Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?” pour la source “Voix parlée”</i>	82
4.13	<i>Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?” pour la source “Guitare”</i>	83
4.14	<i>Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?” pour la source “Darbouka”</i>	84
4.15	<i>Boxplot de la question “Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?” pour différentes positions dans la salle, en regroupant les réponses des trois sources sonores</i>	84
4.16	<i>Répartition des réponses des participantes et participants par sources sonores</i>	86

Liste des tableaux

1.1	<i>Caractéristiques acoustiques recommandées pour une salle munie d’un système multicanal, tiré de HOLMAN (2008)</i>	39
1.2	<i>Tableau de correction en dB de la courbe “X” en fonction du nombre de sièges</i>	40
3.1	<i>Tableau comparatif de deux termes qui caractérisent la perception sonore selon BLAUERT (1996)</i>	62
3.2	<i>Tableau de répartition des sources sonores et des mécanismes perceptifs par session</i>	67

Table des Annexes

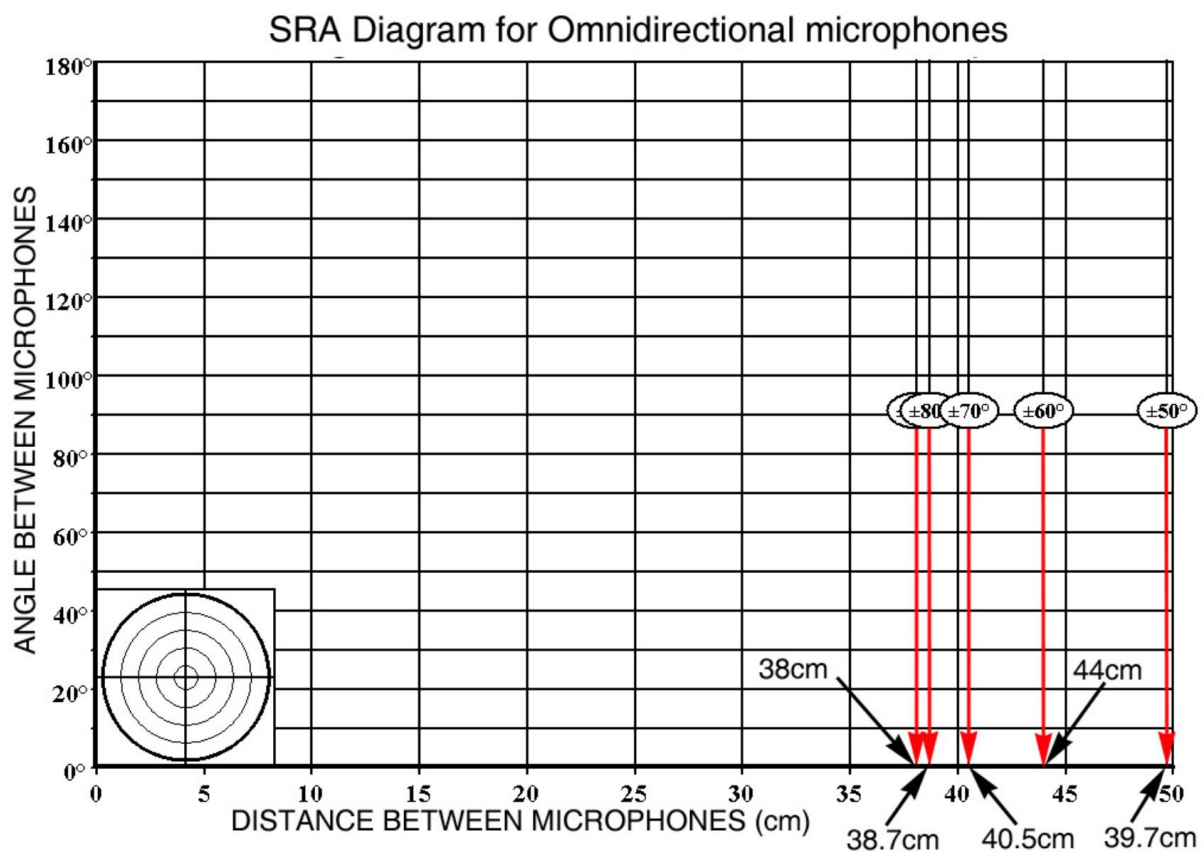
Annexe A : Frise des systèmes sonores au cinéma, PÉRIAUX <i>et al.</i> (2015)	103
Annexe B : Abaques de Michael Williams	104
Annexe C : Arbres microphoniques supplémentaires	108
Annexe D : Lois de Panoramiques, “SPAT Revolution - Panning Algorithms”, 2024	115
Annexe E : Questionnaires des différents tests	118
Annexe F : Pouvoir séparableur selon différents plans, BLAUERT, 1996	128

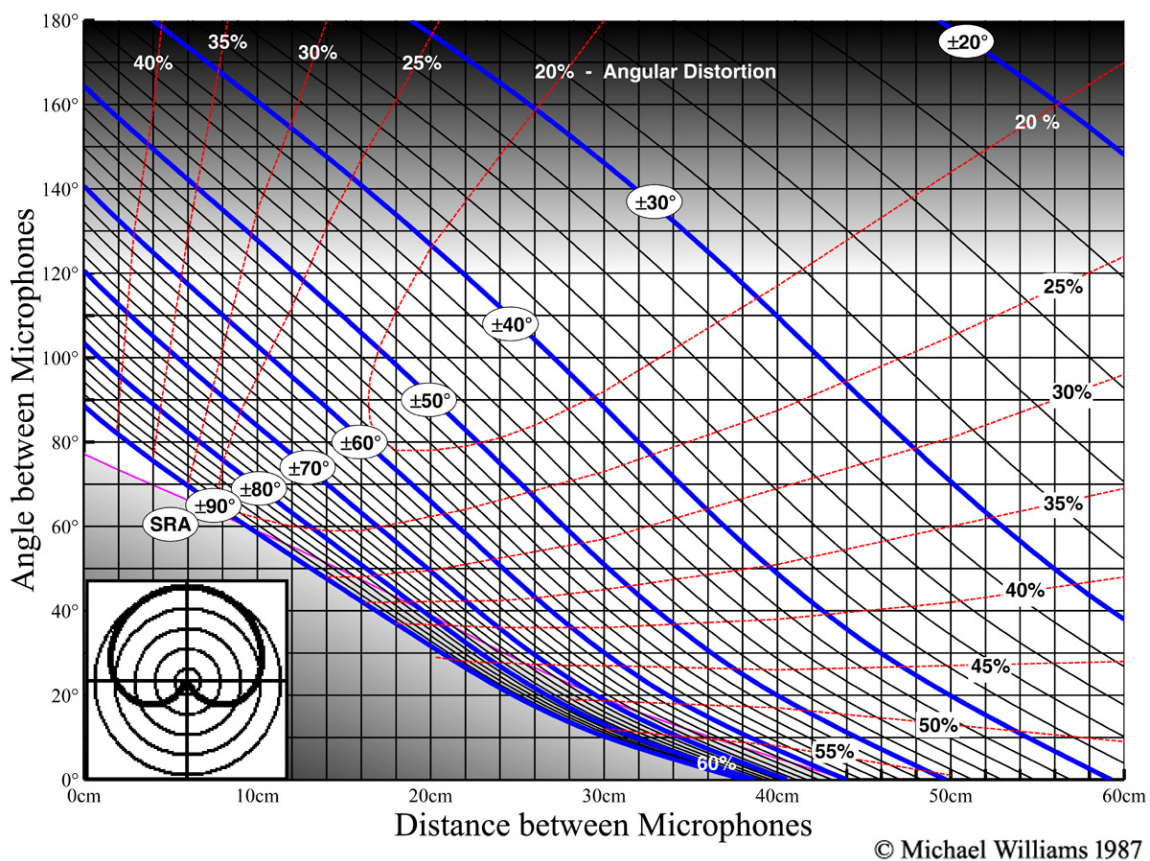
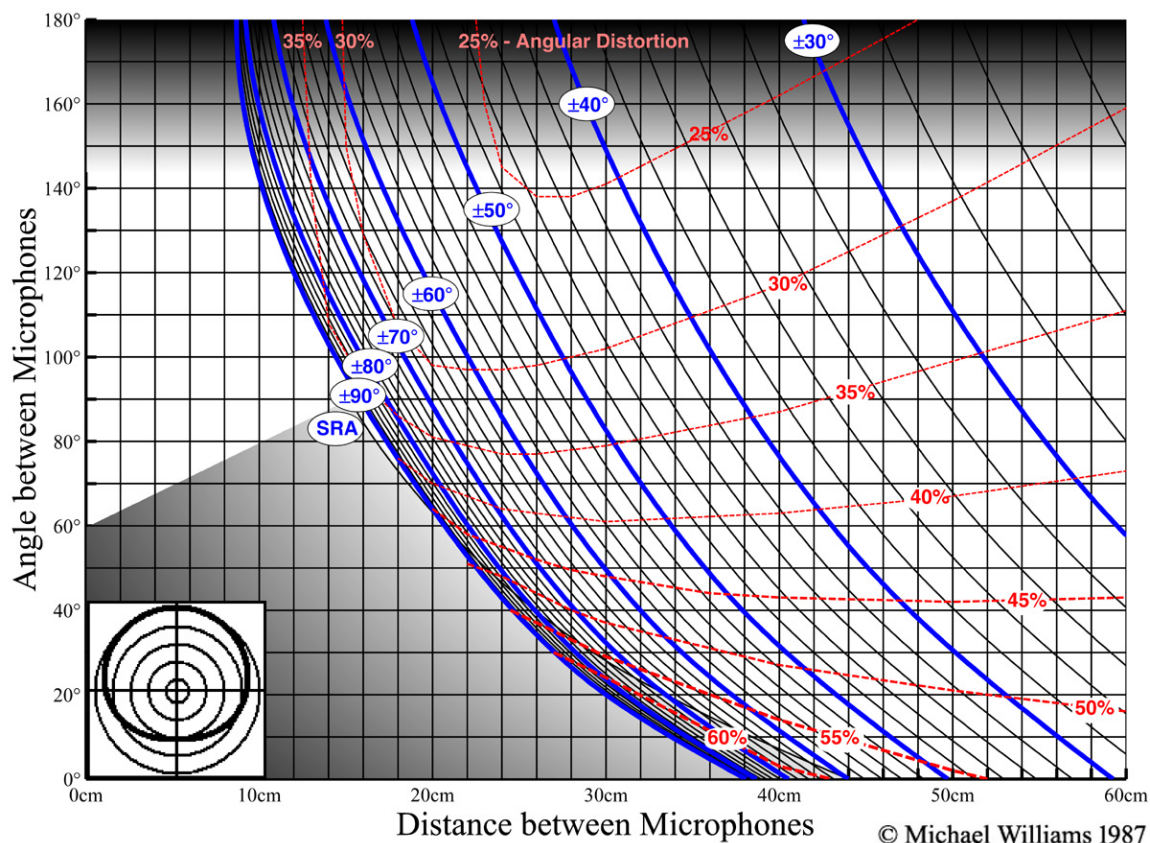
Annexe A : Frise des systèmes sonores au cinéma, Périaux *et al.* (2015)

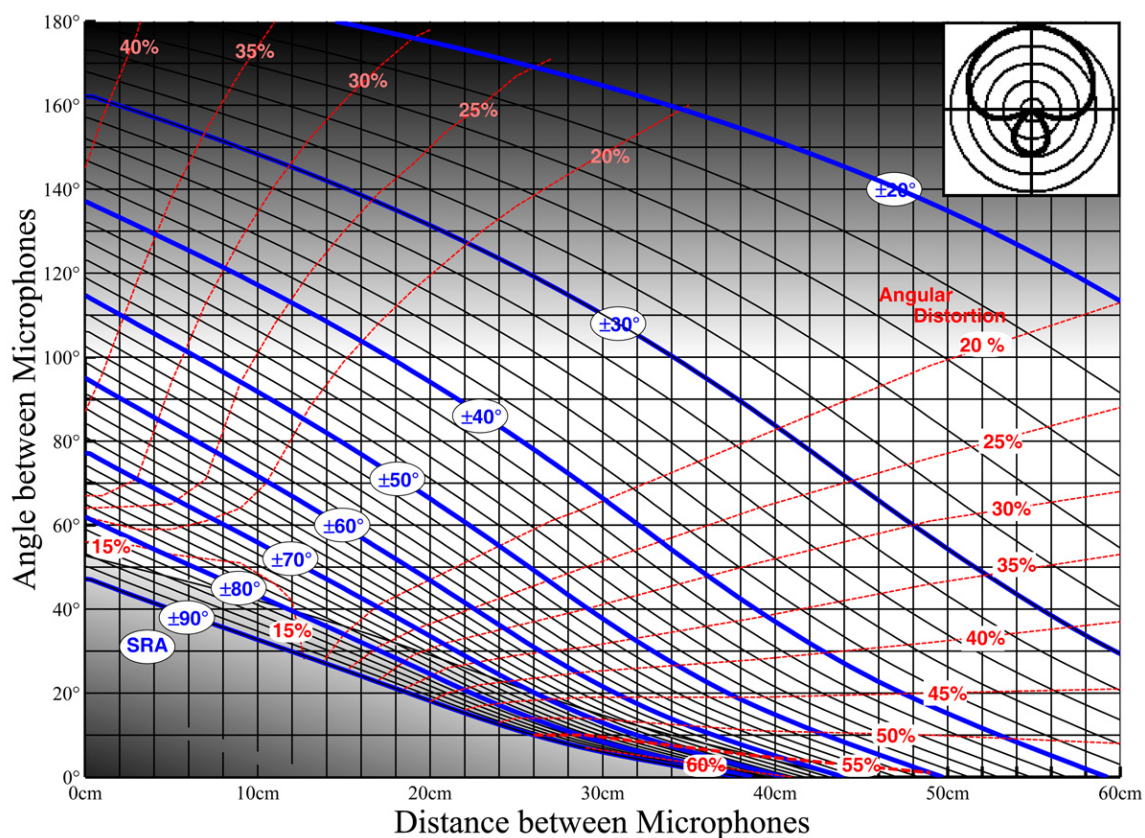
HISTORIQUE CINÉMA		
1931	Blumlein, 2 canaux L/R stéréo	analogiques et Dolby Stereo Digital, proposée par Cinemeccanica
1933	Bell Labs, 3 canaux L/C/R stéréo	
1940	<i>Fantasia</i> , procédé Fantasound 3 canaux diffusés en LCR/LsRs	1993 <i>Jurassic Park</i> , DTS 5.1 <i>Last Action Hero</i> , SDDS 7.1
1952	<i>This is Cinerama</i> , procédé Cinérama 7 canaux LLcCRcR/LsRs	1995 Présentation au Show East d'Atlantic City du processeur de cinéma numérique Dolby CP500 capable de lire à la fois les pistes analogiques optiques et la piste Dolby D
1953	<i>The Robe</i> , procédé Cinemascope 4 canaux LCRS, 20th Century Fox	
1955	<i>Oklahoma !</i> , procédé Todd AO 6 pistes LLcCRcR/S	1996 Environ 4 000 salles sont équipées dans le monde pour lire le Dolby Digital
1970	<i>Tiger Child</i> , le premier film IMAX, est projeté durant l'exposition universelle de 1970, à Osaka	Le Dolby Drive, système entièrement basé sur disque dur, conçu pour remplacer à terme les défileurs 35 mm magnétiques perforés, est présenté à l'AES de Copenhague
1974	<i>Earthquake</i> procédé Sensurround, apparition des Subwoofers	
1976	<i>A star is born</i> (version Barbara Streisand, Kris Kristofferson), premier film 35 mm avec des effets surround et première utilisation de la matrice 4 : 2 : 4 du Dolby Stereo	1998 15 000 salles de cinéma équipées en Dolby Digital dans le monde, 14 000 en DTS
1977	<i>Star Wars</i> , première utilisation du canal Lfe	1999 <i>La Menace fantôme</i> , Dolby Digital Surround EX, format 6.1 co-développé par Dolby et Lucas Film THX
1977	<i>Close Encounters of the Third Kind</i> , Canal Lfe dédié	20 000 salles équipées en Dolby Digital
1978	<i>Superman</i> , premier film magnétique 70 mm expérimentant des canaux arrière stéréo	2000 DTS ES 7.1 Plus de 25 000 salles équipées Dolby Digital
1979	<i>Apocalypse Now</i> premier film en Dolby Stereo six pistes magnétiques 70 mm à être distribué commercialement avec un son surround stereo	2006 Cinéma Numérique, <i>Chicken Little</i> Dolby Digital Cinema
	<i>Don Giovanni</i> , premier mixage film Dolby Stereo en France	2009 <i>Avatar</i> , regain d'intérêt pour le film en 3D
1983	<i>Return of the Jedi</i> , Norme THX	2010 <i>Toy Story 3</i> , Dolby 3D 7.1
1986	Le format Dolby Stereo SR (spectral recording) 35 mm optique est présenté à l'Academy of Motion Picture Arts and Sciences	2011 <i>Mission impossible 4</i> , sortie au format IMAX
1987	Premiers films en Dolby Stereo SR <i>Innerspace</i> , <i>Robocop</i> . Le format 5.1 est normé pour le film	2012 <i>Red Tails</i> , Lucas Film, sortie du premier film Auro-3D 11.1 en salle de cinéma
1990	Sortie du centième film en Dolby Stereo SR <i>Robocop 2</i>	2012 <i>The Impossible</i> , sortie au format Imm Sound 14.1
1991	Premières démonstrations du format Dolby Stereo Digital (Dolby SRD) intégrant l'AC3 en complément des 2 pistes LtRt	2012 Dolby ATMOS, 128 canaux, 64 haut-parleurs, sortie de <i>Rebelle</i> , Disney/Pixar, premier film en Dolby Atmos
1992	<i>Batman Returns</i> , Dolby Digital 5.1 (Dolby SRD)	2012 Dolby acquiert Imm Sound
1993	Première tête de lecture optique 35 mm capable de lire à la fois les bandes son	2012 Inauguration des 2 premières salles de cinéma équipées Dolby Atmos en France, à Paris et à Bordeaux
		2013 Développement du format d'échange MDA par DTS pour le mixage objets
		2015 Installations Dolby Atmos en France : 17 salles de cinéma, 2 salles de vision, 5 auditoriums, un labo DCP. 170 films produits en Dolby Atmos de 2012 à 2015.

Annexe B : Abaques de Michael Williams

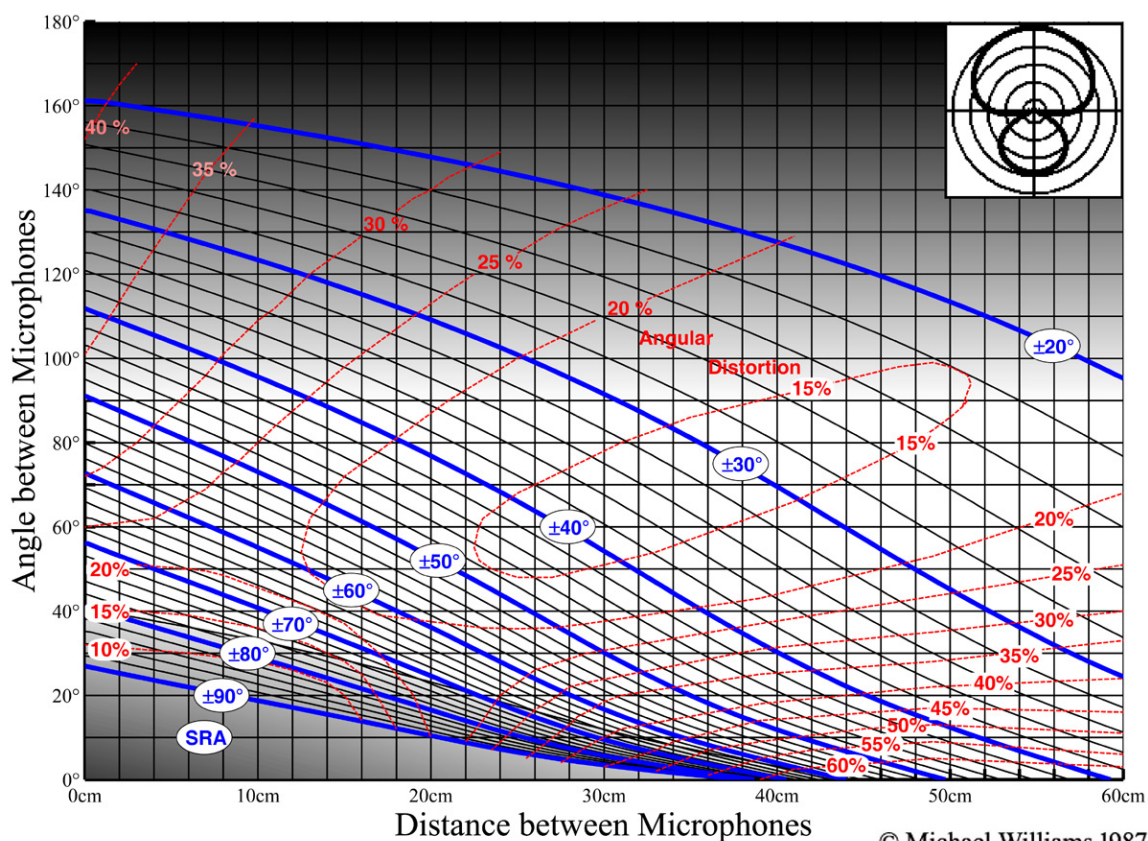
Voici dans cette annexe les différents abaques de Michael Williams (www.williamsmmad.com/MAD/2ch/2ch.htm) adaptés pour différentes directivités microphoniques. Nous avons en abscisses la distance entre les microphones, en ordonnée l'angle appliqué entre les microphones et nous pouvons observer sur le schéma les différents angles de prise de son, SRA en anglais.



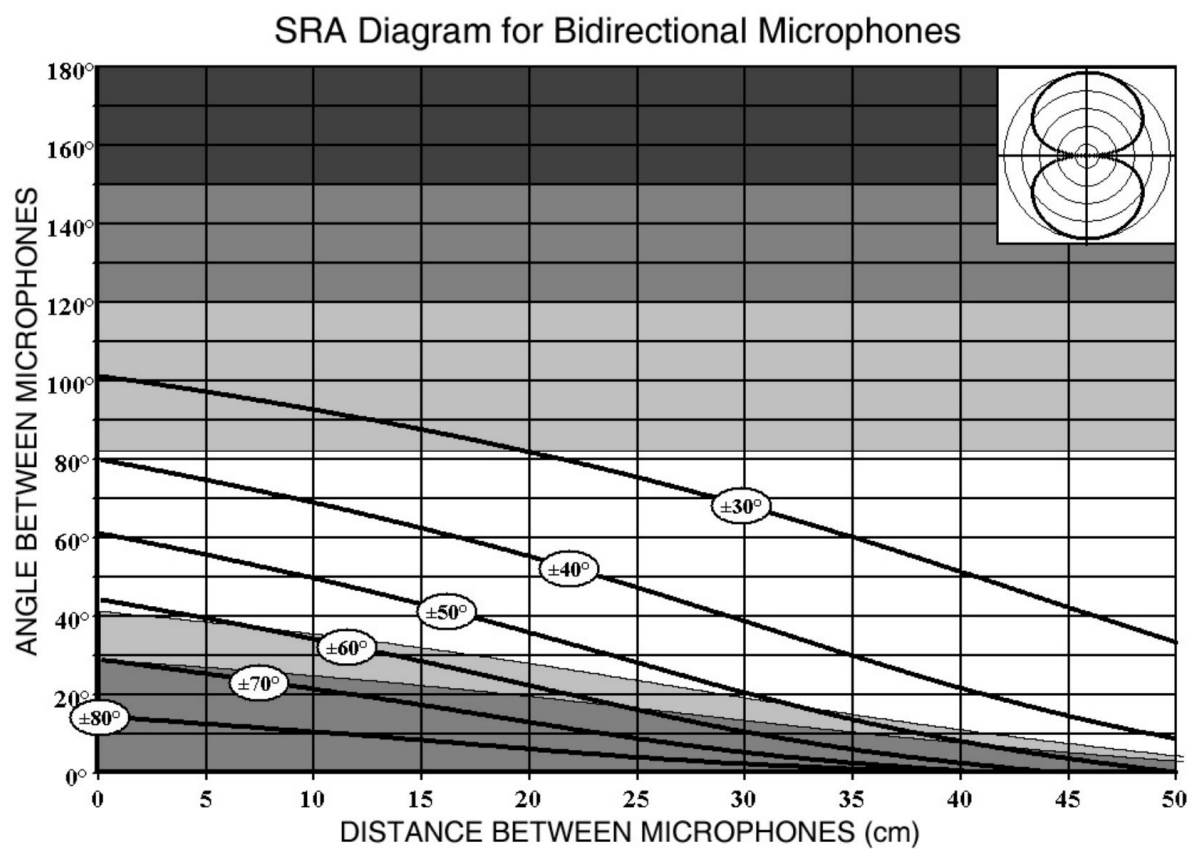




© Michael Williams 1987



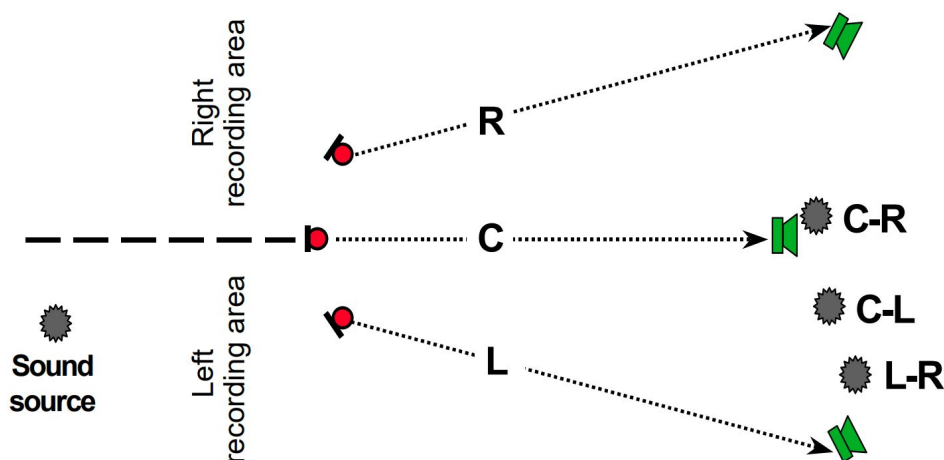
© Michael Williams 1987



Annexe C : Arbres microphoniques supplémentaires

Système OCT

Selon Günther Theile de l'IRT, une configuration classique composée d'un trio de microphones placés à l'avant crée des images fantômes (LC, CR, LR) qui peuvent nuire à la précision de localisation, au timbre et à la stabilité spatiale des sources, surtout lorsque l'auditeur ou l'auditrice se déplace.



Diaphonie acoustique qui résulte de la superposition de trois images fantômes (THEILE, 2001)

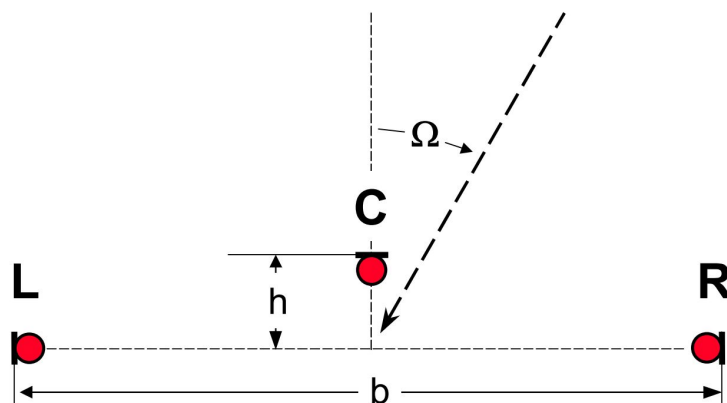
Le système OCT (*Optimised Cardioid Triangle*) est un système de captation à trois microphones qui a été conçu dans le but d'améliorer la stabilité de l'image centrale tout en réduisant la diaphonie⁵ entre canaux. Il se compose de deux capsules supercardioides (L et R) orientés à $\pm 90^\circ$, et d'un microphone cardiïde central (C). Cette disposition permet une bonne séparation des canaux et une reproduction cohérente de la scène sonore, même

⁵La diaphonie acoustique, *crosstalk* en anglais, se produit lorsque qu'un microphone capte partiellement un signal audio destiné à une autre capsule

si les capsules supercardioïdes sont utilisées hors de leur axe optimal.

Les distances typiques sont :

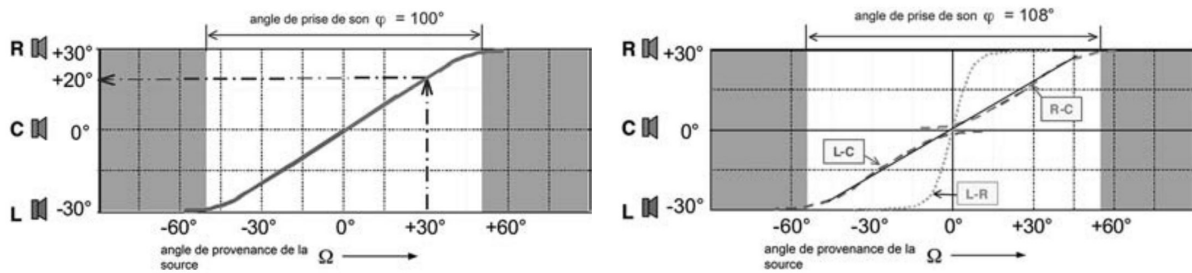
- b : entre L et R, ajustée selon l'angle de prise de son souhaité,
- h : décalage du micro central vers l'avant (8 cm avec des capsules L et R supercardioïdes, 12 cm avec des capsules L et R cardioïdes).



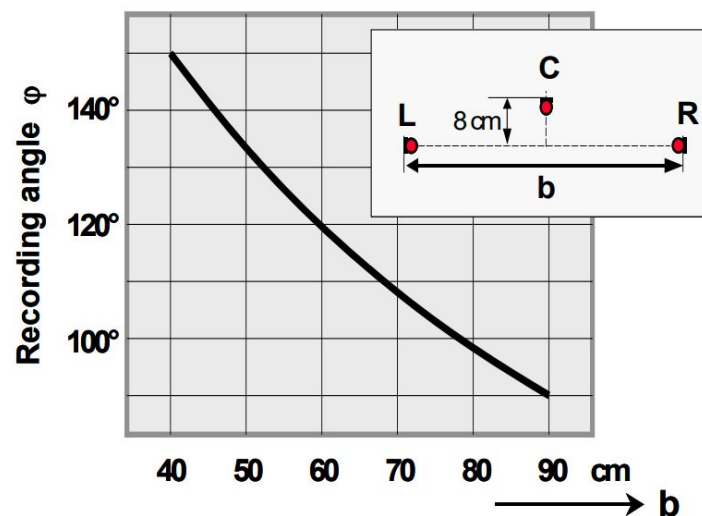
Configuration des trois capsules frontales du système OCT (THEILE, 2001)

Une variante, l'OCT2, pousse ce concept encore plus loin : le microphone central est avancé de 40 cm, ce qui nécessite un retard de 1 ms pour respecter les temps relatifs de l'OCT. Cette configuration, qui se rapproche du système Decca, augmente la décorrélation entre les microphones. Grâce à une diaphonie acoustique presque inexistante, l'OCT offre d'excellentes performances en downmix stéréo, performances qui sont encore améliorées avec l'OCT2.

Enfin, pour compenser les pertes en basses fréquences à cause de l'utilisation de capsules supercardioïdes, deux omnidirectionnels sont placés aux mêmes emplacements, filtrés au-dessus de 100 Hz et mélangés avec les signaux directionnels, garantissant ainsi une restitution sonore complète sur toute la largeur du champ.



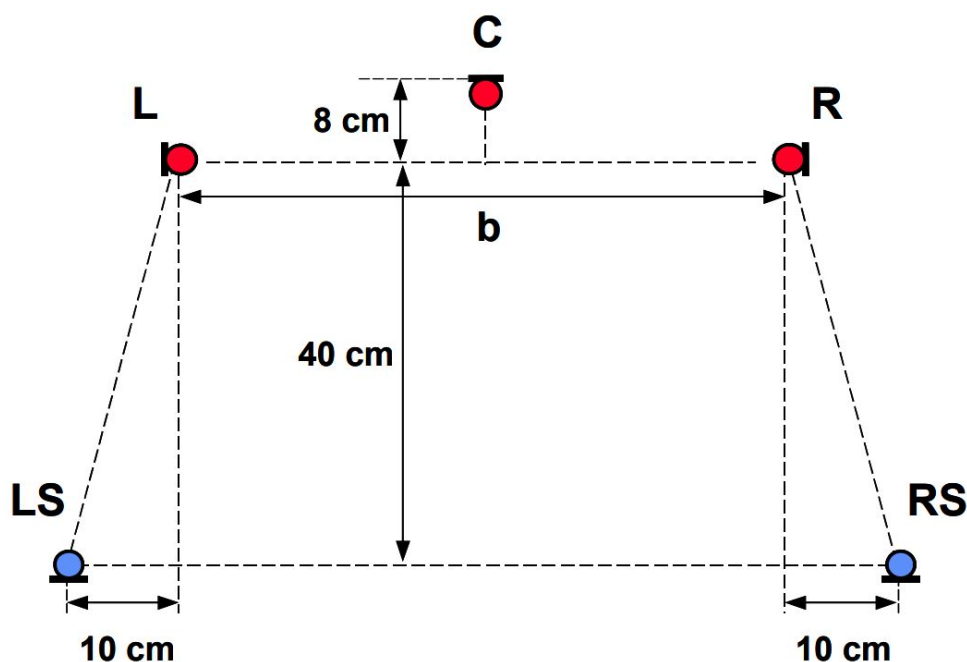
Courbe de localisation pour un système de prise de son stéréo (à gauche) et du système OCT où $b=70\text{cm}$ (à droite) (PÉRIAUX et al., 2015)



Angle de prise de son frontal en fonction de la distance b du couple L/R (THEILE, 2001)

OCT Surround

Pour obtenir un système de captation à cinq capsules, le système OCT Surround ajoute deux capsules cardioïdes, orientées à 180° et espacées de 40 cm du couple L/R. La distance entre les capsules Ls et Rs est 20 cm plus grande que la distance L/R. L'optimisation d'un arbre multicanal comme l'OCT Surround implique de trouver un équilibre entre son direct, réflexions précoces et réverbération, sans compromettre l'image sonore. Pour cela, les microphones avant (L, C, R) doivent restituer fidèlement à la fois les sons directs et les premiers échos dans leurs directions d'origine, tandis que les microphones surround (LS, RS), orientés vers l'arrière, doivent minimiser la capture du son direct frontal. Cette orientation permet de mieux isoler les réflexions latérales et arrière, essentielles pour une spatialisation réaliste. L'ensemble crée ainsi une image cohérente, avec une séparation claire des zones stéréophoniques et un bon équilibre entre champ direct et diffus, sans nécessiter de traitement complexe.



Configuration de l'OCT Surround (THEILE, 2001)

On peut obtenir une autre configuration multicanale du système OCT en le complétant par un système de captation dédié à la prise de son d'ambiance (croix IRT ou carré Hamasaki) placé en fond de salle. Il remplace ainsi les deux microphones arrière de l'OCT Surround et permet de renforcer l'ambiance sonore de la salle.

OCT SURROUND

- Micro central cardioïde avancé de 8 cm.
- Micros L et R supercardioïdes à $\pm 90^\circ$, distance b variable en fonction de l'angle de prise de son frontal souhaité.
- Micros arrière cardioïdes orientés à 180° , à 40 cm du triplet avant, distance entre eux = base $b + 10$ cm de chaque côté.
- Utilisation de deux omni pour le renfort des graves des canaux L et R.
- Utilisation fréquente de l'OCT avec un système d'ambiance Croix IRT ou Carré Hamasaki.

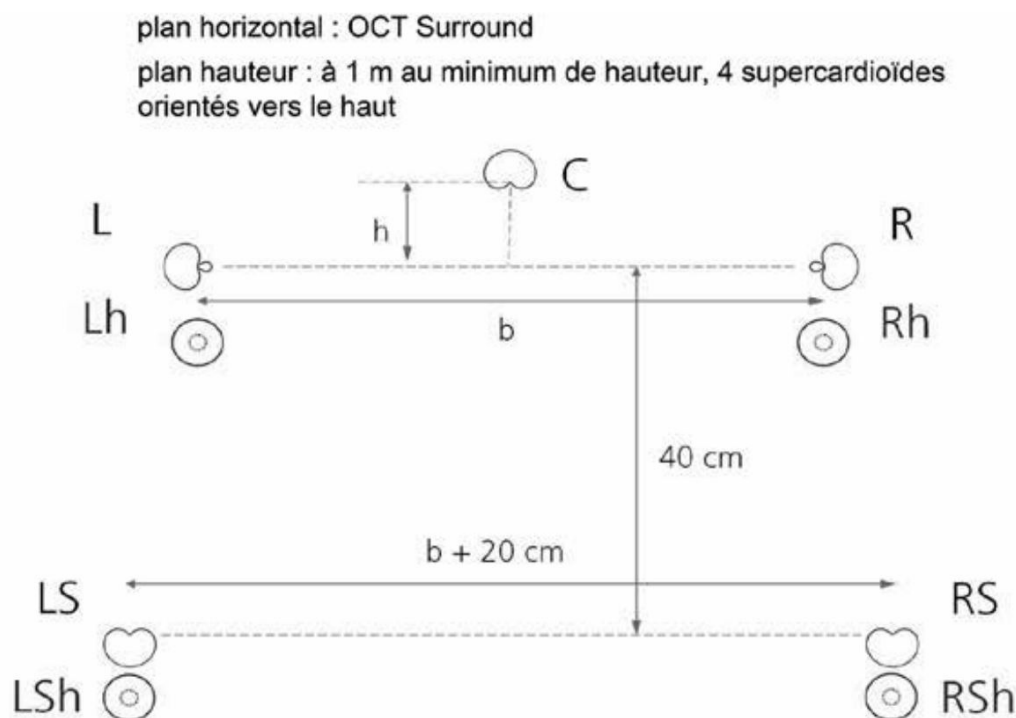
ANALYSE

- Très bonne localisation frontale, plans sonores bien restitués.
- Bonne impression spatiale.
- Manque de basses fréquences.
- Système peu modulable : orientations des micros fixes, directivités imposées.
- Utilisation pour la salle de concert.
- L'OCTS est plutôt utilisé pour des esthétiques frontales.

Récapitulatif des données de l'OCT Surround (PÉRIAUX et al., 2015)

OCT9 (3D)

L'OCT9 est une extension du système OCT Surround conçue pour la captation sonore avec élévation. Nous complétons le système OCT Surround par quatre microphones supercardioïdes placés au moins à 1 mètre au-dessus de la base horizontale. Ces capsules d'élévation (Lh, Rh, Lsh, Rsh), orientées vers le haut, sont positionnées respectivement au-dessus des microphones L, R, Ls et Rs.



Configuration du système OCT9 (PÉRIAUX et al., 2015)

Carré Hamasaki

Le carré Hamasaki est un système de microphones conçu pour capter principalement les sons latéraux et ambiants, tout en minimisant la présence du son direct. Il se compose de quatre capsules bidirectionnelles disposées en carré, espacés d'un, deux ou trois mètres et les lobes positifs sont orientés vers l'extérieur. Cette configuration, souvent placée en hauteur dans la salle pour capter un champ sonore diffus, est particulièrement efficace pour restituer l'ambiance naturelle sans altérer l'image frontale. Elle offre une image sonore homogène et apporte une *"bonne impression spatiale"* (PÉRIAUX et al., 2015). La sensation d'espace dépend de l'écartement des microphones : plus le carré est grand, plus l'image paraît large.

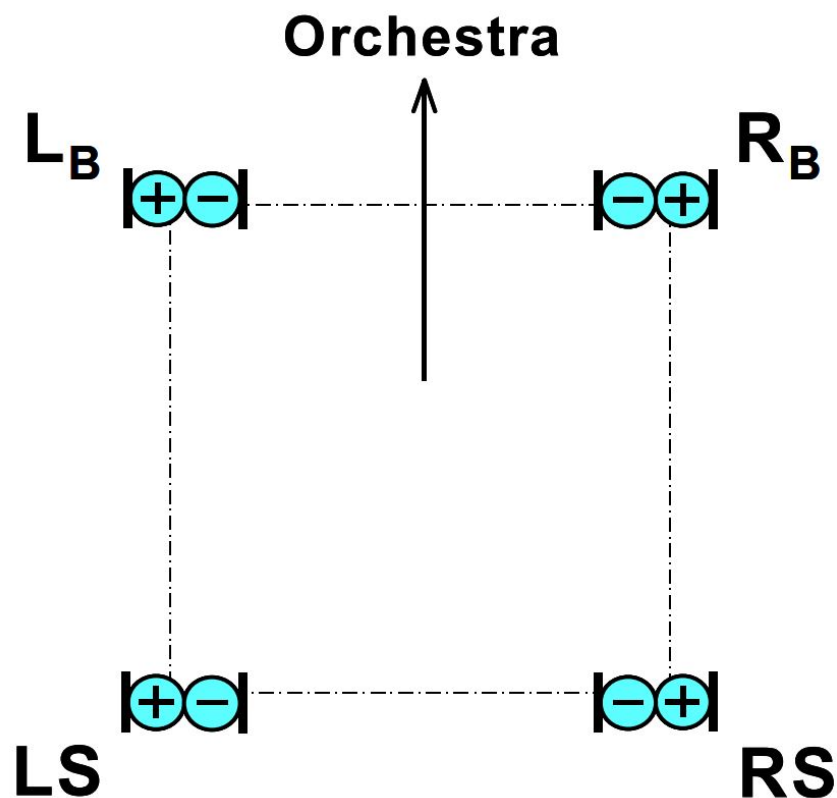
CARRÉ HAMASAKI

- 4 bidirectionnels disposés en carré, dimension du côté de 2 à 3 m, les lobes positifs étant orientés vers les murs.

ANALYSE

- Pas de son direct.
- Très bonne impression spatiale.
- Construction du champ réverbéré très homogène.
- Basses fréquences atténuées.
- Utilisation pour la captation d'ambiances en salle de concert.

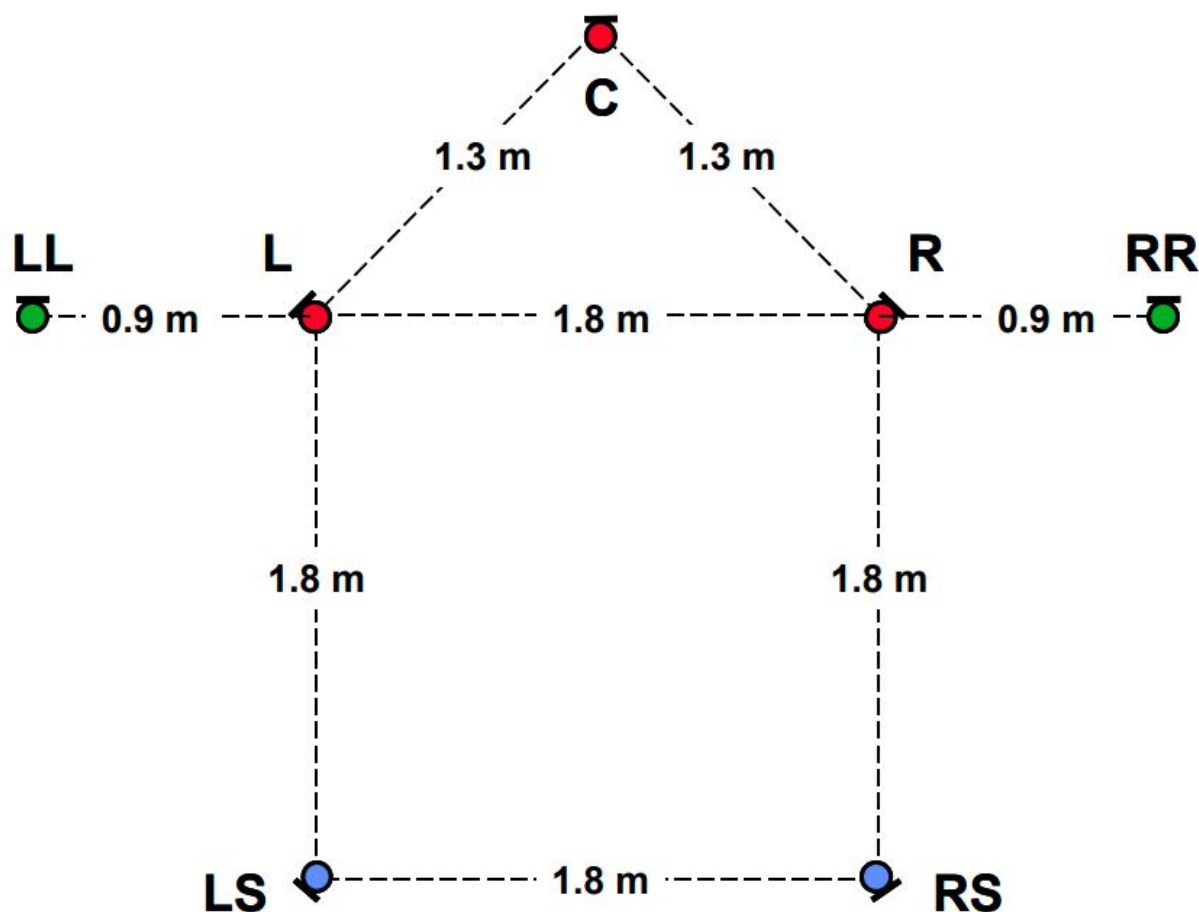
Récapitulatif des données du carré Hamasaki (PÉRIAUX et al., 2015)



Configuration du carré Hamasaki (THEILE, 2001)

Arbre Fukada

L'arbre Fukada, inspiré du Decca Tree, utilise cinq microphones cardioïdes (L, C, R, LS, RS) disposés en carré espacé d'environ 1,8 m entre les capsules L, R, Ls et Rs. Deux microphones omnidirectionnels (LL, RR) placés en renfort latéral sur l'axe du couple L/R élargissent la scène et favorisent la transition entre les canaux avant et arrière. Cette configuration, adaptable selon l'acoustique de la salle et la disposition de l'orchestre, ne vise pas une image stéréo précise, mais plutôt une sensation d'espace avec une image centrale stable. Elle est généralement positionnée quelques mètres au-dessus et derrière le chef d'orchestre, avec des réglages affinés à l'écoute. Les ouvertures des couples microphoniques L/R et Ls/Rs sont variables.



Configuration d'un arbre Fukada, les capsules cardioïdes sont représentées en rouge et bleu et les capsules omnidirectionnelles en vert (THEILE, 2001)

Annexe D : Lois de Panoramiques, “SPAT Revolution - Panning Algorithms”, 2024

8.3 Vector Base Intensity (VBIP)

Vector Base Intensity Panning is a similar variation to the VBAP technique. It can also reproduce a 2D or 3D immersive sound field with sharply localized virtual sound sources.

How does it work?

VBIP was designed to improve on VBAP when calculating the high-frequency (above 700 Hz) localization criteria. The selection of which speakers to use to render a virtual sound source is similar to VBAP, only the gain calculations differ.

The same four dependencies mentioned for VBAP, also apply to VBIP. You will need to listen for quite a nuanced difference between these two panning algorithms. Try to compare how each panning type handles the higher frequency content of your material.

8.4 Dual Band Vector Based Panning (VBP Dual-Band)

Both Intensity and Amplitude Vector Based panning have an ideal frequency range of action:

- Localization of low frequencies is better with Amplitude Panning.
- Localization of high frequencies is better with Intensity Panning.

A hybrid approach of vector-based panning has been developed in this way: the Dual Band Vector Based Panning. This panning type merges the two approaches in order to combine the best of both worlds and to reach a better localization. Amplitude panning is applied below the crossover frequency, while intensity panning is applied above. The crossover frequency has been defined to 700 Hz by default.

The dependencies mentioned in the VBAP section also apply to Dual Band Vector Based Panning.

8.5 Layer-based amplitude panning (LBAP)

Layer-based amplitude panning can be explained as **multiple 2D VBAP layers**: The speaker setup is split into several layers, depending on the speaker elevation. The panning used between speakers on the same layer is the VBAP 2D. Between these layers, a crossfade is applied between the two nearest layers.

Note

The difference between VBAP 3D and LBAP is the number of speakers that will be active between the layers: three in VBAP versus four in LBAP.

8.7 K-Nearest Neighbor (KNN)

KNN is another panning type that does not depend on a *Sweet Spot* to be perceived correctly. It is a version of a ‘Nearest-neighbor’ interpolation algorithm. This family of algorithms is also used in the fields of complex systems, 3D graphics and network science to name a few. In *SPAT Revolution*, you can sonically explore a network of loudspeakers using this panning type and some virtual sound sources.

How does it work?

An interesting parameter of KNN is that the user gets manual control over one of the main coefficients in the underlying algorithm. The parameter is called *Nearest Neighbor spreading*. It sets a maximum limit to speakers’ number that the algorithm can use as neighbors. The parameter becomes available as a continuously variable percentage for each virtual source in a *SPAT Revolution* room.



Figure 8.1: Source Spreading

What makes this particularly interesting is that different sources can activate less or more of the sound system dynamically and in a very smooth way. For example, one virtual sound source might seem to pop in and out of individual speakers because its *Nearest Neighbors Spread* parameter is set at a low percentage, while another sound source could seem to diffuse over the entire sound system because its spread variable is set to 100%. For example, on a 10-speaker arrangement: 1-10% will use 1 speaker, 11% to 20% 2, and so on.

i Note

Try automating the Nearest Neighbors Spread in a relationship with another source property of the same sound sources such as room presence.

8.8 Speaker-Placement Correction Amplitude (SPCAP)

SPCAP is a 3D panning algorithm that takes its inspiration from VBAP. SPCAP selects not just 2 or 3, but any number of speakers to render a virtual source and weights signal gains according to how much each selected speaker is actually contributing to the overall power

output of the speaker configuration. Using this method SPCAP guarantees conservation of loudspeakers power output across any speaker arrangement. Its strengths lie in the down-mixing and up-mixing of virtual scenes from very different channel-based speaker arrangements, and in being able to render wider sound sources by smartly using more speakers.

How does it work?

The result will still be *Sweet Spot* dependent, although it will be a wider listening area. SPCAP inherits some dependencies of VBAP to get successful spatial imaging.

1. Speakers must be placed *around* the listening position.
2. 2D speakers should be on the same horizontal plane as the ears.

i Note

SPCAP panning can do a good job of translating surround audio mixes from one speaker configuration to another.

8.9 Ambisonic Equivalent Panning (AEP)

In common with the channel-based panning types we have covered so far, Ambisonics is a technology that also distributes virtual sound sources in space. Yet it achieves this in a fundamentally different way. Ambisonics rely on a two-step process.

1. **Encoding** Audio sources along with their positional information are wrapped up together using signal mathematics to create encoded Ambisonic audio. Ambisonic scenes are always carried on at least three channels of audio. They are not intended to be *listened to directly* they are intended to be *decoded*.
2. **Decoding** Ambisonic audio signals are unwrapped and the positional information contained within them is decoded *specifically* for one type of speaker configuration. What we get is an immersive sound field that should accurately render the original spatial composition in 2D or 3D on the specified speaker configuration.

Keeping these two steps separate has a number of advantages. Primarily, that of being able to record the encoded Ambisonic audio signals independently of any fixed speaker arrangement. On the other hand, it is possible to “fuse” the two stages of the process together resulting in what appears to be the output of a generalized channel-based type of panning. That is the AEP panning type in a nutshell.

How does it work?

AEP has certain computational and ambisonic mixing advantages and exhibits very different behavior from the VBAP/VBIP pairwise approaches. It is up to you to decide whether to work with pure Ambisonic rooms (more about that in the later section) or to use AEP as

a channel-based panning law. Both approaches are valid and could be useful. As we have mentioned a few times already, the choice of panning type depends on what sounds best in the context of your material, your compositional goals and the acoustics of the system you are working with.

Annexe E : Questionnaires des différents tests

Cette annexe regroupe les questionnaires des deux tests réalisés dans le cadre de ce mémoire : d'abord le test perceptif, puis le test pratique.

TEST DE PERCEPTION SONORE

Questions préliminaires :

- Date :
- Nom/Prénom ?
- Quel âge avez-vous ?
- Quel est votre emplacement ?
- Session (Vous sera indiquée en début de test, veuillez entourer la bonne réponse)

A / B / C

Consigne

Je tiens tout d'abord à vous remercier de participer à ce test !

Vous allez écouter une série de sons traitée spatialement de quatre manières différentes. Le but de ce test est d'évaluer votre capacité à localiser un son dans différentes conditions de restitution. Vous allez entendre le même son présenté de quatre manières différentes, chacune associée à une même valeur de panoramique. Une seconde écoute vous sera proposée. Cette série sera ensuite répétée pour deux autres sons. Nous ne prendrons pas en compte les variations de timbre dans ce test.

Le test se remplit en entourant les réponses de votre choix ou en grisant l'une des parts du cercle autour des silhouettes (vous pouvez dessiner plusieurs cases si la source sonore vous paraît plus large).

Je vous souhaite un bon test !

Voix parlée

Cas V1 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

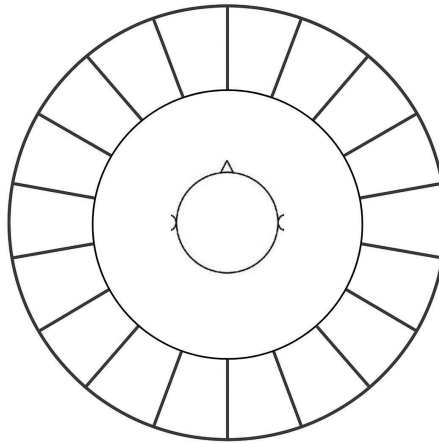
2

3

4

5

Dans quelle direction avez-vous
perçu la source ?



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Cas V2 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

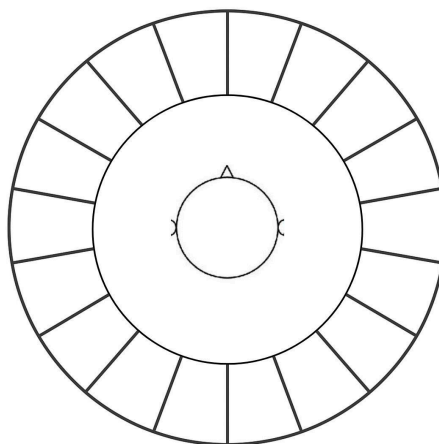
2

3

4

5

Dans quelle direction avez-vous
perçu la source ?



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Cas V3 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

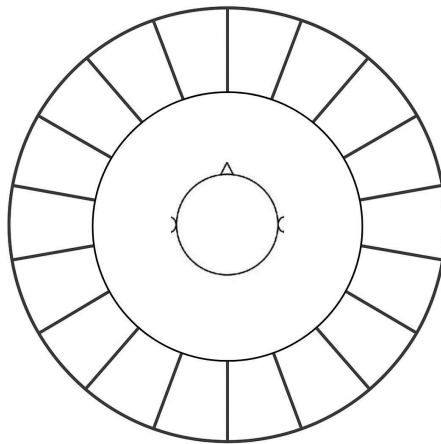
2

3

4

5

**Dans quelle direction avez-vous
perçu la source ?**



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Cas V4 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

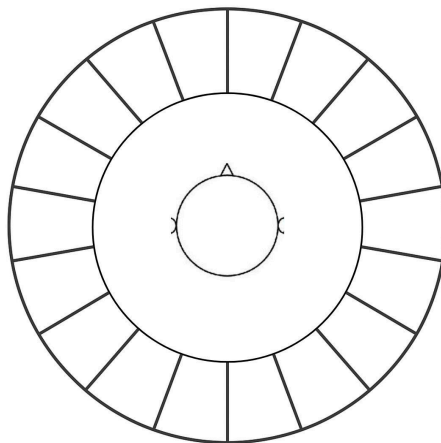
2

3

4

5

**Dans quelle direction avez-vous
perçu la source ?**



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Avez-vous préféré un cas en particulier ?

Cas V1

Cas V2

Cas V3

Cas V4

Aucun des cas

Guitare

Cas G1 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

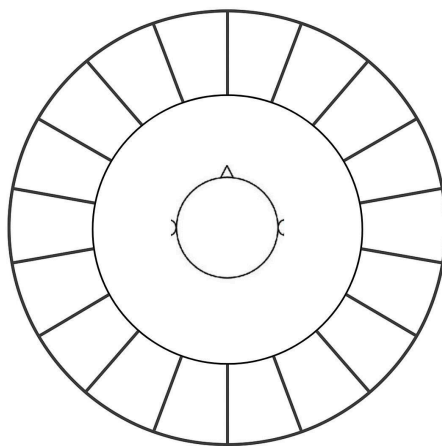
2

3

4

5

Dans quelle direction avez-vous
perçu la source ?



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Cas G2 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

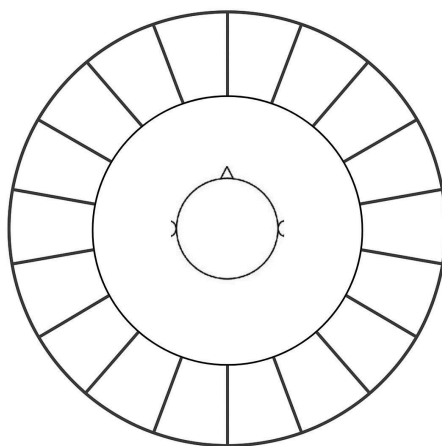
2

3

4

5

Dans quelle direction avez-vous
perçu la source ?



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Cas G3 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

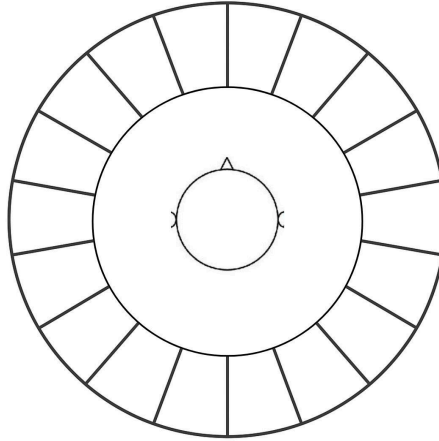
2

3

4

5

**Dans quelle direction avez-vous
perçu la source ?**



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Cas G4 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

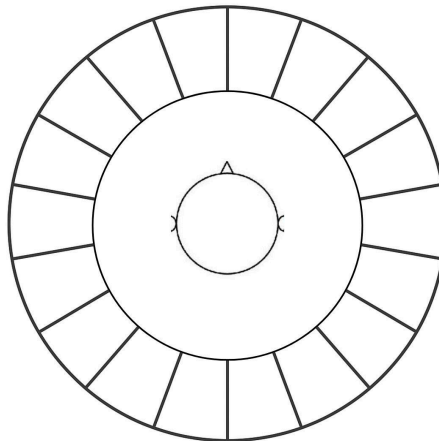
2

3

4

5

**Dans quelle direction avez-vous
perçu la source ?**



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Avez-vous préféré un cas en particulier ?

Cas G1

Cas G2

Cas G3

Cas G4

Aucun des cas

Darbouka

Cas D1 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

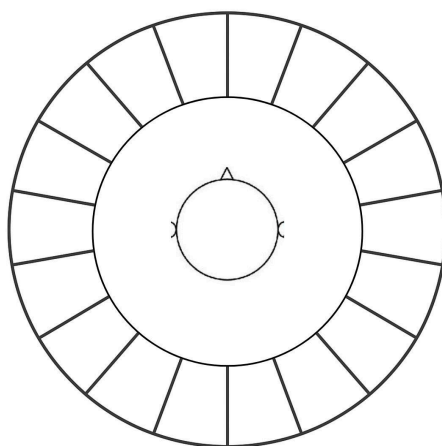
2

3

4

5

**Dans quelle direction avez-vous
perçu la source ?**



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Cas D2 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

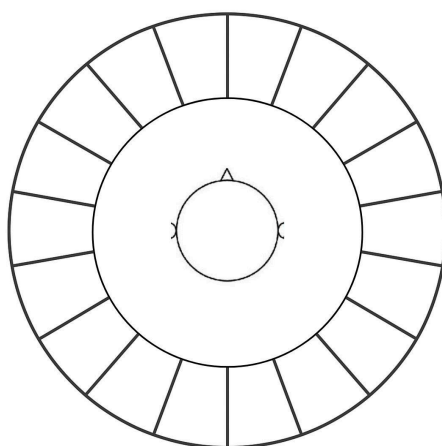
2

3

4

5

**Dans quelle direction avez-vous
perçu la source ?**



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Cas D3 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

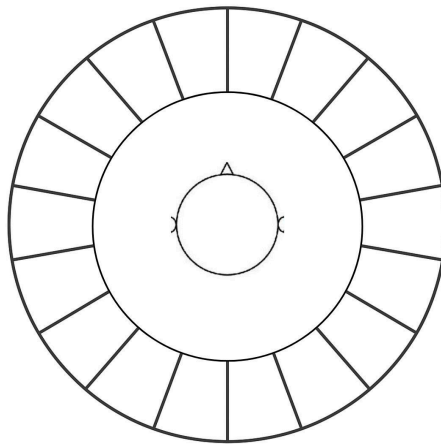
2

3

4

5

**Dans quelle direction avez-vous
perçu la source ?**



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Cas D4 :

Sur une échelle de 1 à 5, à quel point la source sonore est-elle facile à localiser ?

Je n'ai pas
pu localiser

J'ai pu localiser
très clairement

1

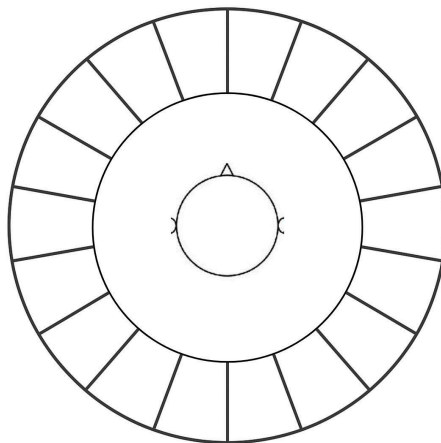
2

3

4

5

**Dans quelle direction avez-vous
perçu la source ?**



Sur une échelle de 1 à 5, à quel point la source sonore est-elle précise ?

Très diffuse/large

Très précise

1

2

3

4

5

Avez-vous préféré un cas en particulier ?

Cas D1

Cas D2

Cas D3

Cas D4

Aucun des cas

Retours Généraux

Avez-vous d'autres commentaires ou observations ?

TEST PRATIQUE

Je tiens d'abord à vous remercier grandement de participer à ce test.

Ce test vous laisse la liberté de latéraliser des sources sonores à partir de l'outil qui vous est mis à disposition.

Informations générales :

Nom et prénom :

Profession :

Questions :

Ergonomie :

- Est-ce que les paramètres de l'outil (déplacement de la source sonore, ajustement des paramètres, etc.) sont facilement accessibles ? Sont-ils suffisants, insuffisants ou trop nombreux ?
- Avez-vous essayé de changer les paramètres de l'outil (positionnement, orientation, directivité) ? Si oui, avez-vous trouvé cela intuitif ?

Utilisation :

- Avez-vous entendu des différences perceptives lorsque vous changez les paramètres de l'arbre microphonique ? Si oui, quelles différences avez-vous perçues ?
- En comparant l'outil proposé à ceux que vous pourriez utiliser au quotidien, avez-vous perçu des différences perceptives notables ? Si oui, trouvez-vous que ces choix pourraient impacter significativement le résultat final d'un mixage ?

Remarque générale :

- Quelles fonctionnalités ou améliorations souhaiteriez-vous voir ajoutées à l'outil ?

Annexe F : Pouvoir séparateur selon différents plans, Blauert, 1996

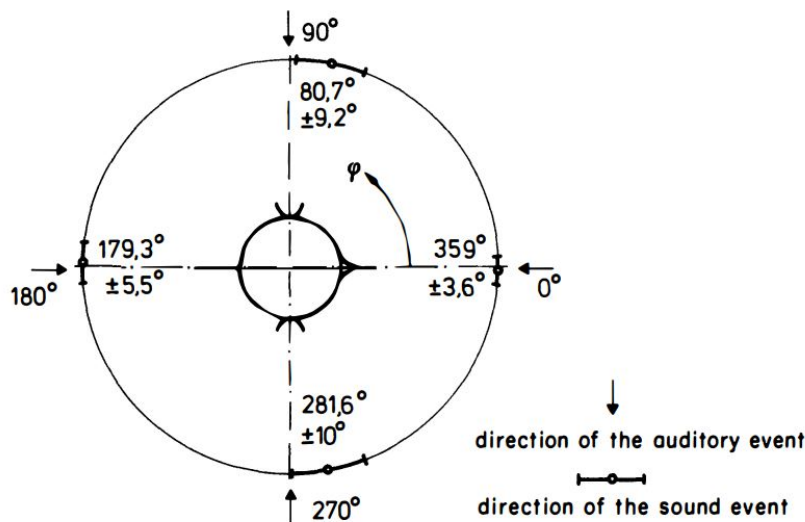


Figure 2.2
Localization blur $\Delta\varphi_{\min}$ and localization in the horizontal plane (after Preibisch-Effenberger 1966a and Haustein and Schirmer 1970; 600–900 subjects, white-noise pulses of 100 ms duration, approximately 70 phon, head immobilized).

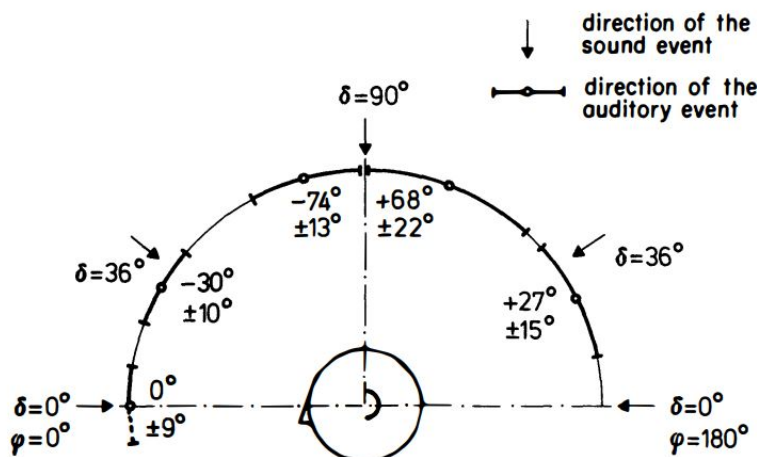


Figure 2.5
Localization and localization blur in the median plane for continuous speech by a familiar person (after Damaske and Wagener 1969; 7 subjects, 65 phon, head immobilized). Note that the view is different from that in figure 2.2.