



Caractériser l'immersion sonore cinématographique : application au cas des ambiances multicanales

Thibaud Carcy

Mémoire de Master 2 - Spécialité Son

Directeur de mémoire interne : Sylvain LAMBINET

Directeur de mémoire externe : Cyril HOLTZ

Examineur : Alan BLUM

Référent académique : Laurent MILLOT

Coordinateur des mémoires : Corsin VOGEL

Juin 2023

Résumé

Ce travail de recherche présente une analyse des rapports existant entre l'immersion cinématographique et l'utilisation de systèmes de diffusion sonore multicanaux. Dans un contexte de déploiement en salles de dispositifs dits « immersifs » depuis les années 2010 (Dolby Atmos, Auro 11.1, DTS:X, etc.), l'étude vise à identifier les indices d'une potentielle immersion chez les spectateurs et à mesurer l'impact d'un format de diffusion sur celle-ci. Ce travail comprend deux parties.

La première présente une revue de la littérature scientifique et technique sur le sujet. Une attention particulière a été portée sur le statut qu'occupent les ambiances au sein de la bande sonore. Nous verrons que celui-ci interroge leur mode de diffusion.

La seconde partie retrace une expérience réalisée pour mesurer l'influence d'un format de diffusion sur la sensation d'immersion d'un sujet dans le cadre de projections audiovisuelles d'ambiances. La perception des caractéristiques spatiales des espaces présentés constituait également un critère d'étude. Quatre formats ont été comparés (LCR, 5.0, 7.0 et 7.0.2) et six critères d'évaluation ont été établis. Un dispositif d'enregistrement multicanal a été développé pour la captation de l'ensemble des stimuli nécessaires à l'expérience. Huit extraits de ces enregistrements ont été diffusés à vingt-huit sujets qui, après chaque visionnage, devaient répondre à un questionnaire reprenant les critères d'évaluation. L'analyse des résultats a démontré des différences de réponses statistiquement significatives entre les formats pour les critères d'enveloppement (sensation de se sentir entouré de son) et de présence (sensation de se sentir projeté dans un espace virtuel). Ces sensations étaient accrues lors de l'augmentation du nombre de canaux autour du sujet. L'expérience a également montré que la nature des espaces projetés, la présence de sources synchrones à l'image et au son, ainsi que la cohérence spatiale entre ces deux champs pouvaient avoir un impact sur l'immersion des sujets.

Mots-clés : immersion ; système immersif ; format de diffusion ; prise de son multicanale ; ambiance sonore ; son spatialisé ; enveloppement ; présence

Abstract

This research presents an analysis of the relationship between immersion in a film and the use of multichannel sound systems. In the context of immersive sound systems in movie theatres since the 2010s (Dolby Atmos, Auro 11.1, DTS:X, etc.), this study aims to identify the criteria for a potential spectator's immersion and to measure the impact of the speaker layout on it. This work consists of two parts.

The first part presents a review of the available scientific and technical findings on the subject. Particular attention has been paid to the role of ambient sounds within the soundtrack. We will see that this raises questions about their diffusion model.

The second part describes an experiment conducted to measure the influence of a sound field configuration on the sense of immersion experienced by a subject in the context of ambient sound screening. The perception of the spatial characteristics of the presented spaces was also a study criterion. Four speaker layouts were compared (LCR, 5.0, 7.0 and 7.0.2) and six evaluation criteria were established. A multichannel recording device for the recording of all the stimuli required for the experiment was also developed. Eight excerpts from these recordings were played to twenty-eight subjects, who were asked to complete a questionnaire after each viewing, regarding the evaluation criteria. Analysis of the results showed statistically significant differences in responses between layout for the envelopment (feeling of being surrounded by sound) and presence (feeling of being in an environment other than that in which we are physically). These sensations were enhanced as the number of audio channels around the subject increased. The experiment also showed that the nature of the screened spaces, the presence of synchronous audiovisual sources and the spatial coherence between these two fields could have an impact on the immersion of the subjects.

Keywords: immersion; immersive sound system; speaker layout; multichannel recording; ambient sound; spatial sound; envelopment; presence

Remerciements

Je tiens à remercier particulièrement Sylvain Lambinet et Cyril Holtz, qui, par leurs conseils et leur implication constante, m'ont permis d'évoluer dans un cadre de travail rigoureux et serein.

Merci également à Antoine Martin, responsable de la marque HAL, ainsi qu'à Mélia Roger, artiste sonore, pour avoir partagé avec moi leur expérience. Dans un contexte de travail particulièrement actif, je les remercie d'avoir pris le temps de répondre avec précision à l'ensemble de mes questions.

Merci à Corsin Vogel et Valentin Bauer pour avoir suivi ce projet tout au long de l'année.

Merci aussi à Florent Fajole, responsable du centre de ressources documentaires de l'école, qui, par sa disponibilité et ses suggestions avisées, m'a permis d'établir un corpus bibliographique scientifique et technique. Ces textes ont fondé ma réflexion personnelle sur le sujet de l'immersion cinématographique et de la diffusion multicanale.

Merci à Nathalie Montier de la société Tapages et Nocturnes, Charles Bussienne et Diane Prieur de la société Poly Son Post Production, Agnès Hominal, Franck Jouanny et Pierre Chevrin de l'École Nationale Supérieure Louis-Lumière ainsi que Michel Casang, ingénieur du son, pour leur accompagnement logistique et technique dans la mise en œuvre des enregistrements sonores et de leur diffusion dans le cadre de l'expérience menée.

Merci également aux mairies de Cahors et de Trespoux-Rassiels ainsi qu'à l'ensemble des personnes qui m'ont accueilli pour la réalisation de ces mêmes enregistrements.

Merci à tous les volontaires qui ont participé aux sessions d'écoute ainsi qu'aux professionnel·le·s avec qui j'ai pu m'entretenir pour la réalisation de cette recherche.

Enfin, merci à ma famille et mes amis pour leur soutien constant durant cette période ainsi que pour leurs retours précieux durant les phases de relecture.

Table des matières

RÉSUMÉ	2
ABSTRACT	3
REMERCIEMENTS	4
TABLE DES MATIÈRES	5
INTRODUCTION	7
PREMIÈRE PARTIE – PRISE DE SON ET DIFFUSION D’AMBIANCES MULTICANALES : IDENTIFICATION DES INDICES D’IMMERSION	9
I – IMMERSION	9
1.1 – <i>L’immersion comme expérience cognitive et sensorielle</i>	<i>10</i>
1.1.1 – Le cinéma affectif : absorption et mise en phase	10
1.1.2 – L’immersion sensorielle.....	11
1.2 – <i>Critères de définition et de mesure</i>	<i>12</i>
1.2.1 – Propositions de critères perceptifs.....	13
1.2.2 – Questionnaires de présence et d’immersion.....	14
1.3 – <i>Les systèmes immersifs : une caractéristique objective</i>	<i>16</i>
1.4 – <i>Immersion sonore et nature des sources</i>	<i>18</i>
II – AMBIANCES	20
2.1 – <i>Définir l’ambiance sonore cinématographique</i>	<i>20</i>
2.1.1 – Définition générale.....	20
2.1.2 – Différents usages	20
2.2 – <i>Production et diffusion d’une ambiance multicanale</i>	<i>23</i>
2.2.1 – Quelques repères historiques.....	23
2.2.2 – Inscription de l’ambiance dans le processus technique d’enregistrement d’une bande son	25
III – MULTICANAL	27
3.1 – <i>Diffusion multicanale : des standards et des nouveaux formats immersifs</i>	<i>27</i>
3.1.1 – Standardisation des formats 5.1 et 7.1.....	27
3.1.2 – Modèles actuels de diffusion immersive.....	29
3.1.3 – Deux études comparatives de formats de diffusion	35
3.2 – <i>Prise de son multicanale : développement et application professionnelle</i>	<i>39</i>
3.2.1 – Procédé de développement à partir d’angles utiles	40
3.2.2 – Exemples de dispositifs dédiés à l’enregistrement d’ambiances	45
3.2.3 – Deux études comparatives de dispositifs d’enregistrement	50

SECONDE PARTIE – RÉALISATION D'UNE EXPÉRIENCE : INFLUENCE D'UN FORMAT DE DIFFUSION SUR LA SENSATION D'IMMERSION.....	53
IV – PRÉSENTATION DE L'EXPÉRIENCE	53
4.1 – <i>Contexte d'étude.....</i>	53
4.2 – <i>Questions de recherche et critères d'évaluation.....</i>	54
4.3 – <i>Résumé de l'expérience</i>	55
V – ENREGISTREMENTS ET PRÉPARATION DES EXTRAITS.....	56
5.1 – <i>Conception d'un dispositif d'enregistrement sonore multicanal</i>	56
5.1.1 – <i>Directivité et influence des microphones d'élévation</i>	56
5.1.2 – <i>Réalisation du dispositif.....</i>	57
5.1.3 – <i>Équipements utilisés.....</i>	58
5.2 – <i>Création des stimuli et méthode de diffusion</i>	59
5.2.1 – <i>Sélection et enregistrement des ambiances</i>	59
5.2.2 – <i>Création des extraits.....</i>	59
5.2.3 – <i>Répartition des extraits par groupe de visionnage</i>	62
5.3 – <i>Caractéristiques techniques de l'environnement de l'expérience.....</i>	64
VI – PROTOCOLE DE TEST.....	66
6.1 – <i>Choix des participants et répartition par groupe.....</i>	66
6.2 – <i>Élaboration du questionnaire</i>	68
6.3 – <i>Présentation du protocole aux sujets.....</i>	71
VII – EXPLOITATION DES RÉSULTATS.....	72
7.1 – <i>Statistiques de l'échantillon.....</i>	72
7.2 – <i>Méthode d'analyse retenue.....</i>	73
7.3 – <i>Analyse par question</i>	75
7.3.1 – <i>Q1 : l'enveloppement (réponse par échelle de Likert).....</i>	75
7.3.2 – <i>Q2 : l'enveloppement (réponse par dessin).....</i>	76
7.3.3 – <i>Q3 : la localisation.....</i>	78
7.3.4 – <i>Q4 : la présence.....</i>	78
7.3.5 – <i>Synthèse de Q5, Q6 et Q7.....</i>	79
7.4 – <i>Analyse par question et par extrait</i>	81
7.4.1 – <i>Q4 et Q5 : la présence et la sensation de se sentir captivé</i>	82
7.4.2 – <i>Q6 : la cohérence image-son.....</i>	84
7.4.3 – <i>Q7 : l'imagination de l'espace</i>	85
7.5 – <i>Discussion autour de l'expérience</i>	86
CONCLUSION.....	88
BIBLIOGRAPHIE	91
ANNEXES.....	96

Introduction

La séance de cinéma est aujourd'hui mise en valeur par l'expérience sensationnelle qu'elle propose. Les tailles d'écran sont toujours plus importantes, la qualité des projecteurs permet la production d'images mieux définies et les systèmes de diffusion sonore ont vu leur nombre de canaux augmenter. Des technologies propriétaires, dites « premiums », sont aujourd'hui déployées en salles et leurs exploitants promettent aux spectateurs une expérience immersive inégalable. Cette quête de l'immersion, déjà amorcée par les travaux sur la réalité virtuelle et l'expérience du jeu vidéo, est devenue un argument de vente pour l'exploitation cinématographique.

Pour autant, il serait illusoire de penser ces artifices comme absolument nécessaires pour apprécier la projection, en l'absence desquels il serait impossible de vibrer au rythme du film. L'identification personnelle aux situations fictives, la compréhension des actions et des enjeux de narration, ainsi que les liens affectifs établis avec les personnages sont tout aussi importants.

Cette présente recherche a pour objectif d'interroger les rapports existant entre la sensation d'immersion cinématographique et l'utilisation de systèmes de diffusion sonore multicanaux. Son périmètre est défini par deux thèmes majeurs : l'identification et la mesure de critères d'évaluation de l'immersion, ainsi que le développement de formats de diffusion immersifs et de dispositifs d'enregistrement multicanaux adaptés.

L'étude sera focalisée sur un type de source rencontré dans la bande son d'un film : les sons d'ambiance. Celles-ci ont un statut particulier : elles élargissent la scène sonore au-delà du cadre de l'image et offrent aux spectateurs la possibilité de deviner l'environnement dans lequel les situations se déroulent et de s'imaginer le hors-champ. Cette fonction spécifique interroge le procédé de diffusion de ces ambiances, ce qui motive le sens de cette recherche.

Enveloppé physiquement par une ambiance sonore tout autour de lui, le spectateur pourrait-il être assisté dans son processus d'analyse et d'imagination de l'environnement présenté ?

Dès 1998, Witmer et Singer définissent le facteur de présence, concept attaché à l'immersion, comme la sensation de se sentir dans un espace lorsque l'on se situe physiquement dans un autre. Sans prise en compte du type de médium, la présence est une caractéristique perceptive pour l'évaluation d'un environnement virtuel. Plus récemment, dans le domaine du jeu vidéo, Bouchara et Okuya (2020) effectuent une revue de la littérature scientifique sur la

thématique de l'immersion, la présence et l'incarnation de soi. De la même manière, nous souhaitons dans la présente étude apporter un ensemble de précisions sur les spécificités de l'immersion au regard du médium cinéma.

Dans la thématique de la diffusion sonore multicanale, plusieurs études comparatives de systèmes s'appuient sur l'évaluation de critères perceptifs. Si certains peuvent être rapprochés des indices potentiels d'immersion, l'objectif principal de ces études ne s'inscrit pas dans ce champ d'action mais vise à caractériser objectivement les capacités des systèmes étudiés. Nous souhaitons, dans une démarche complémentaire, considérer d'autres variables que nous estimons nécessaires à une juste mesure de l'immersion cinématographique, comme l'influence de l'image et la présence ou l'absence de narration.

Ainsi, comment mesurer l'immersion ? Un format de diffusion sonore influence-t-il cette sensation ? La perception de l'espace présenté à l'image est-elle également impactée ?

La première partie de ce travail de recherche sera dédiée à l'examen des thèmes exposés, ainsi qu'à une analyse commune des liens qui les regroupent. Une mise en commun des travaux scientifiques et théoriques, relative aux propositions de définition de l'immersion et à l'identification de ses critères d'évaluation, sera d'abord proposée. Puis, l'étude approfondie des fonctions qu'occupe l'ambiance sonore cinématographique permettra d'interroger les méthodes de production au regard des formats de diffusion multicanaux dits « immersifs ». Ces derniers seront présentés et mis en comparaison dans un chapitre dédié, au même titre que les dispositifs d'enregistrement multicanaux.

En seconde partie, nous interrogerons les rapports existant entre l'immersion et les systèmes de diffusion en rapportant une expérience que nous avons réalisée. Celle-ci vise à mesurer l'impact de quatre formats de diffusion sonore (LCR, 5.0, 7.0 et 7.0.2) sur l'immersion d'un sujet sur la base de six critères d'évaluation. Cette expérience prend la forme de sessions d'écoute avec un recueil de données au moyen d'un questionnaire. L'élaboration de ce dernier ainsi que l'implémentation du test seront d'abord présentés, puis une analyse des résultats sera proposée.

Première partie – Prise de son et diffusion d'ambiances multicanales : identification des indices d'immersion

I – Immersion

Le terme « immersion » est polysémique. Dans le cadre de la production de médias, sa définition varie en fonction du domaine d'étude (musique, réalité virtuelle, films...) et des caractéristiques des supports de consommation.

La notion d'immersion croise par exemple celle de l'illusion sensorielle. Le plus souvent par la vue (utilisation de casques VR) et l'ouïe (popularisation du son 360°), les auteurs de contenus jouent avec les sens des utilisateurs, pour les plonger dans un nouvel environnement. Dans la production audiovisuelle et du jeu vidéo, l'immersion s'aborde également d'une manière différente. Engagés narrativement et émotionnellement, les spectateurs, ou les utilisateurs, sont comme transportés dans l'histoire.

Ce terme est récupéré de manière répétée par différentes sociétés de développement technologique, citons Dolby Laboratories et Auro Technologies pour la diffusion cinéma, qui en font dorénavant un argument commercial : « Il s'agit de la technologie la plus efficace pour obtenir un véritable son 3D, qui maintient le plus haut niveau de qualité audio et d'immersion souhaité par les créateurs. »¹ (Auro Technologies, 2015, p. 6, à propos de leur technologie Auro 3D)

Face à ces multiples acceptions, plusieurs groupes de recherche ont proposé une catégorisation de ces définitions. Agrawal *et al.* (2020) proposent dans le cadre de la production d'expériences audiovisuelles une séparation de l'idée d'immersion en deux thèmes majeurs : l'état psychologique et la propriété objective d'un système ou d'une technologie. Le concept d'état psychologique est lui-même subdivisé selon trois axes que sont l'immersion sensorielle, l'engagement du spectateur dans la narration et l'absorption face à un défi. Ce dernier thème

¹ « It is the most efficient technology for true 3D sound, maintaining the highest level of audio quality and immersion that was intended by the creators » (Auro Technologies, 2015, p. 6, notre traduction).

fait référence à l'usage de médias interactifs, notamment les jeux vidéos, dans lesquels le joueur, par ses choix, ses stratégies et ses calculs se retrouve impliqué et attentif à ses tâches. Ce dernier point est écarté par Williams *et al.* (2022) qui propose une catégorisation similaire de l'immersion dans le cadre la production cinématographique : l'expérience subjective (développée en « absorption narrative » et « immersion spatiale ») et la propriété d'un système.

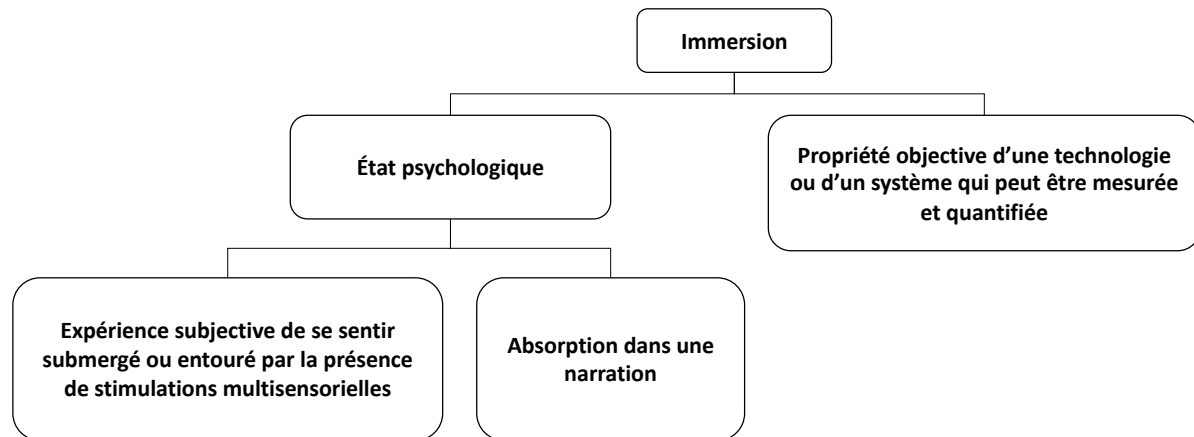


Figure 1 : Schéma récapitulatif de la classification de l'immersion selon Williams *et al.* (2022)

Ainsi, il convient de mettre en lumière les spécificités de l'immersion au regard du médium cinéma. Pour cela, nous utiliserons la catégorisation présentée par Williams *et al.* (Figure 1) pour ordonner nos propos.

1.1 – L'immersion comme expérience cognitive et sensorielle

1.1.1 – Le cinéma affectif : absorption et mise en phase

« L'immersion est un phénomène vécu par un individu lorsqu'il se trouve dans un état d'implication mentale profonde dans lequel ses processus cognitifs (avec ou sans stimulation sensorielle) provoquent un changement de son état attentionnel tel qu'il peut expérimenter une dissociation vis-à-vis de sa conscience du monde physique. »² (Agrawal *et al.*, 2020, p. 6)

² « Immersion is a phenomenon experienced by an individual when they are in a state of deep mental involvement in which their cognitive processes (with or without sensory stimulation) cause a shift in their attentional state such that one may experience disassociation from the awareness of the physical world » (Agrawal *et al.*, 2020, p. 6, notre traduction).

Cette définition générale proposée par Agrawal *et al.* suggère que l'immersion est avant tout un mécanisme cognitif. L'expérience dissociative du monde physique et virtuel, ainsi que la modification de l'attention en sont les éléments clés. Le terme absorption fait alors figure de synonyme dans la littérature scientifique (Bouchara & Okuya, 2020 ; Zhang *et al.*, 2017). Par analogie, au cinéma, le spectateur peut se retrouver absorbé par le film et engagé dans la narration.

L'immersion, en tant qu'état psychologique défini ci-dessus, intègre également la charge émotionnelle. Transporté dans l'histoire au visionnage d'un film, le spectateur peut alors se retrouver impliqué de manière affective. Il s'inscrit ainsi dans ce que Roger Odin appelle le processus de mise en phase : « Par mise en phase, j'entends le processus qui me conduit à vibrer au rythme de ce que le film me donne à voir et à entendre. » (Odin, 2000, p. 38)

Le spectateur établit un lien avec des personnages et des situations, il se retrouve projeté et engagé dans l'histoire. Parfois il s'attache, s'identifie ou reste neutre et à l'écart. Le ressenti de ces émotions reste une motivation majeure dans la volonté du public à se rendre en salle.

1.1.2 – L'immersion sensorielle

En se référant à l'origine latine du mot immersion, il se révèle une caractéristique liée cette fois-ci à nos sens. *Immergo - in mergo*, renvoie à l'acte concret d'immerger, de plonger dans un liquide, en mer ou, de manière imagée, dans un nouvel environnement. Nous nous rattachons à l'usage que nous faisons de nos sens pour appréhender celui-ci. Biocca et Delaney (1995) définissent notamment l'immersion de la manière suivante : « Le terme 'immersif' désigne le degré auquel un environnement virtuel submerge le système perceptif de l'utilisateur. »³ (p. 57)

Cette définition décrit la capacité d'un sujet à ressentir une expérience multisensorielle. Au visionnage d'un film, le spectateur sollicite sans cesse (et de manière majoritairement inconsciente) sa vue et son ouïe. Il accepte, par l'analyse des informations provenant de son

³ « Immersive is a term that refers to the degree to which a virtual environment submerges the perceptual system of the user » (Biocca & Delaney, p. 57, notre traduction).

système perceptif, la création d'un nouvel espace virtuel, celui du film, avec ses propres lois. Plus il se sent transporté physiquement dans cet environnement, plus l'immersion est élevée.

L'immersion sensorielle, ou spatiale, est à rapprocher de la notion de présence. Cette sensation peut se définir comme l'expérience subjective de se sentir dans un environnement lorsque l'on se situe physiquement dans un autre (Witmer & Singer, 1998). On la compare régulièrement à la « sensation d'être là » en référence à l'espace fictif qui nous est présenté (Bouchara & Okuya, 2020). Eaton et Lee (2019) qualifient d'immersion « passive » le facteur de présence, au sens où le spectateur n'a aucune interaction directe avec le film. Son attention est dirigée seulement par ses sens, à l'inverse du jeu vidéo où il se concentre également sur ses actions.

Nous nous écartons ainsi de la notion d'immersion en tant que processus psychologique d'engagement dans une narration où notre attention est portée sur l'histoire. Cependant, ces deux définitions restent corrélées. Agrawal *et al.* (2020) précisent dans leur étude que la sur-présentation des informations sensorielles visuelles et sonores en salle (écran large, niveau sonore élevé), ainsi que la mise à distance de l'environnement extérieur (sas d'entrée, isolation acoustique, lumières éteintes) peut contraindre le spectateur à se concentrer sur ce qui lui est présenté à l'image et au son. Cet état d'attention spécifique constitue un contexte favorable à un engagement psychologique dans le film et au développement de la mise en phase.

Plusieurs recherches ont eu pour objectif d'identifier les critères qui définissent l'immersion qu'un sujet peut ressentir à l'écoute d'une bande sonore. L'objectif de ces études est de proposer des critères de définitions pour la qualifier et la quantifier. Nous verrons que leurs domaines d'application peuvent être généraux ou s'appliquer à un médium spécifique.

1.2 – Critères de définition et de mesure

Nous aborderons dans cette partie plusieurs recherches qui ont traité de la capacité de l'être humain à analyser un environnement sonore, réel ou virtuel, dans lequel il se sent immergé. Dans un premier point, nous présenterons des critères de définition potentiels de l'immersion sensorielle, puis nous porterons notre regard sur les questionnaires de mesures existants prenant en compte l'implication du spectateur dans la narration.

1.2.1 – Propositions de critères perceptifs

Différents adjectifs de la littérature scientifique anglophone viennent qualifier l'immersion sensorielle. Nous retrouvons régulièrement les termes *perceptual immersion*, *sensory immersion*, *spatial immersion* ou encore *quality of spatial audio* (Williams *et al.*, 2022).

Lindau (2015) propose d'établir un glossaire spécifique pour l'évaluation et la comparaison d'environnements sonores virtuels à travers un ensemble de critères. Avec son équipe, il en classe quarante-sept, accompagnés de leur description, en huit catégories :

- **Timbre:** *tone color bright-dark; high-frequency tone color; mid-frequency tone color; low-frequency tone color; sharpness; roughness; comb filter coloration; metallic tone color*
- **Tonalness:** *tonalness; pitch; doppler effect*
- **Geometry:** *horizontal direction; vertical direction; front-back position; distance; depth; width; height; externalization; localizability; spatial disintegration*
- **Room:** *level of reverberation; duration of reverberation; envelopment (by reverberation)*
- **Time behavior:** *pre-echoes; post-echoes; temporal disintegration; crispness; speed; sequence of events; responsiveness*
- **Dynamics:** *loudness; dynamic range; dynamic compression effects*
- **Artefacts:** *pitched artifact; impulsive artifact; noise-like artifact; alien source; ghost source; distortion; tactile vibration*
- **General:** *clarity; speech intelligibility; naturalness; presence; degree-of-Liking; other*

Certains critères de ce glossaire ne sont en lien direct ni avec la notion d'immersion ni avec le domaine filmique de manière générale. Cela n'est pas l'objet de la recherche. Cependant, nous retrouvons déjà le critère *presence* lié à la sensation subjective de se sentir transporté dans un environnement virtuel. Nous observons également que les critères regroupés dans les catégories *geometry* et *room* (*horizontal direction, depth, envelopment*, etc.) permettent de situer une source dans l'espace autour de soi. Les auteurs de l'étude estiment que ces critères peuvent être quantifiables avec une échelle de notation (*envelopment* : *less pronounced – more pronounced* ou *presence* : *lower – higher* par exemple).

En se concentrant maintenant sur le médium cinéma, dans une étude de 2019, Eaton et Lee identifient dix critères perceptifs pour les domaines du jeu vidéo, de la musique, du film et du paysage sonore. Les chercheurs ont ensuite proposé à des sujets d'évaluer ces critères en fonction de leur degré de liaison avec la sensation d'immersion. Les critères principaux selon les résultats correspondant au domaine filmique ont été l'enveloppement (la sensation subjective d'être entouré de son) procuré par les sources sonores et leur réverbération, ainsi que la localisation (direction de la source sonore sur le plan horizontal).

L'ensemble des critères présentés précédemment sont associés à une immersion sensorielle : le spectateur cherche à analyser d'un point de vue spatial l'environnement qui l'entoure (localisation, enveloppement, etc.). Cependant, nous avons évoqué le fait que cette immersion spatiale, bloquant les stimuli extérieurs et forçant le spectateur à se concentrer sur ce qu'il voit et écoute, engendre également un mécanisme cognitif d'engagement dans la narration. C'est une mise en phase qui s'installe sur la durée et selon la nature du stimulus présenté. C'est pourquoi il est pertinent de recueillir le ressenti des spectateurs à travers d'autres systèmes de mesures que des critères perceptifs, comme par exemple des questionnaires de présence.

1.2.2 – Questionnaires de présence et d'immersion

Bouchara et Okuya (2020) expliquent qu'il est possible de recueillir des données relatives à l'immersion des sujets sous différentes formes :

- Des mesures objectives physiques (« suivi oculaire ... fréquence cardiaque », p. 24) ou comportementales (« temps de réaction aux tâches secondaires ... mesure attentionnelle », p. 25)
- Des mesures subjectives par des questionnaires de présence dans des environnements virtuels qui « incluent très généralement des questions plus spécifiques à l'immersion » (p. 24) tel que le *Igroup Presence Questionnaire* (IPQ) ou l'*ITC-Sense of Presence Inventory* (L'ITC-SoPI).

Parmi les questionnaires existant, *The Film Immersive Experience Questionnaire for film and TV* (*the Film IEQ*) a été proposé par Rigby *et al.* en 2019 pour mesurer l'immersion des spectateurs au cinéma. Il a été construit à partir d'un questionnaire préexistant mesurant

l'expérience subjective d'immersion de joueurs de jeux vidéos (Jennett *et al.*, 2008). Les chercheurs ont considéré que le fondement théorique utilisé pour ce questionnaire pouvait s'appliquer à d'autres environnements, notamment aux visionnages de contenus audiovisuels. Ainsi, ils ont adapté les questions originales et remplacé certaines pour que le questionnaire puisse être fonctionnel dans ce nouveau champ d'application. Celui-ci se compose de trente-et-une questions dont les réponses sont mesurées sous la forme d'échelle de Likert (1-7). Ce questionnaire est disponible en Annexe A.

Après son développement, les chercheurs ont réalisé une analyse factorielle exploratoire pour regrouper les questions en plusieurs thèmes et comprendre de quelle manière le concept d'immersion pouvait être mesuré. Celle-ci fût conduite sur quatre-cent-quatorze participants.

Quatre thèmes furent retenus de cette analyse et définis comme suit :

- « (1) La capacité à se sentir captivé. Douze questions (Qs 1-4, 13, 17-19, 21-24) concernant le plaisir du spectateur, son intérêt et sa motivation à regarder le programme.
- (2) La dissociation du monde réel. Trois questions (Qs 5-7) concernant le degré de conscience du spectateur par rapport au monde réel qui l'entoure.
- (3) La compréhension. Quatre questions (Qs 12, 14-16) concernant la compréhension des concepts et des thèmes du programme.
- (4) La présence (« *transportation* »). Cinq questions (Qs 8-11, 20) décrivant à quel point l'utilisateur a eu l'impression de vivre les événements pour eux-mêmes et à quel point il a eu l'impression de se trouver dans le monde présenté dans le programme. »⁴ (p. 7)

Ces chercheurs ont observé que l'immersion pouvait être mesurée selon quatre thématiques que sont l'intérêt et la motivation du spectateur pour le document audiovisuel qui lui est présenté, sa compréhension, sa conscience du monde réel durant la projection, et son

⁴ « (1) Captivation. Twelve items (Qs 1–4, 13, 17–19, 21–24) regarding the viewer's enjoyment, how interested they were, and their motivation to watch. (2) Real-world dissociation. Three items (Qs 5–7) regarding how much the viewer was aware of their real-world surroundings. (3) Comprehension. Four items (Qs 12, 14–16) asking about how well the concepts and themes of the video were understood. (4) Transportation. Five items (Qs 8–11, 20) describing how much the user felt like they were experiencing events for themselves, and how much they felt they were located in the world portrayed in the video » (Rigby *et al.*, 2019, p. 7, notre traduction).

sentiment d'avoir été présent dans le monde virtuel. Les auteurs précisent que le terme *transportation* dans leur étude est synonyme de présence (telle que définie en page 12).

Ce questionnaire a l'avantage de ne pas se baser sur des critères uniquement perceptifs. Il prend en compte l'analyse subjective que les spectateurs en font, leur attention et la dimension cognitive apportée par la narration.

Notre capacité à analyser ces thèmes et ces critères dépend étroitement de la qualité technique des dispositifs de projection sonore et visuelle mis en jeu. En salle de cinéma, cette immersion sensorielle est encouragée par un confort visuel (écran large, lumières éteintes) et sonore (niveau et système d'écoute adéquat, isolation acoustique empêchant la survenue de bruits extérieurs) du spectateur. Par exemple, Lombard *et al.* (2000) démontrent dans une étude que lors de projections de documents audiovisuels, la taille de l'écran influe sur la sensation de présence des sujets. Ainsi, l'immersion peut aussi être considérée comme une propriété inhérente au dispositif technique de projection.

1.3 – Les systèmes immersifs : une caractéristique objective

En s'écartant des facultés cognitives et sensorielles d'un individu, le potentiel immersif d'un dispositif peut enfin décrire sa capacité à favoriser l'immersion, notamment par « le nombre et le réalisme des canaux sensoriels stimulés » (Bouchara & Okuya, 2020, p. 26). De cette manière, Mel Slater, chercheur en science informatique et en environnement virtuel, met en avant la différence entre immersion et présence : « Réservez le terme 'immersion' à ce que la technologie apporte d'un point de vue objectif. Plus un système de diffusion propose des dispositifs (dans toutes les modalités sensorielles) et un suivi qui préserve sa fidélité par rapport aux modalités sensorielles équivalentes du monde réel, plus il est 'immersif'. »⁵ (2003, p. 1) Selon cette proposition, il serait possible de mesurer le degré de sophistication d'un système afin d'estimer son potentiel immersif.

⁵ « Let's reserve the term 'immersion' to stand simply for what the technology delivers from an objective point of view. The more that a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more that it is 'immersive' » (Slater, 2003, p. 1, notre traduction).

Pour l'exploitation de salles de cinéma, ce potentiel immersif renvoie alors aux procédés techniques et technologiques mis en œuvre et censés favoriser l'immersion du spectateur.

Une multitude de technologies a pu voir le jour et l'immersion est devenue un argument commercial. Selon le bilan 2021 du CNC, quatre-vingt-quinze établissements disposent d'au moins une salle premium, soit 4,7 % des cinémas actifs en France. Parmi les technologies décrites comme premium : l'IMAX (écran de 22 m × 16 m), considérée comme étant une « projection immersive » (p. 64), et ScreenX (projection à 270°) permettent de visionner des films sur des écrans toujours plus imposants. Concernant les systèmes de diffusions sonores, nous observons une multiplication de canaux audios et d'enceintes avec le développement de systèmes propriétaires tel que Dolby Atmos ou DTS:X.

En France, le label « immersion » est proposé par la Commission Supérieure Technique de l'image et du son aux salles de cinéma répondant à un cahier des charges précis. Il garantit une « qualité d'expérience ... favorisant l'immersion » (CST, s. d.) en se basant sur des critères tels que les caractéristiques de l'écran et du projecteur, le confort visuel et acoustique, la présence d'un système sonore immersif (selon la définition proposée ci-après) ou encore l'accessibilité.

En s'intéressant à la production sonore pour le cinéma plus spécifiquement, l'expression « système immersif » est utilisée de manière abstraite pour présenter à la fois un dispositif de prise de son, de diffusion, une technologie particulière ou un format d'écoute.

Les membres de la Society of Motion Picture and Television Engineers (SMPTE), organisation établissant des standards internationaux de diffusion pour la télévision et le cinéma numérique, se sont intéressés à ce sujet et ont posé une définition de l'expression : « Système immersif : un système de diffusion sonore capable de produire un son immersif. Son immersif : comprend le son qui émane des enceintes situées au niveau du *base layer* et au-delà. »⁶ (SMPTE, 2018)

Ici, *base layer* correspond à la disposition usuelle des groupes d'enceintes au-dessus des spectateurs en salle de cinéma, retrouvée notamment en configuration 5.1 et 7.1 (cf. 3.1.1 – *Standardisation des formats 5.1 et 7.1*).

⁶ « Immersive Sound System: A sound system capable of producing immersive sound. Immersive Sound: Includes sound that emanates from sources at and beyond the base layer » (SMPTE, 2018, noter traduction).

Cette disposition est à rapprocher de *height Layer* et *top Layer* qui définissent respectivement les groupes d'enceintes disposés encore plus en hauteur sur les murs au-dessus des spectateurs ainsi qu'au plafond de la salle (cf. 3.1.2 – *Modèles actuels de diffusion immersive*).

Pour conclure, l'immersion peut désigner, selon notre classification, l'implication du spectateur dans une histoire ou une narration, une illusion sensorielle – appelée aussi immersion spatiale - et désigne enfin les propriétés physiques d'un système de diffusion. Ces catégories ne sont pas adoptées collectivement. Des nuances, apportées par les auteurs et synthétisées en partie ci-dessus, empêchent l'élaboration d'une définition claire et partagée de ce terme et de ces thèmes associés.

Pourtant, on peut considérer que ces derniers restent interdépendants. Si Agrawal *et al.* (2020) n'adhèrent pas à l'idée que l'immersion puisse être une caractéristique objective d'un système, ils rappellent cependant que l'ajout de canaux audios augmente les informations sensorielles et participe à l'immersion spatiale du spectateur. Cette même immersion spatiale peut amener à une attention accrue de sa part envers le récit. En considérant ce nouvel environnement comme référence sonore et visuelle, il se concentre sur la narration et apparaît alors le phénomène de mise en phase.

1.4 – Immersion sonore et nature des sources

Si dans cette première partie nous avons évoqué les multiples critères qui peuvent caractériser l'immersion, il est maintenant nécessaire de savoir ce qui la motive. Au cinéma, le spectateur est sans cesse guidé par des stimuli sonores et visuels. C'est bien leur nature, leur signification et leur agencement qui permettent de créer une unité filmique dans laquelle le spectateur pourra potentiellement être absorbé.

A propos de la bande son d'un film : quels éléments la constituent permettent la compréhension de l'histoire ? Existe-il une hiérarchisation des sources ? Quels éléments de cette bande, si tant est qu'il en existe, favorisent l'immersion ?

Pour tenter de répondre à ces questions, il faut d'abord examiner la composition de cette bande sonore. Ses éléments essentiels sont communément regroupés en trois catégories : les paroles, les musiques et les effets. Cette dernière catégorie, aussi nommée « bruit » (Chion,

2017, p. 150), regroupe l'ensemble des bruitages réalisés pour le film, les effets spéciaux et les ambiances ajoutées.

Il est admis que la musique, selon son genre, son rythme et ses instruments, procure des émotions au spectateurs. Les dialogues font avancer l'histoire et les personnages sont identifiés par leur physique et leur voix. Par ailleurs, il existe une tendance naturelle à donner aux effets sonores une fonction de soutien ; leur utilisation consiste justement à orienter la sélectivité du spectateur vers les sons strictement nécessaires à la narration.

L'ambiance sonore, plus spécifiquement, invite aussi le spectateur à imaginer un environnement autour de l'action. Elle ouvre un hors-champ au-delà du décor visible et, par un jeu d'identifications et d'associations, le spectateur devine ou imagine les composantes de ce nouvel espace.

Relativement à une immersion sensorielle, le statut particulier de l'ambiance pose alors question : une ambiance multicanale diffusée autour du spectateur permet-elle à ce dernier de mieux se figurer l'espace qu'il voit ? Pourrait-il se sentir plus présent dans l'environnement qui lui est donné à entendre ? Dans la suite de cette étude, nous nous concentrerons plus spécifiquement sur la nature et les usages des ambiances au sein de la bande sonore comme acteurs d'une potentielle immersion.

II – Ambiances

2.1 – Définir l'ambiance sonore cinématographique

2.1.1 – Définition générale

« Dans la production de médias, une ambiance sonore pourrait être décrite comme l'ensemble des sons qui constituent un arrière-plan sonore, presque un second plan, et construisent un environnement sur lequel peuvent s'appuyer les autres sons, ceux destinés principalement à la progression du récit. Les ambiances caractéristiques sont celles des sons de la nature (le vent, la mer, la forêt, la pluie, etc.) et celles des activités humaines (foule, circulation, site industriel, parc, école, plage, gare...). » (Adjiman & Denizart, 2015, p. 3)

Dans cette définition, Adjiman et Denizart exposent le concept d'une division de la bande sonore en plusieurs plans et dont l'ambiance est un soutien fondamental. Cette dernière définit un « environnement », qu'il soit physique (illusion réaliste du lieu présenté à l'image) ou narratif (dichotomie audiovisuelle pour un effet de mise en scène). L'ambiance est perçue de manière plus ou moins consciente selon les plans sonores qui l'accompagnent et l'engagement du spectateur dans la narration.

Elle « n'est pas nécessairement synchrone avec l'action qui se déroule à l'écran » (de Nanteuil, 2008, p. 24). Sans relation temporelle avec le cadre de l'image, l'ambiance sonore invite l'imagination du spectateur à ouvrir l'espace et à se projeter dans un hors-champ sonore. Ce dernier ne se construit pas instantanément, il évolue parallèlement à l'écoute de l'ambiance. Dans ce contexte, le spectateur pourra ressentir l'effet de présence et suivre les situations présentées à l'écran avec une attention accrue.

2.1.2 – Différents usages

L'ambiance a pour vocation la mise en place d'une unité sonore. Celle-ci s'inscrit sous plusieurs registres qui définissent ensemble ses fonctions et ses usages. Elle peut être fonctionnelle, indice de lieu ou narrative (Adjiman, 2018).

L'ambiance fonctionnelle

L'ambiance a avant tout une fonction pratique. De par sa longueur temporelle, elle fait le lien entre les plans et gomme les ellipses nécessaires à la dynamique du film. En escamotant les coupes produites par l'étape du montage, elle permet au champ attentionnel du spectateur de ne plus être fixé sur les sautes de plans mais sur la continuité narrative de la séquence.

L'ambiance fonctionnelle oriente l'ouïe et l'engagement du spectateur vers les sons nécessaires à la progression du récit. Dans des contextes particuliers, comme par exemple une séquence se déroulant dans un appartement complètement isolé des événements extérieurs, cette ambiance peut être réduite à la notion de fond d'air. Celui-ci, par essence enregistré lors d'un silence général sur le lieu du tournage, se caractérise par l'absence quasi-totale de source ponctuelle et par conséquent son extrême monotonie. Ces fonds d'air ont « une vocation technique ... [et] n'apportent rien narrativement ou émotionnellement » (Valérie Deloof citée dans Adjiman, 2018).

L'ambiance territoire

« Territoire » est un qualificatif introduit par Michel Chion dans son expression « son-territoire », synonyme de son ambiant, comme servant « à marquer un lieu, un espace particulier de leur présence continue et partout épandue » (Chion, 2017, p. 84). En effet, les sons d'ambiance, de par la nature des sources qu'ils contiennent, peuvent représenter l'empreinte sonore d'un environnement (vent sifflant dans une forêt primaire, brouhaha d'un marché animé, sifflet de train sur un quai de gare, etc.). Ces marqueurs sonores créent une cohérence, plus ou moins réaliste, sur laquelle le spectateur va se reposer pour identifier le caractère d'un lieu. Le spectateur ne cherche pas à localiser ou visualiser ces sources. Au lieu de cela, son imagination vient, sans effet de distraction, compléter le hors cadre image selon les indices qu'il entend. Il peut ainsi exprimer un sentiment de présence en acceptant de se sentir dans le lieu sonore qui lui est proposé.

Notons que nous utilisons le pluriel pour présenter ces ambiances car il est souvent complexe, voire impossible, d'intégrer à la bande sonore une seule ambiance réunissant l'ensemble des indices spécifiques au lieu fictif traité. En résulte un travail d'intégration en plusieurs couches, nommées *stems*, où ces indices sonores sont ajoutés les uns après les autres.

Ce travail est effectué à l'étape du montage son. Ces sources peuvent être ponctuelles (un chien aboie à l'extérieur d'une maison, une moto passe au loin, etc.) comme continues (léger bruissement des feuilles dans les arbres, rumeur de ville en après-midi, etc.).

L'ambiance narrative

Selon Adjiman (2018), une dernière fonction qu'occupent les ambiances au sein de la bande son d'un film est leur vocation à servir le récit. Denizart indique, pour qualifier leurs rôles, qu'elles : « incarnent une appréciation sensible, esthétique et différée ... de la dimension sonore de l'univers filmique, associée à une écoute affective, émotive et contemplative » (2017, p. 40).

En effet, au-delà des caractéristiques descriptives, il émerge entre l'ambiance et l'image une relation complexe qui tend à produire une infinité de significations possibles selon les intentions de mise en scène du réalisateur. Chaque source sonore possède une palette lexicale de sensations (Adjiman, 2015) et peut être appliquée, selon le sens que l'on souhaite lui donner, à des images différentes.

En prenant l'exemple de cris d'oiseaux, Adjiman (2018) pose les questions nécessaires à l'élaboration d'un parti pris sonore pour accompagner la narration : « Les cris des oiseaux sont-ils mélodieux ou nasillards, harmonieux et musicaux ou stridents et agressifs ? Sont-ils discrets ou piaillants ? Sont-ils dispersés ou menaçants par leur nombre et leur densité ? Rappellent-ils des rossignols ou des corbeaux ? ... Le vent est-il paisible et porteur d'une douce quiétude en faisant doucement osciller les feuilles frêles, ou sifflant, glacial et agressif ? » (p. 147)

Pour des effets de dissonance audiovisuelle, on s'aperçoit que la cohérence entre son (l'ambiance) et image (le lieu) n'est plus nécessaire. Prenons l'exemple d'une ambiance sonore d'aire de jeux avec des enfants chantant des comptines. Si nous l'associons à l'image sombre d'une vieille porte en bois entre-ouverte, cela donnerait, assez efficacement, un ton mystérieux, voire angoissant à la séquence.

Enfin, les ambiances peuvent être considérées comme une transcription sonore des émotions du personnage. Il peut être instauré une « relation de mimétisme entre leur état psychologique (les relations diégétiques) et l'environnement sonore qui est donné à entendre au spectateur (les relations filmiques) » (Denizart, 2017, p. 242). Prenons l'exemple d'un

personnage isolé se baladant en forêt qui, à la nuit tombée, se met à paniquer quand il se rend compte qu'il s'est perdu. La forêt de jour lui apparaît comme familière, ressourçante ; le monteur son pourrait choisir de la matérialiser par un vent doux à travers le bruissement des feuilles et des oiseaux chantants. La nuit, cette même forêt est perçue d'une autre manière. Elle peut être oppressante ou effrayante. Un vent sifflant et des craquements de branches intempestifs pourraient alors être utilisés pour signifier l'angoisse du personnage.

Les ambiances narratives permettent un processus d'attachement, parfois d'identification, du spectateur envers le personnage par un engagement émotionnel et participent ainsi à l'immersion narrative.

2.2 – Production et diffusion d'une ambiance multicanale

Nous avons défini plusieurs usages de l'ambiance sonore selon la nature des sources qu'elle contient, sa relation avec l'image et sa capacité narrative. C'est bien une ambiance cognitive qu'analyse le spectateur lorsqu'il est absorbé par le film.

Nous n'avons pas pris en compte son mode de diffusion. Celui-ci a-t-il une importance ? Une diffusion multicanale de l'ambiance peut-elle servir une ou plusieurs de ses fonctions citées précédemment ?

Prenons l'exemple d'un usage fonctionnel de l'ambiance, qui consisterait à créer un simple liant dans le cas d'une séquence dialoguée. Les voix et les présences des acteurs étant restituées à travers le canal central, la diffusion similaire d'une ambiance monophonique qui les accompagne serait cohérente. Pour caractériser un lieu cette fois-ci, nous pouvons questionner la pertinence de tendre continuellement vers une diffusion multicanale de l'ambiance.

Si aujourd'hui cette question est d'actualité et que le cinéma multipiste est désormais omniprésent, l'usage actuel des ambiances au cinéma reste lié aux évolutions techniques et technologiques concernant les supports d'enregistrement et de diffusion.

2.2.1 – Quelques repères historiques

L'usage de la monophonie sur son optique comme référence dans les premières décennies du cinéma sonore limitait le mixage d'un nombre conséquent de sources de par le masquage fréquentiel important imposé par une faible bande passante (Chailloleau, 2016). Afin

d'éviter le désordre d'un amas sonore et de permettre aux spectateurs d'apprécier leur séance sans gêne, les éléments strictement nécessaires à la narration – notamment les musiques puis les dialogues à l'ère du cinéma parlant - étaient priorisés.

Plusieurs évolutions successives ont permis de faire entendre dans la bande de nouvelles couches d'éléments sonores, en laissant apparaître les bruitages et des ambiances plus définies. Parmi ces évolutions, nous pouvons citer l'accroissement progressif de la bande passante et de la dynamique du son optique, puis le passage vers la bande magnétique ainsi que les innovations de la société Dolby en termes de réduction de bruit (Chion, 2017).

Cette dynamique plus élevée et les évolutions technologiques des systèmes de diffusion, « autorisent de travailler sur un grand nombre de niveaux et différentes couches de son qui viennent qualifier les plans sonores et les différents cercles autour de l'écran, et ce avec beaucoup de subtilité » (Adjiman, 2015, p. 1).

L'avènement de la diffusion sonore multipiste, initiée dès les années 30 par le procédé Fantasound puis développée lors du passage au son magnétique sur pellicule, permet de minimiser la contrainte de hiérarchisation des sources lors de mixages monophoniques. En espaçant les sources entre elles à travers différents haut-parleurs, il est dorénavant possible de créer un « démasquage spatial » (Bailblé, 2004). En conséquence, des ambiances, autrefois ôtées de la bande son par souci de synthèse, purent émerger en partie par leur diffusion sur de nouveaux canaux de diffusions dédiés : les canaux *surround*.

Ces canaux ouvrent physiquement le champ sonore autour des spectateurs et permettent d'intégrer l'ambiance ainsi que d'autres éléments de la bande sonore, par spatialisation, dans ce que Michel Chion appelle le super-champ : « Nous appelons super-champ le champ dessiné, dans le cinéma multipistes, par les sons ambiants de nature, de rumeurs urbaines, de musique, de bruissements, etc., qui environnent l'espace visuel et sonore et peuvent provenir de haut-parleurs situés hors des limites strictes de l'écran, récemment même depuis le plafond. » (Chion, 2017, p. 155)

Ces phénomènes de spatialisation relèvent de choix esthétiques et d'intentions de mise en scène mis en œuvre à l'étape du mixage. Alors, la nature des sources, mais aussi leur mode de captation, influe sur leur capacité à s'intégrer dans la bande. Certaines ambiances sont parfois déjà pensées et définies en amont comme pouvant être enregistrées par des systèmes de prise de son multi-microphoniques. Mais sur un tournage, il est parfois difficile de trouver un temps

dédié pour leur captation. Ainsi, il convient de rappeler quelques enjeux fondamentaux dans le cas de l'enregistrement d'ambiances sonores à destination de la production cinématographique.

2.2.2 – Inscription de l'ambiance dans le processus technique d'enregistrement d'une bande son

Lors du tournage d'un film, le chef opérateur du son et son équipe ont pour fonction principale de recueillir les sons directs de la séquence, le plus souvent les répliques des acteurs. Il leur est aussi nécessaire de réaliser des silences raccords (enregistrements silencieux du plateau) nécessaires au montage des dialogues pour que les spectateurs aient une impression de fluidité à l'écoute de ces derniers. Ces silences forment, une fois montés, une ambiance essentielle pour donner une unité temporelle à la séquence (c'est l'usage fonctionnel).

Il arrive également que l'équipe son du film puisse enregistrer des ambiances pour caractériser des espaces directement pendant le tournage (rumeur de foule avec les figurants sur place, ambiance du trafic lors d'un tournage en ville, etc.). L'appellation « sons seuls » est alors employée pour identifier ces sons.

L'enregistrement des ambiances sur un plateau demande de l'organisation et une certaine rigueur. Le plan de travail organisé par l'équipe de mise en scène limite ces enregistrements à des temps parfois trop serrés. De plus, cela demande un silence absolu sur le plateau, condition parfois difficile à obtenir.

Ainsi, il est par moment préférable pour l'équipe son de capturer des ambiances lorsqu'elle se trouve seule sur le lieu d'enregistrement.

Les enregistrements d'ambiances peuvent aussi s'effectuer en totale séparation des tournages et des contextes de production. Ces enregistrements de terrain, ou *field recordings*, peuvent être regroupés selon différentes classifications, souvent par thème (paysages urbains, vents et forêts, foules, etc.) et par format. Ces enregistrements sont commercialisés sous la forme de bibliothèques sonores. Citons par exemple Boom Library, Sound Ideas et Pro Sound Effects comme trois marques de référence dans la distribution de bibliothèques pour la post-production sonore cinématographique.

Pour conclure, les ambiances occupent des rôles multiples au sein de la bande sonore. Elles soutiennent le montage et ses enjeux narratifs. Elles aident le spectateur à se focaliser sur des séquences sans discontinuité et favorisent tant que possible son engagement dans l'histoire. Elles marquent enfin - c'est leur utilisation territoire - un lieu fictif d'une empreinte sonore dans laquelle il peut se retrouver transporté par son imagination.

Cette dernière fonction rappelle dans un registre plus sensoriel la sensation d'enveloppement. Entouré physiquement par une ambiance sonore omniprésente, le spectateur pourrait-il être aidé à se sentir plus présent dans l'espace projeté ? Nous avons proposé une réponse à cette question dans la seconde partie de cette recherche qui présente une expérience que nous avons menée en comparant plusieurs formats de diffusion sonore.

Des systèmes aujourd'hui dits « immersifs » incluent des enceintes autour mais également au-dessus et au-dessous des spectateurs. Si l'ambiance possède un statut particulier dans la bande son, celui d'ouvrir le champ à ce que l'on ne voit pas, la présence d'enceintes au-delà de l'écran pourrait s'avérer être un appui matériel pour diffuser ces sons.

Dans ce prochain chapitre, nous nous intéresserons aux possibilités de prise de son et de diffusion multicanale relativement à notre objet d'étude. Nous aborderons les spécificités de l'enregistrement et de la restitution d'ambiances dans un cadre pratique, puis nous présenterons les liens et les limites existantes avec la notion d'immersion.

III – Multicanal

« Il y a aussi cette vieille idée - négation de l'imaginaire en même temps qu'affirmation triomphale du perceptif - selon laquelle le stade ultime du cinéma serait de composer une réalité-bis, soit un dispositif en lequel champ visuel et champ auditif seraient totalement reconstitués. » (Bailblé, 2004, pp. 56-57)

3.1 – Diffusion multicanale : des standards et des nouveaux formats immersifs

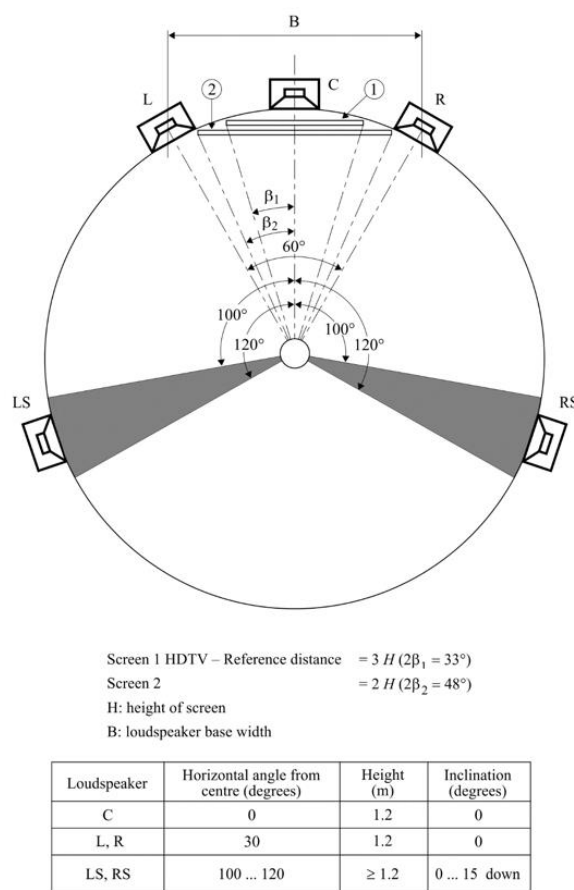
L'approche multicanale dans la diffusion d'une bande sonore filmique, d'abord limitée par les dispositifs techniques mis en place jusque dans les années 40, puis déployée par la popularisation du son magnétique sur film, fait aujourd'hui l'objet de standardisations par la publication de normes et de recommandations.

Un nombre important de documents écrits aborde les principales évolutions techniques des dispositifs de prise de son et de diffusion sonore au cinéma. Nous invitons par exemple le lecteur à se diriger vers l'ouvrage de Lerouge (1996) pour découvrir ces procédés, des premiers essais de la fin du XIX^{ème} siècle jusqu'à l'avènement du cinéma numérique. Nous allons ici nous attarder sur les principaux procédés et technologies exploités aujourd'hui.

3.1.1 – Standardisation des formats 5.1 et 7.1

L'Union Internationale des Télécommunications (2022) encadre actuellement le format référence de la diffusion multipiste 5.1. Dans sa recommandation *ITU-R BS.775-4 Multichannel stereophonic sound system with and without accompanying picture*, l'organisme présente les caractéristiques techniques de ce système à cinq canaux autour du spectateur avec un canal optionnel dédié aux basses fréquences. Celles-ci sont résumées en Figure 2.

Cette recommandation précise que les canaux surround Ls et Rs peuvent être divisés en plusieurs enceintes à disposer à intervalles réguliers dans un angle allant de $\pm 60^\circ$ à $\pm 150^\circ$ par rapport à l'axe frontal.



BS.0775-01

Figure 2 : Schéma du positionnement des enceintes pour le format de diffusion 5.1 (ITU, 2022)

Le champ d'application de cette recommandation n'a pas de limite précise. Si elle tire son origine d'une standardisation liée à l'arrivée de la télévision numérique par satellite, plusieurs albums de musique sont aujourd'hui mixés en respectant cette configuration. De la même manière, la majorité des studios de montage son pour le cinéma l'a également adoptée.

Cependant, il est important de notifier que les premiers essais de diffusion sonore incluant un canal *surround*, ou d'ambiance, au cinéma eurent lieu bien avant une quelconque standardisation. Dès les années 40, avec les expérimentations du Fantasound, puis, plus tard, par le CinemaScope et le format Todd-AO, l'utilisation de ces canaux s'est généralisée. Le Dolby Stéréo 70mm et le Dolby Digital sont par exemple deux procédés utilisant le format 5.1 tel qu'on le présente aujourd'hui, en séparant les canaux *surround* gauche et droit.

En 2010, pour la sortie du film *Toy Story 3* réalisé par Lee Unkrich, la société Dolby Laboratories dévoile son nouveau système de diffusion cinéma : le Dolby Surround 7.1. Ce système ajoute deux canaux supplémentaires au format 5.1. Chaque canal *surround* est divisé

entre le côté et l'arrière de la salle. Cette configuration est présentée en Figure 3. À ces canaux matérialisés par une ou plusieurs enceintes, appelées lorsqu'elles sont en groupe *array*, viennent s'ajouter deux canaux HI (*Hearing-Impaired*) et VI-N (*Visually Impaired-Narrative*) qui correspondent respectivement aux pistes mixées pour les personnes malentendantes et malvoyantes (Dolby Laboratories, 2011).

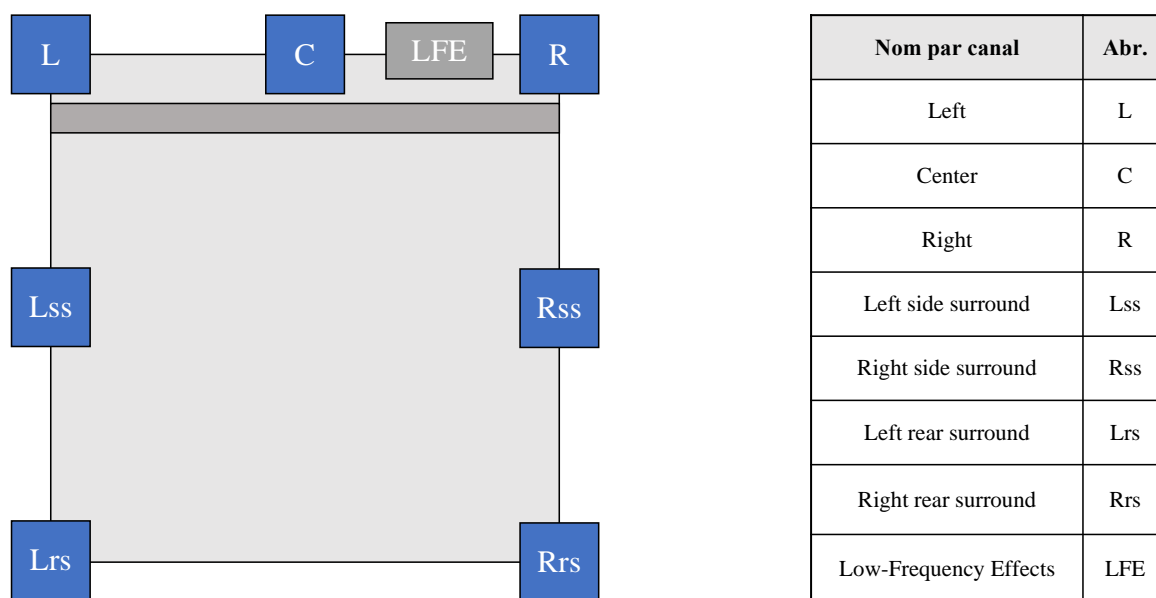


Figure 3 : Répartition des canaux audios pour le format de diffusion Dolby Surround 7.1 (Dolby Laboratories, 2011)

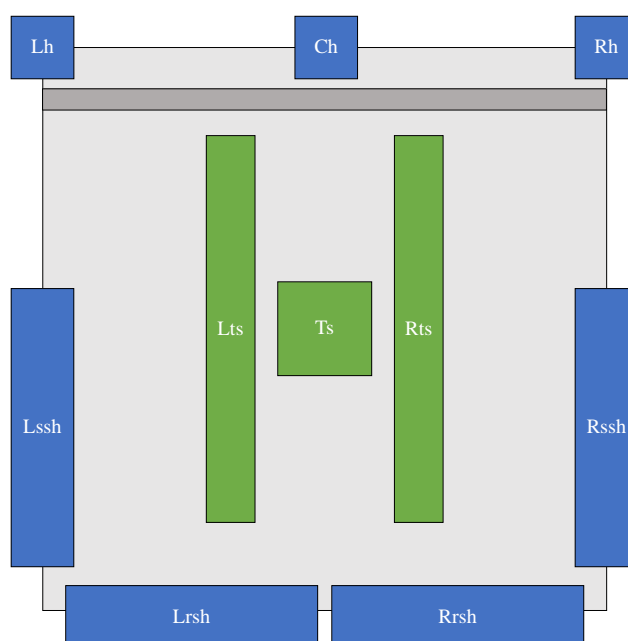
3.1.2 – Modèles actuels de diffusion immersive

Encore aujourd'hui, les formats 5.1 et 7.1 restent un standard de diffusion. Dans les années 2010, de nouveaux formats propriétaires dits « immersifs » et contenant un nombre encore plus élevé de canaux ont vu le jour, promettant une approche plus localisée de la spatialisation et une sensation accrue de réalisme. Des enceintes sont ainsi parfois positionnées encore plus haut sur les murs de la salle, au plafond, et parfois même au-dessous des spectateurs. Cyril Holtz, mixeur son français, dit au sujet des ambiances mixées dans ce type de format : « Ce qui m'a frappé dans mes premières utilisations du système, [c'est que cela permet] des bandes sons plus aérées, moins denses, moins chargées. Ça permet de disposer d'une répartition plus agréable. » (Lambinet, 2022, p. 136)

Ces formats immersifs intègrent plusieurs couches d'enceintes, appelées *layers*, positionnées à différentes hauteurs. La SMPTE (2018) qui, nous l'avons exposé dans le premier

chapitre, a proposé de regrouper ces différents formats en une définition commune, définit trois *layers* : *base*, *height* et *top*.

Base layer correspond au placement usuel des enceintes retrouvé dans le format 5.1 et 7.1. Les canaux principaux sont ceux cités en Figure 3 auxquels s'ajoutent parfois des canaux intermédiaires entre les voies gauche, centre et droite, nommés *left center* (Lc) et *center right* (Rc). Le format 7.1 SDDS de la société Sony utilise par exemple ce type de canaux. *Height layer* définit un nouvel espace au-dessus du *base layer* et des spectateurs. *Top layer* regroupe enfin les enceintes positionnées au plafond. Ces nouveaux canaux recensés par le standard SMPTE sont présentés en Figure 4.



Nom par canal <i>height layer</i>	Abr.
Left height	Lh
Right height	Rh
Center height	Ch
Left surround height	Lsh
Right surround height	Rsh
Left side surround height	Lssh
Right side surround height	Rssh
Left rear surround height	Lrsh
Right rear surround height	Rrsh

Nom par canal <i>top layer</i>	Abr.
Left top surround	Lts
Right top surround	Rts
Top surround	Ts

Figure 4 : Répartition des canaux audio pour les couches *height* et *top* selon la recommandation SMPTE ST 2098-5:2018

Auro 11.1

L'Auro 11.1 est un format de diffusion multicanal développé par la structure Auro Technologies qui s'est associée à Barco, fabricant de projecteurs, pour sa distribution en salle de cinéma. Il est présenté à la 40th *International AES Conference of spatial audio* de Tokyo en 2010 en même temps qu'un ensemble de technologies associées pour sa mise en exploitation (Auro Technologies, 2015). Si un grand nombre de films sont actuellement distribués dans ce format, très peu de salles sont équipées en France.

L'Auro 11.1 s'inscrit aujourd'hui dans un environnement plus global de la marque, nommé Auro-3D, qui désigne un ensemble d'outils de travail et de formats de diffusion allant du 9.1 au 13.1 pour des installations *home cinema*.

La configuration spatiale des canaux pour ce format est divisée selon les trois couches présentées par la SMPTE. Le *base layer* ainsi que le *height layer* sont composés de cinq canaux disposés en un format 5.0. Le *top layer* est lui composé d'un seul canal central, parfois appelé *voice of god*, situé au centre du plafond.

Une version 13.1 est également déclinée en divisant les canaux surround du *base layer* en côtés et arrières.

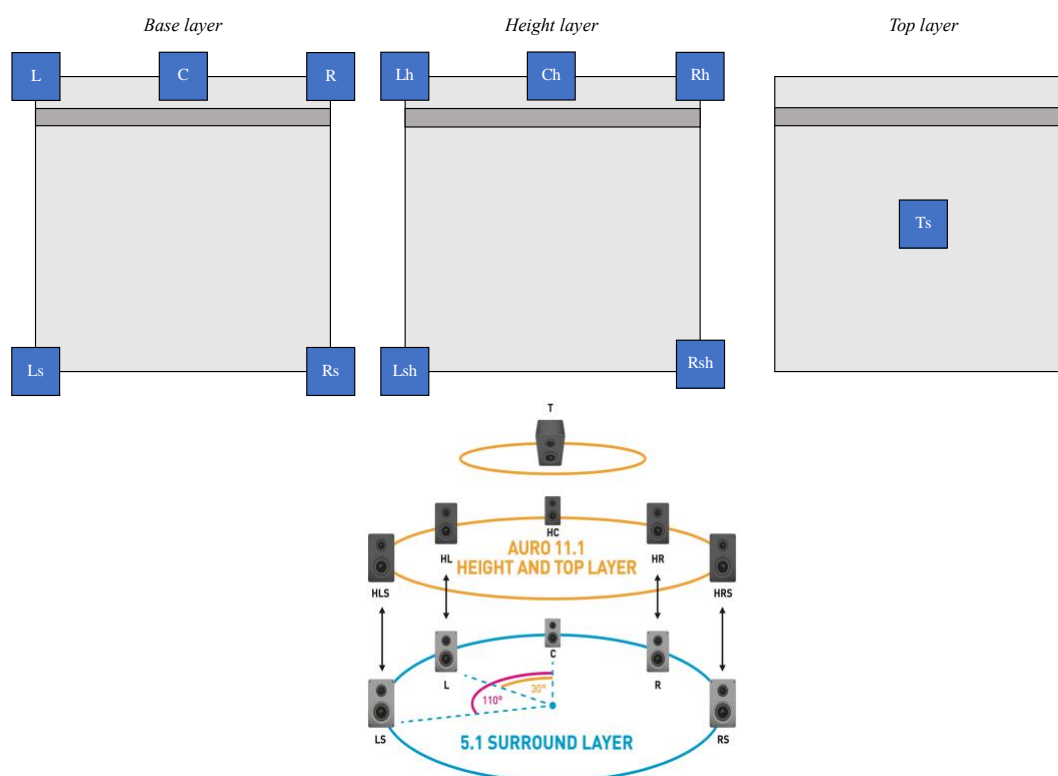


Figure 5 : Répartition par couche des canaux audios pour le format Auro 11.1 (Auro Technologies, 2015)

NHK 22.2

NHK 22.2 est un format de diffusion multicanal développé par Hamasaki *et al.* (2006) du laboratoire de recherche science et technologie de la NHK, compagnie publique de diffusion télévisuelle et radiophonique japonaise. Il fut présenté au grand public lors de l'exposition universelle de Aichi en 2005 pour accompagner le nouveau standard UHDTV Super Hi-Vision (4320p × 7680p).

Ce système s'inscrit donc dans un contexte de diffusion télévisuelle. Il peut être déployé en matérialisant chaque canal par une enceinte autour du spectateur ou par des systèmes de réduction binaurale ou transaurale (Matsui, 2015).

Le positionnement de ses canaux diffère du standard SMPTE, dont le champ d'application se limite à l'exploitation de salle. Le *base layer* constitue un format 7.1 auquel sont ajoutés les deux canaux intermédiaires *left center* et *center right*. Le *top layer*, parfois aussi appelé *upper*, est composé de neuf canaux au-devant, au milieu et au-derrrière de la salle. Ce format intègre enfin une dernière couche *lower* en dessous du spectateur par rapport au plan horizontal. Y sont positionnés trois canaux en façade ainsi que les deux canaux LFE.

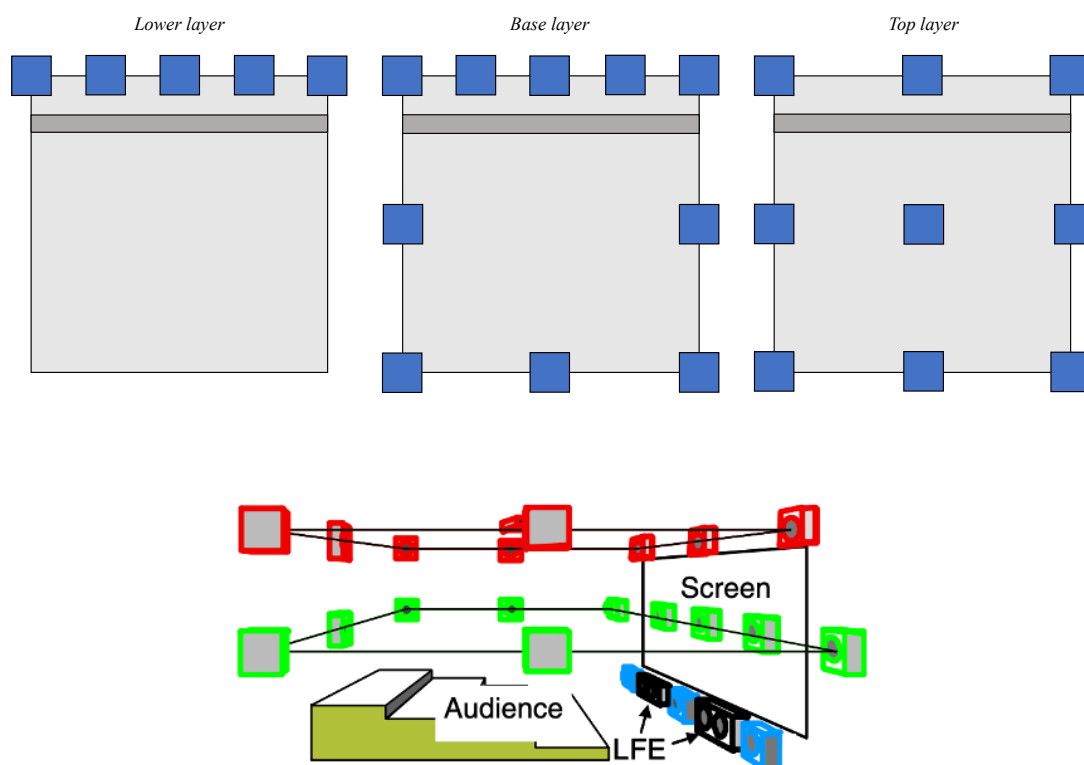


Figure 6 : Répartition par couche des canaux audios pour le format NHK 22.2 (Hamasaki *et al.*, 2006)

Dolby Atmos

Dolby Atmos désigne un environnement de travail et de diffusion développé par la société Dolby Laboratories. Il fut introduit en 2012 avec le film d'animation *Brave (Rebelle* pour son titre français), premier long-métrage à être mixé et diffusé dans ce format d'écoute. C'est un système dit immersif utilisant des canaux zénithaux et qui se distingue par l'utilisation hybride des modes de mixage *bed* et *objet*. Nous différencierons ces deux notions dans le paragraphe suivant. En France, il faudra attendre 2013 avec *En Solitaire* de Christophe Offenstein pour écouter un film en Dolby Atmos. Ce dernier est aujourd'hui très répandu : fin 2020, ce sont deux-cent-quarante-deux salles françaises qui sont équipées de ce système et plus de cinq-cents films qui ont déjà été mixés dans ce format (Robert, 2021).

L'appellation mode objet, *object-based* en anglais, ou mixage orienté objet, s'oppose au mode canal ou *bed*, système de spatialisation communément utilisé lors de l'étape du mixage qui consiste à répartir des sources sonores autour du spectateur par un jeu de panoramiques entre des canaux de diffusion définis (*left, center, right, etc.*). En mode objet, le nombre de canaux ou leur configuration ne sont plus pris en compte. Une source sonore est déclarée dans un espace virtuel par des coordonnées x, y et z. Cet espace virtuel est restitué en temps réel, en auditorium comme en salle, à l'aide d'un processeur qui tient compte des caractéristiques dimensionnelles de la salle et du placement des enceintes.

Ces méthodes de calculs s'appuient sur des principes de panoramique multicanal comme le VBAP (*Vector-Based Amplitude Panning*) et le DBAP (*Distance Based Amplitude Panning*) (Denizot, 2016).

Ces procédés ne seront pas présentés ici ; les formats que nous étudions dans cette recherche s'appuient sur un mixage orienté canal et ne sont pas soumis à ce type de calculs. Pour plus de précisions concernant le mixage en mode objet au cinéma, nous invitons le lecteur à se diriger vers le travail de recherche de Denizot (2016).

Une particularité du format Dolby Atmos est donc sa transportabilité, c'est-à-dire que chaque processeur tiendra compte des spécificités du système de diffusion déployé et adaptera la spatialisation des objets sonores en conséquence.

La configuration spatiale des canaux en mode canal pour le format Dolby Atmos suit un format 7.1, auquel viennent s'ajouter :

- deux canaux zénithaux *left top surround* et *right top surround*
- deux subwoofers en complément des canaux surround gauche et droit, dont les signaux diffusés sont construits dans le processeur et inaccessibles à l'étape du mixage
- un groupe d'enceintes, appelées *wide* : leur nombre varie en fonction de la taille de l'auditorium ou de la salle ; elles sont disposées entre les canaux de façade et les canaux surround et ne sont accessibles qu'en mode objet (AFSI, 2020).

Cette configuration est présentée en Figure 7. Les deux haut-parleurs représentés en vert entre les canaux gauche, centre et droit sont auxiliaires. Il est recommandé qu'ils soient présents lorsque la taille de l'écran de projection est supérieure à douze mètres (Dolby Laboratories, 2012) pour que des effets de panoramiques soient perçus de manière plus homogène. Ils ne sont pas à confondre avec les canaux Lc et Rc, utilisés nativement dans le format NHK 22.2 par exemple.

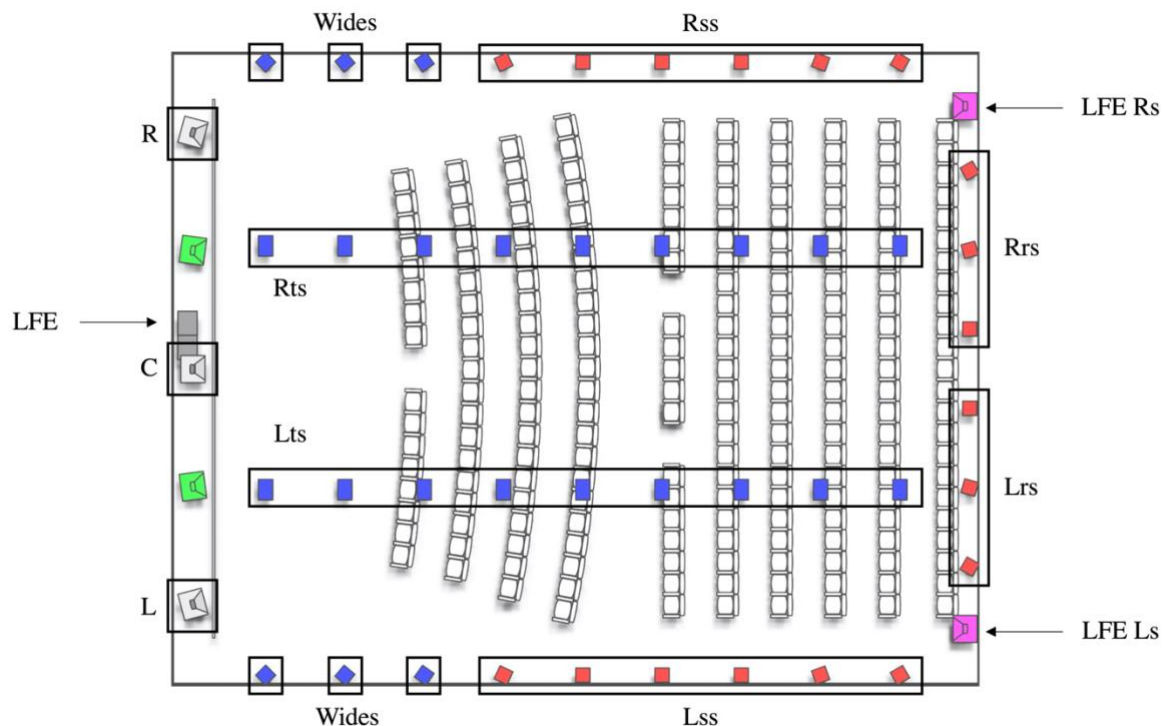


Figure 7 : Répartition des canaux audio pour le format de diffusion Dolby Atmos (Dolby Laboratories, 2012)

3.1.3 – Deux études comparatives de formats de diffusion

Hamasaki et al. (2006)

En 2006, Hamasaki *et al.* publient une étude dans laquelle ils souhaitent mesurer l'impact des informations de hauteur transmises par un système de diffusion sonore sur le ressenti global du spectateur, dans le cadre de projections audiovisuelles. Ils ont pour cela déterminé trois formats de diffusion à comparer : 2.0, 5.0 et NHK 22.2.

L'expérience s'est déroulée dans deux environnements différents. Le premier (E1), que nous présenterons ici, était une salle de cinéma des laboratoires de la NHK avec une diagonale d'écran de 11,43 m. Le second (E2) était une salle de post-production audio avec une diagonale d'écran de 127 cm. La configuration des deux systèmes de diffusion sonore, représentée en Figure 8, était identique ; les enceintes étaient calibrées au même niveau.

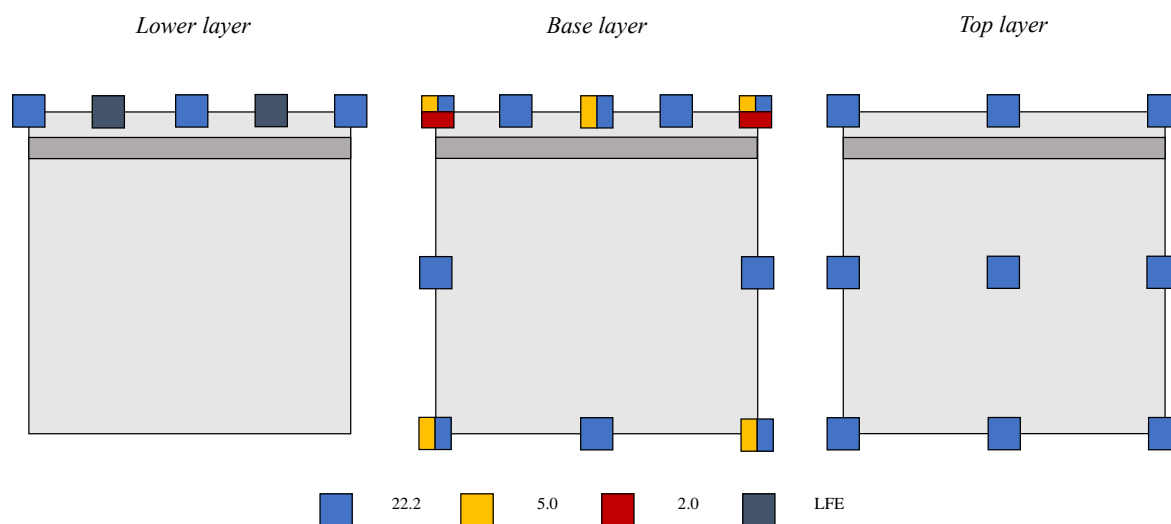


Figure 8 : Répartition des canaux du système d'écoute utilisé dans l'étude de Hamasaki et al. (2006)

Les chercheurs ont sélectionné huit extraits audiovisuels de nature différente (lancement d'une fusée, chœur d'enfants, etc.) et les ont déclinés sous les trois formats étudiés. Vingt-quatre stimuli ont été créés. Ils ont été proposés à l'écoute pour cinquante-trois sujets lors de E1 et quarante-et-un sujets pour E2.

Chaque sujet était invité à noter l'ensemble des stimuli en évaluant vingt-neuf critères à travers une échelle de Likert (1-7). Parmi eux, nous retrouvons *interesting* et *impressive* qui interrogent le ressenti global du sujet. D'autres critères, comme *real*, *presence* et *envelopment*

s'appuient directement sur leurs sens, et traduisent un sentiment d'immersion spatiale. Enfin, des critères tels que *direction* et *up/down* permettent de mesurer directement l'impact d'un format de diffusion sur la sensation de hauteur des sources (p. 6).

Lors de l'analyse des résultats pour l'environnement E1, Hamasaki *et al.* ont constaté que les sujets avaient attribué au système 22.2 une plus grande faculté à procurer une sensation de hauteur des sources que les autres formats pour les extraits présentant un grand nombre de sources mouvantes en trois dimensions. Ces résultats étaient corrélés à une meilleure appréciation du champ sonore et une sensation d'enveloppement, de présence et de localisation plus élevée. Cependant, sur certains programmes n'utilisant pas d'effets sonores particuliers, les chercheurs n'ont pas observé de différence notable entre les formats 22.2 et 5.0.

À partir de cette analyse, les chercheurs ont validé l'hypothèse que les informations de hauteur pouvaient être correctement restituées à travers un système utilisant plusieurs couches de haut-parleurs et que celles-ci permettaient d'accroître le sentiment de présence du spectateur. Dans un second temps, cette expérience met en valeur l'influence de la nature des stimuli présentés sur l'analyse de critères réalisée par les sujets. Ceux exploitant à une plus forte intensité les possibilités de spatialisation offertes par le système 22.2 ont été considérés comme nettement plus immersifs que les autres.

Oramus et Neubauer (2019)

Oramus et Neubauer comparent dans une étude de 2019 plusieurs extraits audios et audiovisuels mixés en Dolby Atmos et réduits par ce même système au format 5.1. Pour la présentation de cette étude, nous nous concentrerons uniquement sur l'analyse des extraits audiovisuels.

Les auteurs ont posé comme question de recherche la potentielle différenciation de ces deux formats pour des auditeurs expérimentés AE (professionnels, étudiants ou amateurs de son) et inexpérimentés AI. Ils ont émis l'hypothèse que l'ensemble des auditeurs serait capable de distinguer les deux formats, notamment grâce à trois avantages supposés du système Dolby

Atmos : « les sons provenant du dessus, l'amélioration de la qualité audio et de la correspondance des timbres, un meilleur contrôle spatial et une plus grande résolution »⁷ (p. 2).

Pour l'implémentation du test, les expérimentateurs ont défini cinq critères : *spatial immersion*, *localization*, *dynamics*, *audio quality* et *preference*. Nous précisons que *spatial immersion* est présenté dans cette étude comme la sensation d'être entouré de son. Il s'agit de la définition que nous avons faite de l'enveloppement en 1.2.1 – *Propositions de critères perceptifs*.

Pour s'inscrire intégralement dans un contexte d'étude cinématographique, les chercheurs ont sélectionné six extraits de films contenant des dialogues, des effets, des ambiances et des musiques. Les films sélectionnés sont sortis en Dolby Atmos avec un downmix 5.1. Les auteurs précisent que les critères d'évaluation préalablement définis ont influencé le choix des extraits. Ainsi, ils en ont choisi certains pour leur capacité à créer un potentiel enveloppement dans le cas de *spatial immersion*, ou avec un nombre important de sources ponctuelles mouvantes pour le critère de localisation.

Les six paires d'extraits Atmos-5.1 ont été diffusées à cent-soixante-sept sujets (soixante-quatre AE et soixante-trois AI) qui, pour chaque paire et à travers dix questions, devaient exprimer leur préférence entre les extraits A et B sur une échelle de notation allant de -4 (préférence pour l'extrait A) à +4 (préférence pour l'extrait B). L'ordre des formats pour chaque paire changeait au cours de l'expérience pour ne pas avoir de biais d'ordre.

Pour l'exploitation des données, certaines valeurs ont donc été inversées pour que -4 corresponde à une grande préférence pour le format 5.1 et la valeur +4 pour la diffusion en Dolby Atmos. Ils ont ensuite réalisé un ensemble de moyennes pour chaque critère Figure 9.

⁷ « sounds originating overhead, improved audio quality and timbre matching, greater spatial control and resolution » (Oramus et Neubauer, 2019, p. 2, notre traduction).

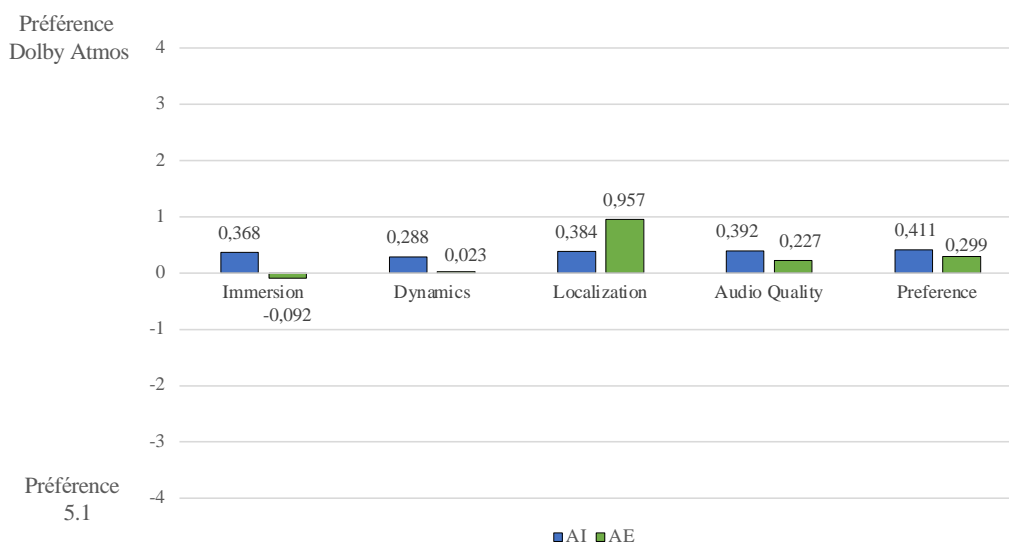


Figure 9 : Histogramme des moyennes des résultats de l'étude à partir des données de Oramus et Neubauer (2019)

Nous remarquons que la plage des moyennes s'étale de $-0,092$ (*spatial immersion AE*) à $0,957$ (*localization AE*). Ces résultats ne permettent pas selon les chercheurs d'observer une différence significative entre les deux formats, quel que soit le critère. Si l'on observe une tendance générale des sujets à préférer les extraits diffusés en Dolby Atmos, les chercheurs précisent qu'ils ont eu des écarts-types relativement importants pour l'évaluation des trois premiers critères. Ceci traduit une grande variation dans les réponses des sujets.

Cette expérience ne permet pas de valider l'hypothèse de départ. Il n'est pas non plus possible de déduire une tendance différente entre les auditeurs expérimentés et inexpérimentés. Les chercheurs ont mis en avant un biais majeur pour expliquer en partie ces résultats : la présence de l'image et sa nature.

« Nous pensons que ces deux formats offrent une expérience différente plutôt 'qu'une meilleure et une moins bonne' et, dans une certaine mesure, c'est aussi une question de contenu diffusé. Un autre facteur à prendre en compte est le fait que ce sont des échantillons audiovisuels qui ont été présentés et que l'image influe sur notre perception et donc sur les résultats. »⁸ (p. 7)

⁸ « We believe that these two formats offer a different experience rather than "better and worse" and, to some extent, it is also a matter of the presented material. Another factor to be considered is that audiovisual samples were presented and that picture influences our perception and thus the results » (Oramus et Neubauer, 2019, p. 7, notre traduction).

Nous émettons alors l'hypothèse que l'effet d'absorption a pu constituer un biais dans cette expérience. Ce processus psychologique, conduisant le sujet à se concentrer sur les situations et les actions des personnages, réduisait peut-être sa capacité à analyser *a posteriori* les critères d'évaluation majoritairement perceptifs étudiés par les chercheurs. Cette étude, ainsi que ses résultats, ont été pris en compte dans l'élaboration de notre propre expérience que nous présentons en deuxième partie de mémoire.

Pour conclure, les deux études présentées ci-dessus illustrent la complexité d'élaboration d'un protocole de test pour évaluer des critères perceptifs sonores en présence d'une image. La nature des sources et l'implication cognitive et émotionnelle du spectateur peuvent constituer des biais majeurs.

L'expérience de Hamasaki *et al.* (2006) permet tout de même de mettre en corrélation le degré de sophistication d'un système, ici en termes de nombre de canaux, et la sensation d'immersion comme indiqué par Agrawal *et al.* (2020).

3.2 – Prise de son multicanale : développement et application professionnelle

Des dispositifs de prise de son multi-microphoniques et prenant en compte l'axe d'élévation ont été développés pour une diffusion de leurs enregistrements directement à travers un système immersif. Ils sont aujourd'hui nombreux et s'adaptent de manière variable aux systèmes de diffusion, en fonction du nombre de microphones utilisés et de leur agencement.

Certains de ces dispositifs sont adaptés à un format de diffusion spécifique. Ono *et al.* (2013) développent par exemple un dispositif d'enregistrement compact et adapté pour une diffusion à travers un système NHK 22.2. D'autres dispositifs d'enregistrement, tels que le *2L-cube* et le *Bowles array* sont quant à eux compatibles avec plusieurs systèmes de diffusion (Roginska & Geluso, 2018).

Si de nombreux dispositifs sont présentés dans des documents de recherche et utilisés localement, comme le dispositif *Hedgehog* de Omoto (dans Kamekawa & Marui, 2019) que nous présenterons plus tard, certains autres sont dorénavant commercialisés. L'OCT-3D *Set*, distribué par exemple par la société Schoeps, est composé d'une base de dispositif 5.0 à laquelle

viennent s'ajouter quatre microphones dirigés vers le ciel pour une restitution à travers des canaux zénithaux.

Il existe aujourd'hui une multitude de systèmes tels que ceux exposés ci-dessus. Afin d'étudier plus précisément ces dispositifs et d'en découvrir de nouveaux, nous invitons le lecteur à se diriger vers les travaux de Roginska et Geluso (2018) ainsi que de Lee (2021).

Dans cette partie, nous nous concentrerons uniquement sur le développement de dispositifs appliqués à l'enregistrement d'ambiances multicanales dans un contexte de production cinématographique.

Nous étudierons dans un premier temps une méthode théorique de développement pour réaliser un tel dispositif, puis nous présenterons deux exemples déjà existants destinés à la captation d'évènements sportifs et à la prise de son d'ambiances. Enfin, à travers deux études, nous explorerons les liens entre l'utilisation de dispositifs d'enregistrements multicanaux et la sensation d'immersion.

3.2.1 – Procédé de développement à partir d'angles utiles

Afin de restituer un espace sonore stéréophonique, deux microphones sont usuellement utilisés pour capter une ou plusieurs sources rayonnantes. Cet espace doit être défini en amont en termes de largeur et de profondeur physique pour pouvoir choisir l'emplacement adéquat du couple de microphones et pour que l'enregistrement puisse être correctement restitué lors de l'écoute. On définit alors un angle utile de prise de son nécessaire, ou *stereo recording angle* (SRA) en anglais.

« L'angle de prise de son d'un système stéréophonique est une caractéristique résultante du choix de directivité, de distance et d'angle entre les capsules microphoniques constituant le système. Il correspond à l'angle solide de captation pour lequel les sources situées à l'intérieur seront restituées entre les deux enceintes acoustiques. » (Noly, 2013, p. 12)

Plusieurs facteurs physiques rentrent donc en compte pour choisir correctement un dispositif stéréophonique. Williams (1984) publie un ensemble d'abaques pour déterminer avec facilité un angle utile en fonction de la distance physique entre les microphones, leur angle et leur directivité. Ces dernières prennent la forme d'un réseau de courbes correspondant aux différentes valeurs d'angle utile souhaitées.

Par exemple, à la lecture du premier abaque de l'Annexe B (pour capsules cardioïdes), à partir d'un angle de $\pm 50^\circ$ souhaité voici quelques combinaisons possibles de distance et d'angle physique entre les microphones : $\approx 40\text{cm}/20^\circ$; $\approx 30\text{cm}/50^\circ$; $\approx 20\text{cm}/90^\circ$; $\approx 15\text{cm}/110^\circ$. Dans le cas limite où la distance serait nulle, nous nous retrouvons dans une configuration XY avec par exemple pour des angles physiques de 90° et 120° , des angles utiles respectifs de $\pm 90^\circ$ et $\pm 70^\circ$.

Williams (2004) a développé ses travaux pour calculer des angles et des dimensions de systèmes d'enregistrement multi-microphoniques, au format 3.0, 4.0, et 5.0. Ces derniers sont présentés sous la forme de plans en vue zénithale, comme celui présenté en Figure 10. Nous précisons que l'étude de ces formats s'effectue dans le respect de la norme ITU - BS.775 relative à la disposition des enceintes pour le format 5.1. Ainsi, le format 3.0 présenté par Williams correspond à une restitution via les canaux L, C et R et le format 4.0 par les canaux L, R, Ls et Rs.

Pour réaliser un système de prise de son 3.0 avec trois microphones, Williams propose de déterminer un angle utile frontal (ou *front triplet coverage*, correspondant aux couleurs rouges et vertes sur la Figure 10) qui, de la même manière que le SRA, vient définir un angle de prise de son dans lequel les sources sonores seront correctement restituées. En complément de ses précédentes abaques, Williams propose plusieurs plans de positionnement de microphones selon l'angle souhaité.

Comme le montre la Figure 10 : Exemple de plan de positionnement de microphones au format 5.0 (Williams, 2004), lorsqu'on ajoute à ce système un couple stéréophonique arrière, il est possible d'élargir cet angle en prenant en compte les angles utiles des microphones L/Ls et R/Rs. Ils sont décrits par Williams comme les angles latéraux (ou *lateral segment coverage* en bleu et en violet) et il est nécessaire de les définir en amont pour pouvoir calculer l'angle physique du couple arrière et la distance entre ce dernier et les microphones frontaux.

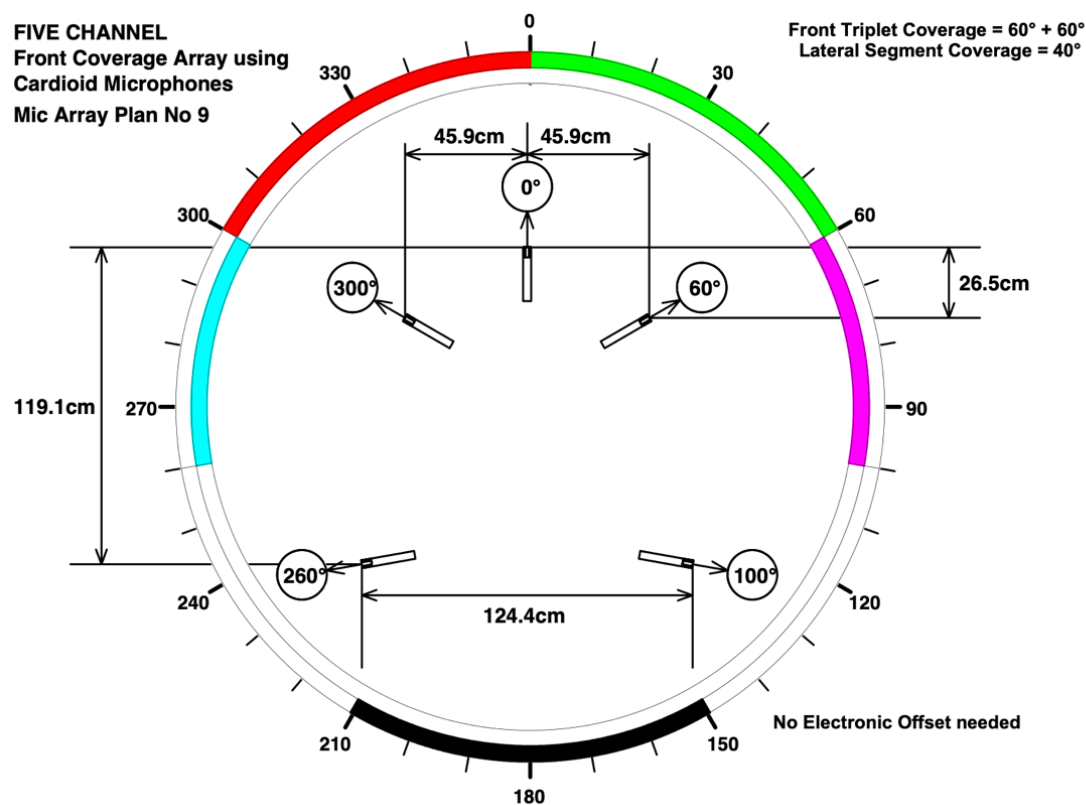


Figure 10 : Exemple de plan de positionnement de microphones au format 5.0 (Williams, 2004)

L'angle utile global (*front sound stage covering* en anglais) est alors défini comme la somme des angles frontaux et latéraux.

Il est possible d'adapter ces configurations à un format 4.0 en remplaçant les trois microphones frontaux par un couple. Dans ce cas, il sera nécessaire de choisir un angle utile pour une restitution stéréophonique via les canaux L et R et le canal central ne sera pas utilisé. La démarche pour intégrer un couple arrière ne change pas.

Les travaux de Williams permettent dorénavant la création de dispositifs d'enregistrement constitués de trois, quatre ou cinq microphones dont les distances et les angles varient en tenant compte de la largeur de la source à capter ou le champ sonore à restituer.

En 2008, en conséquence du développement du son 7.0 (Blu-ray, Dolby Surround 7.1, etc.), Williams souhaite augmenter une nouvelle fois le nombre de canaux à enregistrer. Il propose plusieurs techniques de construction pour arriver jusqu'au format 8.0, dont le principe du « Back to back array design » (Williams, 2008, p. 4). Ce dernier repose sur la mise dos à dos de deux plans de formats 4.0 et 5.0 pour créer, par addition, un format 6.0 (2*4.0 dos à dos), 7.0 (1*4.0 + 1*5.0) ou 8.0 (2*5.0).

Par exemple, la Figure 11 présente un dispositif (a) 5.0 avec un angle global de prise de son de 200° . Les micros Ls et Rs sont séparés d'une distance $L_A=160,8$ cm. Le dispositif (b) 4.0 a, quant à lui, un angle global de 160° . Les micros Ls et Rs sont séparés d'une distance $L_B=170$ cm.

Selon Williams, les deux plans sont compatibles pour être mis dos à dos car le total des angles globaux est égal à 360° et les deux distances entre Ls et Rs sont suffisamment proches. Ainsi le plan 7.0 créé aura l'allure suivante :

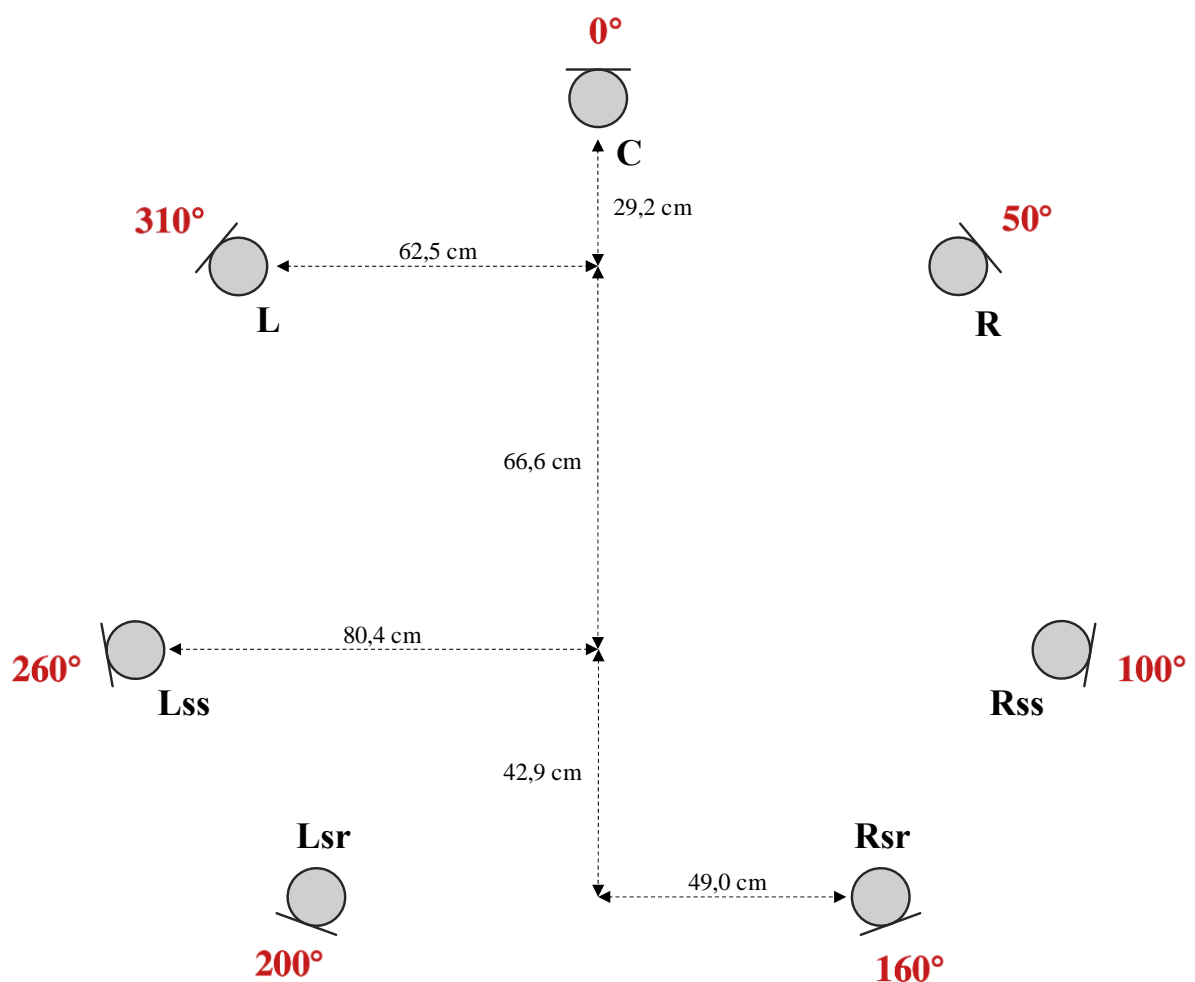


Figure 12 : Plan de positionnement de microphones au format 7.0, créé à partir des plans complémentaires de la figure 11

3.2.2 – Exemples de dispositifs dédiés à l'enregistrement d'ambiances

Le dispositif commercialisé ORTF-3D

L'ORTF-3D est un dispositif de prise de son multi-microphonique développé par la société Schoeps en 2012. Il s'agit d'un système compact, dont le nom fait référence au couple ORTF stéréophonique. Il est aujourd'hui majoritairement utilisé pour la captation d'événements sportifs et la prise de son d'ambiances (Wittek & Theile, 2017). Le dispositif se décline sous deux versions en fonction de son utilisation en intérieur (*Indoor-set*) ou en extérieur (*Outdoor-set*). Bien que leurs dimensions varient légèrement, l'angle utile de prise de son reste le même. Nous présenterons ici le système *Outdoor-set*.



Figure 13 : Photographie de l'ORTF-3D en configuration outdoor-set (©Schoeps)

L'angle et le positionnement de ces microphones peuvent être déterminés à partir des caractéristiques du couple ORTF. Ce dernier désigne un couple stéréophonique dont les capsules des microphones sont cardioïdes, espacées de 17 cm et avec 110° d'angle physique. Cela permet d'obtenir un angle utile de 90° . Sur cette base, Schoeps a dans un premier temps développé un dispositif d'enregistrement multicanal, compatible 5.0, nommé *ORTF Surround*. Celui-ci est fabriqué en espaçant de 10 cm deux couples de microphones dos à dos. Ces derniers ont le même angle utile qu'un couple ORTF mais la directivité de leurs capsules, leur distance et leurs angles diffèrent. Cette fois-ci, ce sont des capsules supercardioïdes, espacées de 20 cm et avec un angle physique de 80° qui sont utilisées.

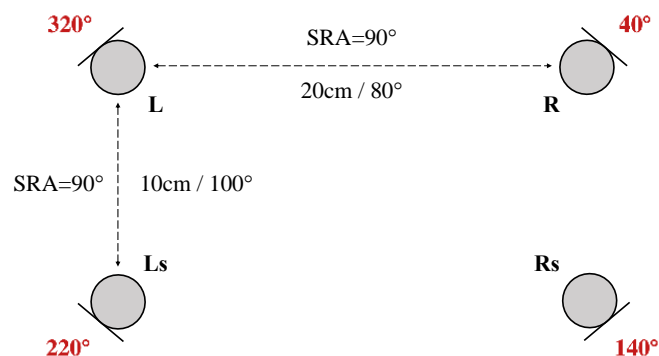


Figure 14 : plan de positionnement des microphones pour une configuration ORTF surround

Selon les abaques de Williams (Annexe B), les quatre couples ainsi formés ont un même angle utile de 90° .

C'est en superposant deux ORTF *Surround* l'un sur l'autre et en inclinant les capsules coïncidentes de $\pm 45^\circ$ qu'est obtenu le dispositif ORTF-3D. Ces capsules forment un angle utile d'environ 130° . Ce dispositif permet alors d'enregistrer huit canaux : *left, right, left surround, right surround, left top, right top, left top surround et right top surround*.

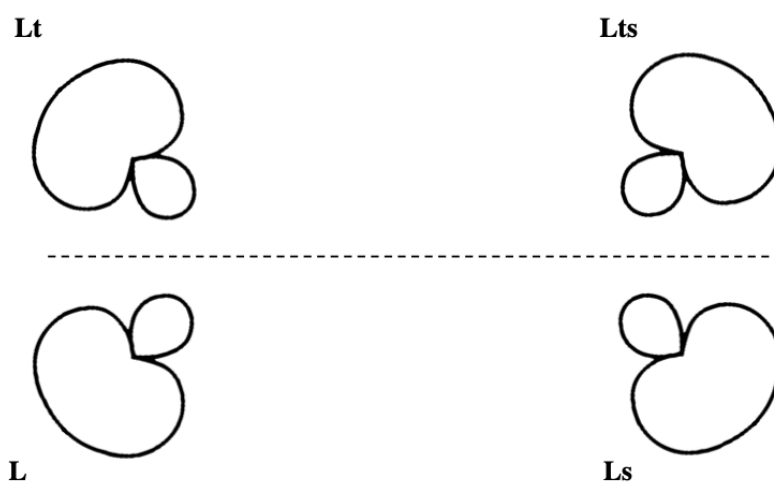


Figure 15 : Photographie et schéma de l'ORTF-3D en vue de côté

L'ORTF-3D, dans sa disposition initiale, est composé de quatre microphones CCM41 (capsules supercardioïdes) pour l'ORTF *surround* du dessous ainsi que de quatre CCM41V (« V » pour vertical) pour l'ORTF *surround* du dessus. Ces derniers microphones possèdent la particularité d'avoir leur capsule pivotée à 90° . Ainsi, l'ensemble du dispositif peut être assemblé de manière compacte dans une cage.

L'ORTF-3D a été testé dès 2014 puis utilisé à de nombreuses reprises pour la captation d'évènements sportifs. De l'année 2018 à 2021, il fut équipé dans l'ensemble des stades de

football accueillant les championnats d'Europe pour capturer les ambiances et les réactions des foules de supporters (Mac, 2021).

Le dispositif est également utilisé depuis 2019 par la société Boom Library, dont l'activité principale est la vente de banques de sons à destination de la post-production cinéma et du jeu vidéo. Un ensemble de fonds d'air ainsi que des ambiances naturelles et urbaines ont déjà été enregistrés de cette manière. Des réductions au format stéréo sont également proposées.

L'ORTF-3D a l'avantage d'être compact et prêt à l'emploi sans aucune préparation des microphones. Ces derniers étant déjà fixés dans leur cage, le dispositif est particulièrement efficace lorsqu'il n'y a que peu de temps pour enregistrer la ou les sources souhaitées. Un autre avantage pratique est qu'il possède sa propre housse imperméable pour être utilisé en cas de mauvais temps.

Il est cependant impossible de modifier sa configuration. Son intégration en post-production est également complexe, étant donné que la configuration de ses canaux diffère du standard SMPTE (2018). Son format n'étant pas nativement intégré dans les logiciels d'édition audio - notamment Pro Tools - il est nécessaire d'éclater ses enregistrements en plusieurs pistes pour travailler. Dans une recherche d'ergonomie et d'efficacité, cette tâche n'est pas souvent réalisée.

Pour diffuser ces enregistrements à travers un système Dolby Atmos par exemple, il est nécessaire d'utiliser quatre objets sonores (approche orientée objet) à positionner aux coins hauts de la salle afin d'y affecter les canaux *left top*, *right top*, *left top surround* et *right top surround*. Le nombre d'objets étant rapidement limités à l'étape du mixage, cette utilisation ne sera donc souvent pas prioritaire.

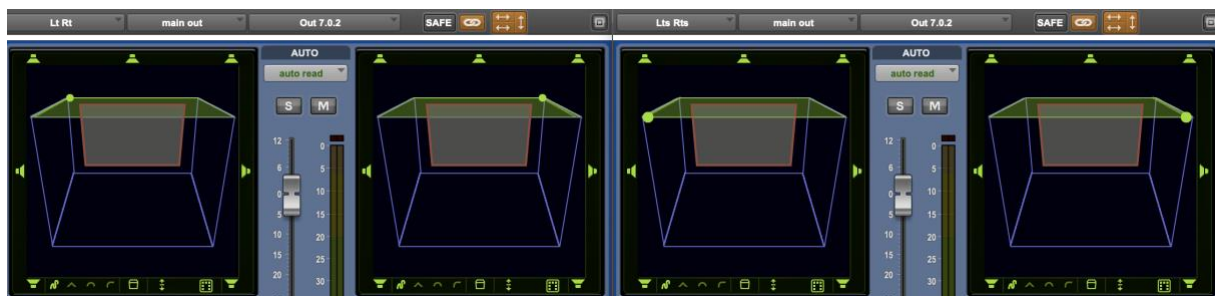


Figure 16 : Spatialisation des quatre sources d'élévation de l'ORTF-3D dans le surround panner de Pro Tools

L'arbre Atmos de HAL Audio

HAL Audio est une marque française fondée par un groupe de professionnels du son à l'image et spécialisée dans la recherche et développement de plug-ins audios, ainsi que la production de bibliothèques sonores. Elle est rattachée à la société parisienne Poly Son, prestataire technique en laboratoire numérique et studio de post-production image et son.

En 2023, la marque commercialise sa première sonothèque d'ambiances urbaines *Urban Atmos*. Pour sa réalisation, HAL a développé un dispositif d'enregistrement multicanal au format 7.0.2 nommé l'arbre Atmos (Figure 17). Ce dernier terme fait référence au format immersif de la société Dolby, avec qui la marque a collaboré pour la construction de leur dispositif.



Figure 17 : Photographie de l'arbre Atmos de HAL Audio

Nous avons pu nous entretenir avec deux professionnels de HAL ayant collaboré à la création du système. Ils nous ont rapporté que l'objectif initial d'une telle recherche était d'apporter aux spectateurs des sensations de présence et d'enveloppement naturelles lors de diffusions en Dolby Atmos, jusqu'alors limitées et réalisées par des artifices de montage et de mixage. Ainsi, la création de ce dispositif avait également une visée pratique, celle de permettre un gain de temps et d'efficacité dans la chaîne de post-production audio en proposant des enregistrements prêts à l'emploi et remplissant l'ensemble des canaux.

Ce dispositif met en jeu un microphone par canal. Il est composé d'une base 7.0 à laquelle viennent s'ajouter deux autres microphones supplémentaires pour alimenter les canaux zénithaux. C'est un système basé sur la présence de différence de temps (Δt), avec des microphones plus ou moins espacés. Les collaborateurs rencontrés nous ont indiqué avoir réalisé plusieurs comparatifs d'écoute en auditorium et que cette configuration permettait de ressentir de manière flatteuse une acoustique ou un paysage sonore.

À l'origine de la recherche, une première configuration fut déterminée à partir d'une proposition de Williams (2008). Celle-ci consistait à créer un dispositif 7.0 en positionnant sept microphones équidistants autour d'un point fictif pour prendre le son à 360°. Les capsules étaient alors omnidirectionnelles.

D'après des tests d'écoute et les recommandations émises par les consultants de Dolby France, la corrélation entre l'ensemble des microphones (y compris ceux des canaux zénithaux) devait être minimisée. Cela nécessitait l'utilisation de capsules cardioïdes et d'avoir une différence de hauteur signifiante entre les couches 7.0 et *top*. Le dispositif a donc été modifié selon ces indications.

Neuf Sennheiser mkh 8040, à la directivité cardioïde, sont actuellement utilisés lors du déploiement de ce dispositif. HAL a travaillé avec la société Dolby pour préciser sa configuration, en prenant notamment en compte les spécificités de placement des enceintes dans le format Dolby Atmos.

Ce système offre donc un avantage pratique d'ergonomie quant à son intégration en post-production. Sa configuration est dédiée à la diffusion en Dolby Atmos et le format 7.0.2 est nativement pris en charge par plusieurs logiciels audios. De par sa configuration, en sélectionnant certains des neuf canaux, il est également possible d'obtenir une compatibilité parfaite 7.0, 5.0 et LCR. Ce dispositif s'inscrit alors parfaitement dans un contexte d'enregistrement d'ambiances pour la production sonore cinématographique.

Enfin, le dispositif est articulé autour d'un mécanisme d'ouverture et de fermeture semblable à celui d'un parapluie. Ainsi, malgré sa taille imposante, son temps de mise en place est réduit ce qui lui permet d'être opérationnel tant en *field recording* que sur des plateaux de tournage pour des sons seuls spécifiques.

3.2.3 – Deux études comparatives de dispositifs d'enregistrement

Kamekawa et Marui (2019)

Kamekawa et Marui réalisent en 2019 une étude dans laquelle ils comparent trois dispositifs d'enregistrement multicanaux pour une restitution dans un format d'écoute NHK 22.2. Ces dispositifs étaient :

- Le *Spaced array* : une disposition de vingt microphones omnidirectionnels espacés et répartis sur trois couches *lower* (trois), *base* (huit) et *top* (huit).
- Le *Hedgehog* : dispositif proposé par Akira Omoto, composé de vingt-quatre microphones canons répartis sur les couches *lower* (huit), *base* (huit) et *top* (huit). Cette fois-ci les bases des microphones sont coïncidentes et les capsules pointent vers des directions différentes. Ce système est qualifié de *one-point array* vis-à-vis du système précédent.
- *Ambisonics* : un système ambisonique format A avec quatre capsules cardioïdes décodées en vingt-deux canaux.

Un solo de violon fut enregistré à l'aide de ces trois systèmes. Pour l'expérience, les expérimentateurs ont demandé à vingt-deux sujets de noter un ensemble de stimuli correspondant à des extraits de ces enregistrements, selon différents critères préalablement définis.

L'analyse des résultats a montré que le système ambisonique était perçu d'un point de vue fréquentiel comme dur (critère de définition *hard*). Le système *Hedgehog* était surtout perçu comme ayant beaucoup de médium (*rich in middle range*) et avait une image sonore large (*wide sound image*). Enfin, le système *Spaced array* constitué de microphones espacés possédait, selon les sujets, une image sonore précise (*clear sound image*) et avec une forte sensation de présence dans l'espace sonore projeté.

Nous observons que ce dernier critère est défini comme un indice d'immersion sensorielle dans les travaux présentés en 1.2.2 – *Questionnaires de présence et d'immersion* (Rigby *et al.*, 2019 ; Bouchara & Okuya, 2020).

On peut observer deux différences entre le champ d'action de cette étude et celui du présent mémoire. Cette étude fut réalisée dans un contexte de prise de son musicale. Dans ce cas, la source sonore (le violon) est ponctuelle et plus proche des microphones que celles retrouvées lors de l'enregistrement d'ambiances. De plus, le contexte de restitution sonore NHK 22.2 n'est pas utilisé au cinéma. Ce format intègre une couche *lower* située en dessous du spectateur qui est inexistante dans les systèmes de diffusion cinéma accessibles aujourd'hui.

Halnet et Ehret (2017)

En 2017, Halnet et Ehret comparent dans leur étude trois dispositifs d'enregistrement multicanaux, chacun mixé à la bande sonore d'un film dans un environnement Dolby Atmos. Ces dispositifs étaient un ORTF-3D, un triple MS et un système proposé par Williams (2008) composé de six microphones supercardioïdes équidistants autour d'un point fictif. À cette base ont été ajoutés deux microphones pour les canaux zénithaux.

Pour préparer cette étude, les auteurs ont enregistré des ambiances à intégrer dans le mixage d'un film. Pour chaque dispositif, une ambiance a été enregistrée dans une aire industrielle et une autre sur un espace de parking. Ensuite, ces ambiances ont chacune été mixées individuellement (adaptation du niveau sonore et égalisation) au reste de la bande sonore.

Des extraits de 00'25'' ont été créés pour réaliser l'étude. Le sujet était invité, pour trente répétitions, à comparer deux extraits aléatoires (correspondant à chaque fois à un dispositif) et à noter ses préférences selon cinq critères *spaciousness, evenness of the distribution, overall quality, realism regarding the scene* et *general preference*.

Si la présente étude n'indique pas d'avantages de précisions sur les conditions d'enregistrement, de diffusion et sur l'exploitation des données recueillies, les auteurs observent que le système proposé par Williams a été majoritairement préféré par les sujets, contrairement au triple MS qui arrive au dernier rang. Le dispositif surpasse sur tous les critères les deux autres systèmes, à l'exception de *realism regarding the scene* où, pour quelques extraits, il ne semble pas y avoir de différence notable entre celui-ci et l'ORTF-3D.

Dans leurs discussions autour de l'expérience, les auteurs font également un retour d'expérience sur l'ergonomie de ces systèmes, lors des enregistrements et en post-production.

Ainsi, le système de Williams était long à mettre en place de par les huit pieds de microphones à ajuster individuellement. Pour enregistrer des ambiances lors de tournages, il ne serait pas donc pas pertinent d'utiliser ce système. Si l'ORTF-3D et le triple MS sont relativement simples à déployer, étant déjà préparés dans leur cage, leur intégration en post-production est difficile. Nous avons vu que le format d'enregistrement cuboïde de l'ORTF-3D ne rentre pas dans les normes habituelles de productions. Le triple MS, quant à lui, nécessite un dématricage chronophage en l'absence d'outils commercialisés.

Pour conclure, dans la promesse d'un cinéma plus sensationnel, comme illusion d'une réalité alternative, les formats de diffusion immersifs enveloppent les spectateurs dans un dôme sonore par l'utilisation d'un nombre toujours plus important de canaux. Ces derniers se matérialisent par différentes couches d'enceintes réparties dans toute la salle ainsi qu'au plafond. Si Hamazaki *et al.* (2006) observent bien en passant du format 5.0 au 22.2 une sensation accrue d'enveloppement, de localisation et de présence, qui sont des indices d'une immersion des sujets, la projection de l'image et ce qu'elle présente influencent leurs résultats. L'immersion du spectateur se construit donc lors d'un processus d'analyse globale du document audiovisuel, perceptif comme cognitif, sans dissociation entre image et son.

Les possibilités de spatialisation offertes par ces systèmes relèvent de choix esthétiques de montage et de mixage. Toujours dans une volonté continue de restituer un champ sonore naturel et enveloppant, par extension aux systèmes de prise de son stéréophoniques, des dispositifs d'enregistrement multicanaux ont été développés spécifiquement pour ces formats.

Nous avons analysé précédemment les usages spécifiques des ambiances au sein de la bande sonore d'un film. Au regard de ces formats immersifs, une ambiance multicanale favorise-t-elle l'immersion spatiale d'un spectateur ? Son attention est-elle toujours focalisée sur l'écran ou se sentira-t-il distrait par les sources autour de lui ? Quels éléments visuels peuvent influencer son processus d'analyse ?

Dans la seconde partie de ce mémoire, nous proposerons une réponse à ces questions en présentant une expérience que nous avons menée pour mesurer l'impact d'un format de diffusion sur la sensation d'immersion d'un sujet au regard de la projection d'ambiances sonores et visuelles.

Seconde partie – Réalisation d'une expérience : influence d'un format de diffusion sur la sensation d'immersion

IV – Présentation de l'expérience

4.1 – Contexte d'étude

Nous avons évoqué en première partie qu'une immersion en tant qu'état psychologique pouvait se définir comme la sensation de se sentir entouré par un ensemble de stimulations multisensorielles ou d'être absorbé par les éléments d'une narration.

Si ces définitions peuvent être interdépendantes, les processus psychologiques mis en jeu sont différents. En effet, l'absorption se traduit par un engagement profond et une attention du spectateur portée entièrement sur les éléments constitutifs du récit (identification aux personnages, réaction émotionnelle aux situations, etc.). C'est un processus cognitif qui s'installe sur la durée du film, à la différence de l'immersion sensorielle qui s'instaure davantage par notre capacité d'analyse en temps réel des stimulations visuelles et sonores que nous recevons.

Dans un contexte d'étude, il nous semble impossible de pouvoir mesurer ces aspects en contrôlant l'ensemble des variables et valider écologiquement un test. Hamasaki *et al.* (2006) ainsi que Oramus et Neubauer (2019) précisent par exemple dans leur étude que la nature des extraits a constitué un biais majeur dans leur expérience.

Nous avons souhaité réaliser une expérience ayant comme variable des formats de diffusions. Nous avons donc choisi de nous concentrer uniquement sur une mesure de l'immersion sensorielle. Ces résultats pourront servir, dans le futur, de soutien pour une expérience visant l'étude du phénomène d'absorption.

4.2 – Questions de recherche et critères d'évaluation

Nous souhaitons réaliser une expérience afin de mettre en valeur l'impact potentiel d'un format de diffusion sur le ressenti d'immersion d'un sujet dans le cadre d'une projection audiovisuelle. Nous souhaitons dans un second temps déterminer si la perception de l'espace présenté à l'écran diffère selon les formats.

Nous émettons alors l'hypothèse que l'augmentation des formats en termes de canaux est corrélée à une sensation accrue d'immersion. Nous faisons la deuxième hypothèse que la perception de l'espace est également impactée.

Comme variable de l'expérience, nous avons déterminé quatre formats de diffusion à comparer : LCR, 5.0, 7.0 et 7.0.2. Nous souhaitons nous inscrire dans un cadre professionnel d'actualité, c'est pourquoi les trois premiers formats ont été retenus en raison de leur utilisation standardisée dans le contexte de la diffusion cinématographique. Nous y avons ajouté le format 7.0.2, aujourd'hui intégré dans plusieurs systèmes de diffusion propriétaires tel que le Dolby Atmos et prenant en compte l'axe d'élévation.

Notre expérience prend la forme de visionnages de plusieurs extraits audiovisuels avec une collecte de données par la réponse à un questionnaire. Nous avons défini six critères d'évaluation pour l'élaboration de celui-ci, quatre pour l'immersion sensorielle et deux pour la perception de l'espace :

- L'enveloppement
- La localisation
- La sensation de présence
- La sensation d'être captivé
- La cohérence des espaces visuels et sonores
- La grandeur de l'espace imaginé

Comme stimuli évalués, nous avons présenté aux sujets huit extraits audiovisuels ne contenant aucune mise en scène pour limiter le biais d'engagement du spectateur. Ces extraits présentent des ambiances sonores caractéristiques que nous avons enregistrées, accompagnées de leurs images associées.

4.3 – Résumé de l'expérience

L'expérience s'est déroulée en auditorium de mixage, le temps de deux journées, dans un environnement Dolby Atmos. Nous ne souhaitons pas tester ce système en tant que tel, mais nous avons choisi de l'utiliser par souci d'ergonomie car il permet la diffusion de sources avec élévation dans le format 7.0.2. Nous avons prêté notre attention à ce qu'aucune altération des signaux sonores proposés à l'écoute ne soit provoquée par le système lui-même (gestion du LFE, downmix, etc.). L'approche orientée objet ne sera donc pas utilisée et les stimuli diffusés seront directement affectés dans les canaux de sortie. De cette manière les conclusions apportées lors de l'analyse des résultats seront indépendantes du système de diffusion utilisé.

Huit enregistrements d'ambiances multicanales, avec vidéo associée, ont été réalisés en amont de l'expérience. Nous avons développé notre propre dispositif de prise de son compatible pour les quatre formats étudiés. Cela nous a permis de pouvoir décliner chaque enregistrement dans les quatre formats souhaités. Pour chaque format, des extraits de 1'30'' ont été sélectionnés. Au total, trente-deux extraits ont donc été créés.

Ces extraits ont été proposés à l'écoute puis évalués par les sujets. Après chaque visionnage, les sujets étaient invités à répondre à un questionnaire reprenant les critères définis dans nos objectifs de recherche. Au total, vingt-huit sujets ont participé à l'expérience. Ils ont été répartis en quatre groupes de visionnage. Onze sessions d'écoute ont eu lieu et chacune d'elle durait en moyenne 30'00'' selon le temps de réponse des sujets.

V – Enregistrements et préparation des extraits

5.1 – Conception d'un dispositif d'enregistrement sonore multicanal

Nous avons développé un dispositif d'enregistrement multicanal spécifique à l'expérience. Nous avons besoin de pouvoir comparer plusieurs formats, sans avoir à réaliser de mixage de niveau ou faire des choix de spatialisation en post-production. Cela n'aurait pas permis de valider écologiquement le test car ces choix esthétiques seraient alors devenus une variable. Nous souhaitons également proposer un dispositif pratique en post-production. Nous avons ainsi écarté les systèmes incohérents dans leur répartition des canaux vis-à-vis du mixage cinéma (ou devant être utilisé en mode objet pour devenir compatibles) tel que l'ORTF-3D et l'OCT-3D.

Nous avons alors choisi de créer notre propre système de prise de son mettant en jeu un microphone par canal : L, C, R, Lss, Rss, Lrs, Lrs, Lts, Rts. Nous utiliserons cette nomenclature pour nommer les différents microphones. Grâce à ce système, il nous suffit au mixage de déterminer quel canal ouvrir pour diffuser les extraits, selon le format de diffusion souhaité.

Comme évoqué en 3.2.1 – *Procédé de développement à partir d'angles utiles*, une méthode de développement de ce système à neuf microphones est proposée par Williams (2008). Selon sa démarche, mettre dos à dos deux dispositifs 5.0 et 4.0 permet de réaliser un système 7.0. Nous avons expérimenté cette démarche, en commençant par prendre en compte les deux microphones d'élévation et leur influence potentielle sur la disposition du système.

5.1.1 – Directivité et influence des microphones d'élévation

N'ayant pas eu accès à un système 7.0.2 avant le premier jour de l'expérience, nous n'avons pas pu mener de comparatifs d'écoute pour définir le placement de ces microphones. Nous avons donc demandé conseil aux professionnels de la société HAL Audio qui, dans leur recherche de développement de l'*arbre Atmos*, se sont posé ces mêmes questions. Nous avons pris en compte leurs indications présentées en page 49 et avons choisi d'utiliser des capsules cardioïdes pour l'ensemble des microphones. Nous avons aussi déterminé une distance de 110 cm comme différence de hauteur entre la couche *top* et la couche 7.0. Lts et Rts seront dirigés vers le ciel avec une distance entre eux fixée à 70 cm.

5.1.2 – Réalisation du dispositif

Pour définir le plan 5.0 à utiliser, nous avons d'abord choisi un angle utile de prise de son pour une configuration LCR. Nous avons réalisé des comparatifs d'écoute d'ambiances urbaines dans un auditorium de mixage pour des angles utiles frontaux allant de $\pm 30^\circ$ à $\pm 70^\circ$. Nous avons défini un angle frontal de $\pm 50^\circ$ et avons répété l'opération pour déterminer les angles utiles latéraux des couples L/Ls et R/Rs. Nous avons alors défini deux angles latéraux de 50° .

Le plan de dispositif 5.0 obtenu est celui présenté en Figure 11. Une fois ce plan déterminé nous avons choisi le plan 4.0 complémentaire, présenté dans cette même figure, selon les données de Williams (2008). Ces deux plans peuvent être utilisés ensemble car la somme de leur FSSC est égale à 360° et la distance entre les microphones Ls et Rs de chaque plan est relativement faible. La Figure 18 représente ainsi le plan final de notre dispositif d'enregistrement multicanal 7.0.2.

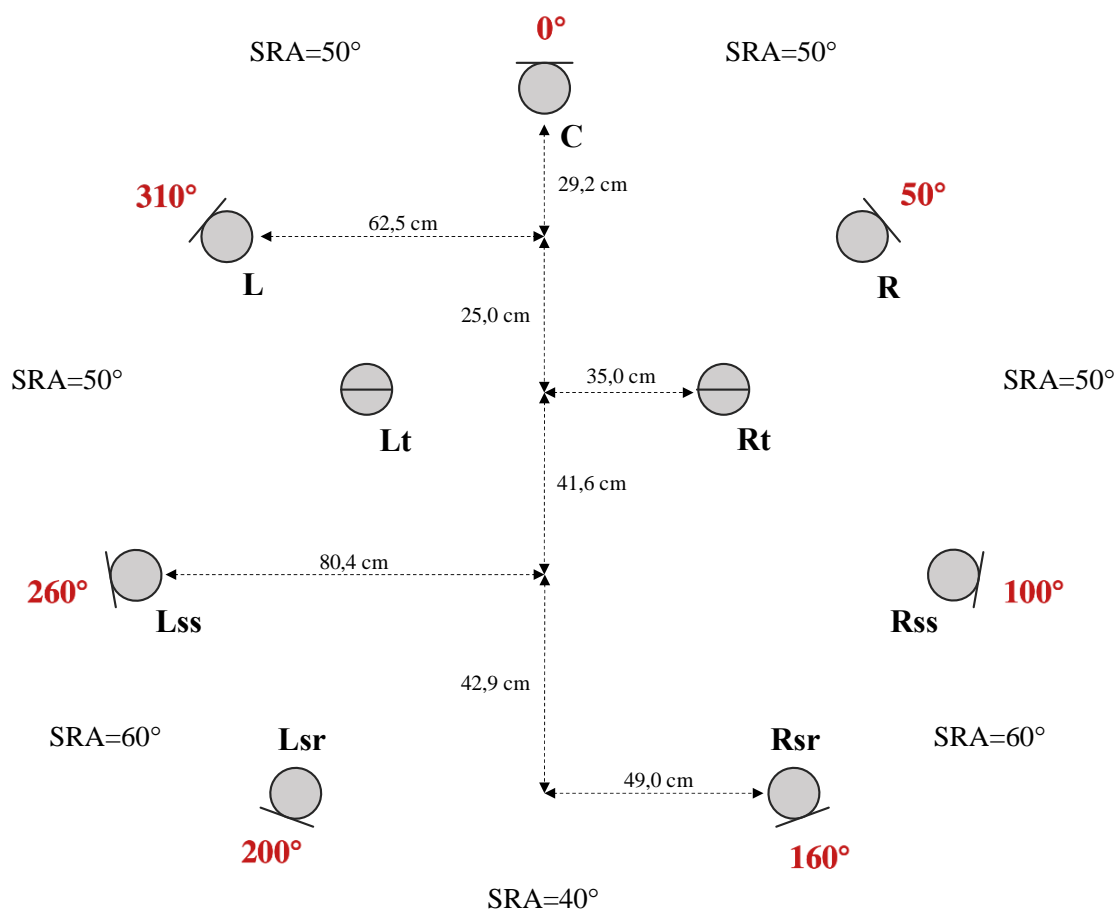


Figure 18 : Plan de positionnement des microphones pour le dispositif d'enregistrement 7.0.2 de l'expérience

5.1.3 – Équipements utilisés

Les neuf microphones utilisés pour la mise en place du dispositif étaient des Schoeps CMC6 + MK4. Ils ont été recouverts de bonnettes pour les prises de son en extérieur. L'enregistrement s'est fait sur un Sound Devices Scorpio permettant de réunir l'ensemble des microphones sur un même appareil. La liste complète du matériel utilisé est disponible en Annexe C. À ce dispositif d'enregistrement sonore, nous avons joint un dispositif de prise de vue constitué d'une caméra sur pied. Nous avons utilisé un boîtier timecode externe Tentacle Sync E pour synchroniser les deux systèmes. Les paramètres d'enregistrement de la caméra ont été adaptés pour une diffusion sur écran large :

Modèle	Résolution	Fréquence d'image	Codec	Gamma
Caméra Sony PXW-Z280	3840x2160 (UHD)	25P	XAVC Long GOP	HG3

Figure 19 : Tableau des paramètres principaux configurés dans la caméra

La durée de mise en place complète du dispositif était comprise entre 30'00'' et 1h00 suivant le nombre de personnes présentes pour l'installation et les réglages nécessaires en préparation de chaque enregistrement. Pour la majorité d'entre eux nous avons été deux personnes à installer et préparer les équipements. Deux photographies de notre dispositif sont présentées en Figure 20 et quatre autres en Annexe D.



Figure 20 : Photographies du dispositif 7.0.2 pour l'enregistrement des ambiances « École primaire » et « Forêt de pins »

5.2 – Création des stimuli et méthode de diffusion

5.2.1 – Sélection et enregistrement des ambiances

Lors de l'élaboration de l'expérience, nous avons eu comme premier objectif de définir un corpus représentatif d'ambiances utilisées au cinéma. Cet objectif fut limité par la pluralité de leurs usages ainsi que leur caractère narratif. Nous nous sommes donc appuyés sur les travaux de Adjiman et Denizart (2015), de Chion (2017), ainsi que sur notre propre expérience professionnelle pour définir un ensemble d'ambiances caractéristiques à enregistrer et à proposer lors du test :

- les rumeurs urbaines : parc ; trafic ; travaux
- les présences humaines : école, foule, marché
- les sons de la nature : vent ; mer ; pluie ; oiseaux
- les ambiances intérieures : fond d'air d'appartement, salle des machines.

Au total, dix enregistrements ont été réalisés sur un temps de cinq jours. Leur durée s'étalait de 13'00'' à 45'00'' en fonction des sources que nous souhaitions obtenir. Pour des raisons de temps et d'autorisation, certaines des ambiances présentées ci-dessus n'ont pas pu être captées.

Nous avons sélectionné huit des dix enregistrements pour créer les extraits diffusés lors de l'expérience. Nous avons en effet souhaité réduire le temps de celle-ci afin de limiter la fatigue du spectateur qui peut constituer un biais dans ses réponses. L'ensemble des extraits sont présentés dans la prochaine partie.

5.2.2 – Création des extraits

Lors de l'étape de dérushage, notre objectif a été de prélever pour chaque enregistrement un extrait contenant image et son synchrones d'une durée de 1'30''. Cette dernière fût déterminée en amont pour permettre aux sujets de prendre le temps d'analyser ce qu'ils écoutent tout en ayant la possibilité de réaliser plusieurs écoutes successives.

Nous avons souhaité faire le choix des extraits en fonction de la qualité et de la pertinence des enregistrements sonores, pour ensuite synchroniser l'image correspondante.

Pour ce faire, nous avons d'abord écouté l'ensemble des enregistrements en notant les passages qui nous paraissaient pertinents. Nous avons rapidement pris conscience qu'il était nécessaire de réaliser des points de montage sur certains extraits afin d'arriver à la durée souhaitée. Ces points étaient par exemple dus à la présence de bruits parasitant l'écoute (interférences, problème de connecteur) ou liés aux sources sonores elles-mêmes (passants proches des microphones, moteur de voiture pour un enregistrement de vent, etc.). Certains de ces points pouvaient entraîner une désynchronisation avec l'image. Nous avons alors évité de faire du montage son sur des enregistrements où les sources sonores étaient identifiables visuellement, comme par exemple des vagues qui s'échouent sur la plage ou des enfants qui jouent dans une cour de récréation.

Sur certains extraits, nous avons utilisé en plus un logiciel spécialisé dans la réduction de bruit pour pouvoir discerner certaines sources comme par exemple des habitants en intérieur qui discutent dans un appartement voisin ou la présence d'insectes et d'oiseaux.

Une fois les huit extraits réalisés, nous avons harmonisé leur niveau et les avons traités fréquemment en auditorium de mixage. Nous avons utilisé un plug-in d'égalisation pour minimiser les effets liés au dispositif d'enregistrement lui-même et que cela ne constitue pas un biais lors de l'écoute. Nous avons notamment utilisé un filtre passe-haut pour réduire certaines basses-fréquences liées au vent sur les capsules ainsi qu'un *shelf* pour réduire la sur-présentation des fréquences aiguës, trop agressives en auditorium.

Une fois les extraits mixés, nous y avons joint leurs images correspondantes à l'aide d'un logiciel d'édition vidéo. Pour les extraits synchrones nous avons utilisé les informations de timecode comme point de repère. Pour les images dont la bande sonore fut l'objet d'un montage son, nous avons choisi librement un passage dans la vidéo en faisant attention à ce qu'aucune source visuelle ne rentre en conflit avec ce qui était donné à entendre.

L'ensemble des extraits ainsi que les sources sonores les constituant sont présentés en Figure 21. A chaque extrait, nous avons exporté un photogramme représentant le lieu capté. Ces derniers sont proposés en Annexe E.

Titre de l'extrait (par ordre chronologique d'enregistrement)	Lieu	Description	Sources sonores
Jardin public	Cahors, Lot.	Parc ; grands arbres ; bâtiments au fond ; homme assis sur un banc au premier plan	Oiseaux chantants ; passantes qui marchent et discutent, passage de voitures au loin, jeunes qui parlent fort
École primaire	Trespoux- Rassiel, Lot.	Cours de récréation ; enfants qui jouent ; aire de pelouse au premier plan	Cris d'enfants qui jouent ; voix portée des animatrices ; coups et rebonds de ballons
Champ	Trespoux- Rassiel, Lot.	Champ ; coucher de soleil ; forêt et collines en fond	Cloches d'églises au loin ; vent léger ; oiseaux proches
Plage océane	Lacanau, Gironde.	Océan calme ; vagues qui s'échouent sur la plage ; ciel nuageux	Rouleaux ; ambiance continue de l'océan
Quartier en travaux	Saint-Denis, Seine-Saint- Denis.	Immeubles en fin de construction ; grues ; arbres au premier plan	Rumeur de trafic ; outils de travaux ; cris d'ouvriers ; avertisseurs de véhicules de travaux ; klaxons ; avion ; oiseaux
Intérieur appartement	Cachan, Val-de-Marne.	Appartement ; volets fermés ; canapé ; fauteuils et table basse	Fond d'air ; clé qui s'insère dans la serrure ; voisins de l'immeuble et passants dehors qui discutent
Forêt de pins	Fontainebleau, Seine-et- Marne.	Pins espacés ; rochers et buissons au sol ; ciel ensoleillé	Oiseaux ; chouettes ; insectes ; vent léger ; voitures au loin
Rumeur urbaine	Saint-Denis, Seine-Saint- Denis.	Vue sur les toits d'un quartier ; Tour Pleyel ; bâtiments en travaux ; grues	Rumeur de ville ; bruits de travaux au loin ; sirène de pompiers ; quelques chants d'oiseaux

Figure 21 : Tableau descriptif des extraits audiovisuels retenus

même ambiance dans plusieurs formats. Au lieu de cela, il écoute l'ensemble des extraits une seule fois, dans des formats différents. Quatre groupes de visionnage ont été créés pour que chaque sujet puisse écouter les huit extraits.

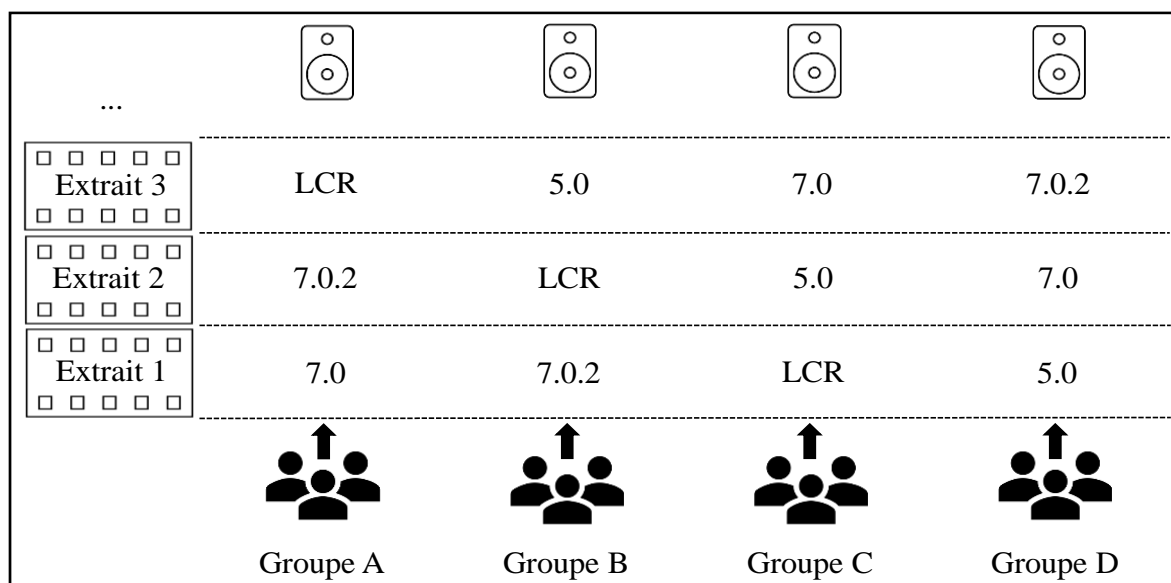


Figure 23 : Schéma récapitulatif de la répartition des sujets et des extraits par groupe

Nous avons réparti les trente-deux extraits en quatre groupes d'écoute (GR A, GR B, GR C ou GR D) de la manière suivante :

	Extrait n°1	Extrait n°2	Extrait n°3	Extrait n°4
	Quartier en travaux	Forêt de pins	Intérieur appartement	Plage océane
GR A	7.0	5.0	7.0.2	LCR
GR B	7.0.2	7.0	LCR	5.0
GR C	LCR	7.0.2	5.0	7.0
GR D	5.0	LCR	7.0	7.0.2

	Extrait n°5	Extrait n°6	Extrait n°7	Extrait n°8
	École maternelle	Rumeur urbaine	Jardin public	Champ
GR A	5.0	LCR	7.0.2	7.0
GR B	7.0	5.0	LCR	7.0.2
GR C	7.0.2	7.0	5.0	LCR
GR D	LCR	7.0.2	7.0	5.0

Figure 24 : Tableau de répartition des formats de diffusion par groupe, selon les extraits

Lors de la préparation de l'expérience, nous n'avons pas identifié de biais d'ordre concernant la diffusion des huit enregistrements. Nous avons donc déterminé en amont et de manière aléatoire un ordre fixe pour les quatre groupes. Concernant la distribution des formats de diffusion au sein d'un groupe, nous avons réalisé un contrebalancement nous permettant de ne pas avoir d'effet d'ordre dans nos résultats. Par exemple, pour le groupe A, nous avons réparti les formats de diffusion entre les extraits selon deux conditions :

- la première, ne pas avoir deux mêmes formats successifs
- la seconde, avoir chaque format représenté deux fois.

Pour les groupes B, C et D, nous avons utilisé le groupe A comme référence et décalé les formats suivant la règle LCR \rightarrow 5.0 \rightarrow 7.0 \rightarrow 7.0.2.

5.3 – Caractéristiques techniques de l'environnement de l'expérience

L'expérience s'est déroulée dans un auditorium de mixage de la société parisienne Poly Son Post Production. Le système de diffusion mettait en jeu une enceinte par canal, à l'exception des canaux zénithaux où quatre enceintes étaient positionnées au plafond. Cette configuration est présentée en Figure 25.

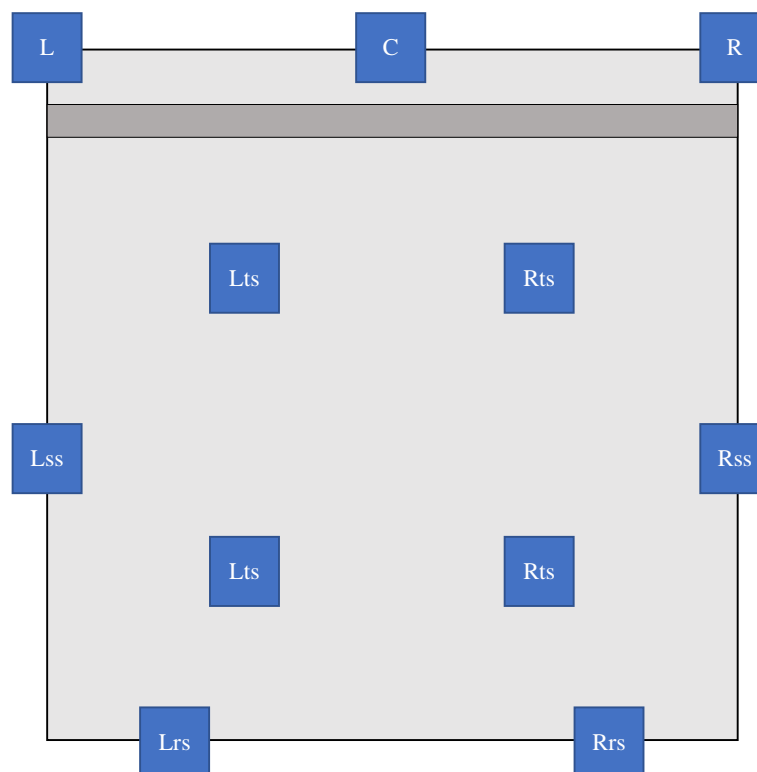


Figure 25 : Schéma représentant le positionnement des enceintes de l'auditorium et l'affectation des canaux

La lecture des extraits s'est faite à travers une session Pro Tools dans laquelle nous avons configuré seize pistes audios (quatre par groupe et par format) et une piste vidéo. Lors de chaque session d'écoute, les quatre pistes correspondant au groupe souhaité étaient activées, les douze autres étaient désactivées.

Chaque piste était affectée en sortie au *Rendering Mastering Unit*, processeur du système Dolby Atmos qui jouait le rôle d'interface avec les amplificateurs.



Figure 26 : Capture d'écran de la session Pro Tools utilisée pour l'expérience - mise en valeur des quatre groupes de visionnage

VI – Protocole de test

6.1 – Choix des participants et répartition par groupe

L'expérience de l'immersion spatiale au cinéma touche un public large qui n'a pas besoin de prérequis concernant l'analyse sonore et filmique pour expérimenter ces sensations. En partant de ce postulat, nous n'avons eu aucun critère de sélection concernant le choix des sujets et avons ouvert le test au grand public.

La prise de contact s'est faite dans un premier temps à partir de l'établissement d'une base de données de potentiels sujets volontaires. Celle-ci fut alimentée tout au long de notre recherche à la suite de présentations et de discussions autour de notre présente étude, notamment aux élèves et enseignants de l'ENS Louis-Lumière. Puis, nous avons ouvert la participation au test par internet à travers des listes de diffusions internes et externes à l'établissement.

L'expérience s'est déroulée sur deux journées complètes. Au total, onze sessions d'écoute ont été réalisées, à raison de une par heure pour la première journée, puis en fonction de la disponibilité des sujets la deuxième journée.

Compte tenu de la dimension de l'auditorium, nous avons fixé lors d'une séance de pré-test à cinq le nombre maximum de sujets pouvant passer le test simultanément. Au total, vingt-huit sujets (ID1->ID28) ont participé à l'expérience.

Selon leurs disponibilités, les sujets ont été chacun affectés à un groupe de visionnage (GR A, GR B, GR C ou GR D) correspondant à une session. Nous avons essayé de les répartir au mieux, au fur et à mesure de leur venue pour se rapprocher d'un même nombre de sujets pour chaque groupe. Le détail de cette répartition est présenté en Figure 27.

À la fin de l'expérience, nous avons proposé aux sujets de remplir un questionnaire supplémentaire pour vérifier s'il existait des différences significatives entre les groupes. Nous nous sommes basés sur trois critères : l'âge, la fréquence à laquelle les sujets vont au cinéma et leur situation professionnelle (Figure 28).

Samedi 15 avril 2023							
	Session n°	Groupe	ID				
10h30 - 11h30	1	GR A	ID1	ID2	ID3		
11h30 - 12h30	2	GR B	ID4	ID5	ID6		
12h30 - 13h30	3	GR C	ID7				
14h30 - 15h30	4	GR C	ID8	ID9	ID10	ID11	
15h30 - 16h30	5	GR D	ID12	ID13	ID14	ID15	ID16
16h30 - 17h30	6	GR B	ID17	ID18			
17h30 - 18h30	7	GR A	ID19	ID20			
18h30-19h30	8	GR A	ID21	ID22			

Samedi 22 avril 2023							
	Session n°	Groupe	ID				
12h30-13h30	9	GR B	ID23	ID24	ID25	ID26	
14h-15h	10	GR C	ID27				
16h - 17h	11	GR D	ID28				

Groupe A	Groupe B	Groupe C	Groupe D	TOTAL
7 sujets	9 sujets	6 sujets	6 sujets	28 sujets

Figure 27 : Tableau du planning des sessions de test avec la répartition des sujets et leur nombre par groupe

-
- Quel âge avez-vous ?
 - Êtes-vous un·e professionnel·le du son, enseignant·e, chercheur·se, artiste, étudiant·e ou ancien·ne étudiant·e dans ce même domaine (rayer la mention inutile) : OUI / NON
 - A quelle fréquence allez-vous au cinéma :
 - au moins 1 fois par semaine
 - au moins 1 fois par mois (et moins d'une fois par semaine)
 - au moins 1 fois par an (et moins d'une fois par mois)
-

Figure 28 : Extrait du questionnaire – questions portant sur les sujets

6.2 – Élaboration du questionnaire

Après le visionnage de chaque extrait, le sujet était invité à remplir un questionnaire. Celui-ci est présenté en Annexe F. Au total, le sujet l'aura rempli huit fois, correspondant au visionnage des huit extraits. Nous avons adapté l'écriture du questionnaire pour garantir sa compréhension auprès du grand public.

Lors d'une séance de pré-test, nous avons estimé à 2'30'' le temps de réponse moyen à ce questionnaire. Cependant, nous avons laissé aux sujets tout le temps qu'ils souhaitaient pour le remplir. Il leur était également possible de poser des questions avant, pendant, ou après chaque questionnaire.

Lors de l'élaboration du test, nous n'avons pas identifié de biais d'ordre concernant les réponses aux questions. Aussi, nous avons précisé aux sujets qu'ils pouvaient répondre aux questions dans l'ordre qu'ils souhaitaient.

Le questionnaire se présente sous la forme de sept questions (Q1->Q7). Nous avons défini en amont de l'expérience six critères de mesure : l'enveloppement, la localisation, la sensation de présence, la sensation d'être captivé, la cohérence des espaces visuels et sonores et la grandeur de l'espace imaginé. Chaque critère est associé à une unique question à l'exception de l'enveloppement qui est traité dans Q1 et Q2.

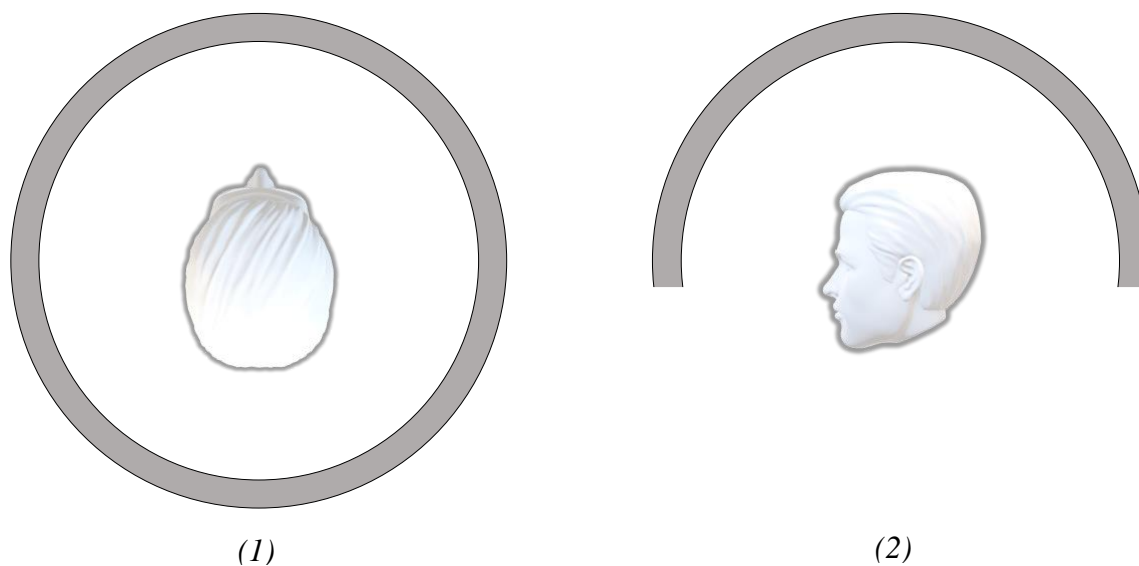
Chaque question, à l'exception de Q2, permet de recueillir des données sous la forme d'échelle de Likert. Le sujet devait entourer un chiffre entre 1 et 7 représentant les deux extrêmes d'une proposition donnée (exemple : (1) pas du tout enveloppant / (7) très enveloppant).

Nous avons choisi cette échelle de notation pour recueillir des informations relatives et convertir en nombre les réponses données par les sujets pour réaliser des statistiques lors de l'exploitation des résultats.

Voici, en suivant, une description détaillée de l'ensemble de ce questionnaire.

<i>Q1 : « Je me suis senti-e entouré-e de sons. » Êtes-vous d'accord avec cette proposition ?</i>	
<i>Pas du tout d'accord (1)</i>	<i>Tout à fait d'accord (7)</i>

Q2 : Les dessins suivants représentent une silhouette de tête, du dessus (1) et de profil (2). Indiquez, en noircissant l'espace grisé, de quelle(s) direction(s) semblaient provenir les sons que vous avez entendus :



Q1 et Q2 ont pour objectif de mesurer l'enveloppement du sujet. L'enveloppement est un critère perceptif mesuré de manière récurrente dans la littérature scientifique pour quantifier de manière globale la façon dont le spectateur se sent entouré de son. Grâce à l'analyse des données recueillies de Q2, nous souhaitons observer de manière graphique l'influence des formats de diffusion sur l'enveloppement du sujet. Si l'enveloppement peut déjà être considéré comme un indice d'immersion spatiale, ce critère nous servira aussi à observer une potentielle corrélation avec les autres questions.

Q3 : « La - ou les – source(s) sonore(s) m'ont paru venir de directions précises. » Êtes-vous d'accord avec cette proposition ?

Pas du tout d'accord (1)

Tout à fait d'accord (7)

Nous étudions dans Q3 la localisation des sources sonores. Ce paramètre est proposé dans Eaton et Lee (2019) pour mesurer l'immersion des sujets dans un environnement virtuel.

Q4 : À quelle intensité vous êtes-vous senti-e présent-e dans l'espace projeté à l'écran ?	
<i>Je ne me suis pas du tout senti-e présent-e dans cet espace (1)</i>	<i>Je me suis pleinement senti-e présent-e dans cet espace (7)</i>

Q5 : Avez-vous eu la sensation d'oublier ce qui vous entourait pendant le visionnage de cet extrait ?	
<i>J'étais conscient-e du monde réel et des éléments autour de moi (1)</i>	<i>J'ai été complètement captivé-e par l'extrait (7)</i>

Q4 est une traduction libre de la onzième question du questionnaire IEQ (Rigby *et al.*, 2019) : « To what extent could you picture yourself in the scene of the events shown in the movie, TV show, or clip ? » (p. 5)

Nous avons rédigé Q5 comme la combinaison des questions huit et neuf du questionnaire IEQ : « Dans quelle mesure étiez-vous conscient de vous-même dans votre environnement ? ... Dans quelle mesure avez-vous remarqué des événements se déroulant autour de vous ? »⁹ (p. 5)

Q4 et Q5 permettent de mesurer le sentiment de présence dans l'extrait ainsi que la focalisation de l'attention du sujet.

Q6 : Selon vous, l'espace présenté à l'image était-il cohérent avec l'espace sonore ?	
<i>Les 2 espaces n'étaient pas du tout cohérents (1)</i>	<i>Les 2 espaces étaient totalement cohérents (7)</i>

Q7 : Un *item* (par exemple une forêt) vous a été présenté à l'écran. En imaginant le hors-champ : quelle grandeur aurait ce dernier par rapport à l'idée que vous vous faites d'un *item* moyen ?	
<i>Beaucoup plus petit qu'un *item* moyen (1)</i>	<i>Nettement plus étendu qu'un *item* moyen (7)</i>

Q6 et Q7 portent sur l'interaction entre le champ visuel et le champ sonore. Plus précisément, ces questions ont été rédigées en tenant compte du statut particulier qu'occupent

⁹ « To what extent were you aware of yourself in your surroundings ? ... To what extent did you notice events taking place around you ? » (Rigby *et al.*, 2019, p. 5, notre traduction)

les ambiances dans la bande sonore. En effet, elles révèlent au son ce qui n'est pas présent à l'image (utilisation territoire). Nous cherchons, à travers ces questions, à observer la manière dont le sujet perçoit l'espace qui lui est présenté à l'image et au son, relativement aux différents formats de diffusion proposés.

6.3 – Présentation du protocole aux sujets

Afin de se familiariser avec le questionnaire et la procédure du test, nous avons proposé aux sujets de visionner un extrait « exemple » pour lequel ils n'avaient pas à répondre aux questions mais pouvaient imaginer leurs réponses et poser des questions en conséquence. De plus, nous souhaitons que cet exemple permette aux sujets de se familiariser avec le système d'écoute. L'extrait a été diffusé au format 7.0.2, cette information n'a pas été communiquée aux sujets. En préparation à leur accueil, nous avons rédigé un document, présenté ci-dessous, détaillant le protocole de l'expérience. Celui-ci a été lu à l'ensemble des sujets avant chaque début de session :

« Bonjour. Vous allez être amenés à visionner huit extraits audiovisuels, constitués de plans fixes et représentant des lieux. Les caractéristiques audios de ces extraits varieront au cours de l'expérience.

À la fin de chaque extrait, vous serez invités à remplir un questionnaire constitué de sept questions. Pour la majorité d'entre elles, il suffira d'entourer un chiffre entre un et sept selon une échelle proposée. Nous allons lire ensemble ces questions par la suite.

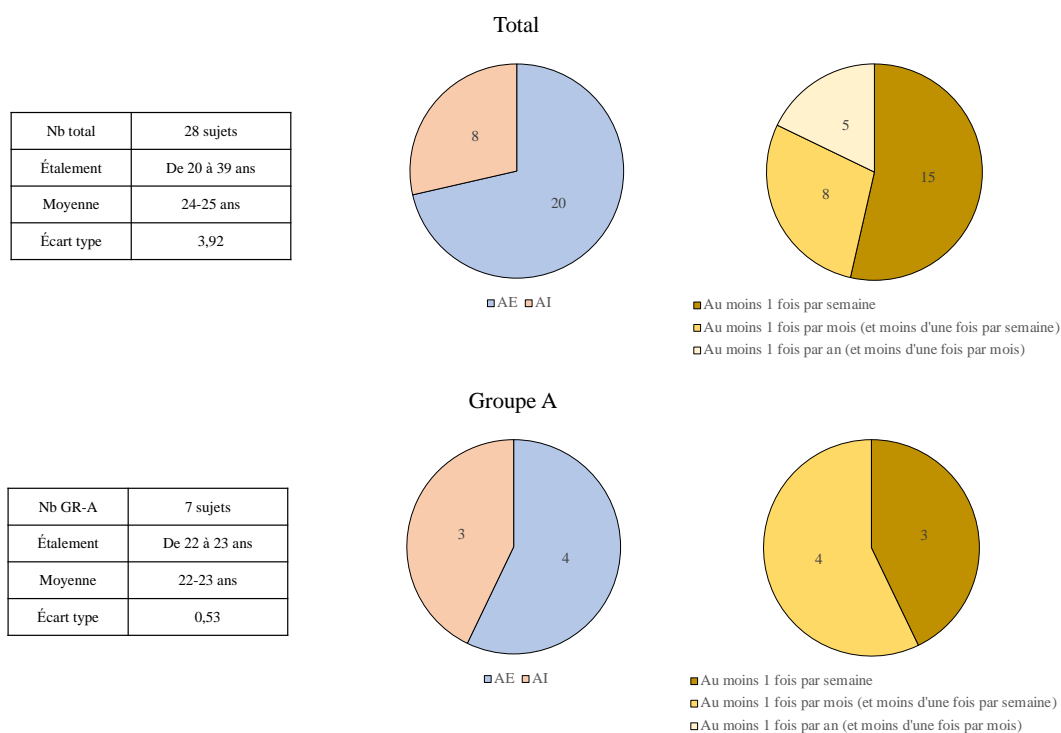
*À l'écran, avant chaque extrait, il sera inscrit un carton titre sur lequel est écrit le numéro et le nom de l'extrait correspondant pour vous repérer dans le questionnaire. Attention, ce-dernier est recto-verso, les numéros de pages sont inscrits en bas à gauche pour vous aider à vous repérer. Chaque extrait dure 1'30'' et vous aurez tout le temps nécessaire pour répondre à chaque questionnaire, avant le lancement du prochain extrait. Vous pouvez répondre aux questions dans l'ordre que vous voulez. À la fin de l'expérience, il y aura un dernier court questionnaire à remplir sur votre identité à des fins statistiques. Le questionnaire est anonyme. Je vous propose que l'on passe à la lecture des questions, puis nous passerons aux questions si vous en avez. *Lecture du questionnaire et réponse aux questions éventuelles*. Je vous propose de visionner un extrait-exemple pour vous familiariser avec le système d'écoute de l'auditorium. Vous n'aurez à répondre à aucune question pour celui-ci. *Diffusion de l'extrait-exemple et début de l'expérience*. »*

VII – Exploitation des résultats

7.1 – Statistiques de l'échantillon

Vingt-huit sujets ont participé à l'expérience et ont été répartis en quatre groupes de visionnage. Nous avons défini trois critères pour vérifier s'il existait des différences significatives entre les groupes : l'âge, la fréquence à laquelle les sujets vont au cinéma et si leur profession s'inscrivait dans le domaine du son. Pour ce dernier critère, nous considérons comme « auditeur expérimenté » (AE) un sujet professionnel, chercheur, enseignant ou étudiant en son. Les autres sujets sont considérés dans cette étude comme « auditeurs inexpérimentés » (AI). L'ensemble de ces informations sont résumées en Figure 29.

Plus de la moitié des sujets ayant participé au test se rendent au cinéma au moins une fois par semaine. Cette population est relativement jeune, avec une moyenne d'âge de 24-25 ans, qui pour la plupart sont des auditeurs expérimentés. Cette répartition est particulièrement variable en fonction des groupes en raison du faible nombre de sujets répartis dans ces derniers. Nous considérons ces groupes comme homogènes pour réaliser un ensemble de statistiques descriptives basées sur les réponses au questionnaire. Il sera cependant nécessaire de prendre en compte le nombre réduit de sujets ayant participé à l'expérience, ce qui peut entraîner de grandes variations dans l'analyse des résultats.



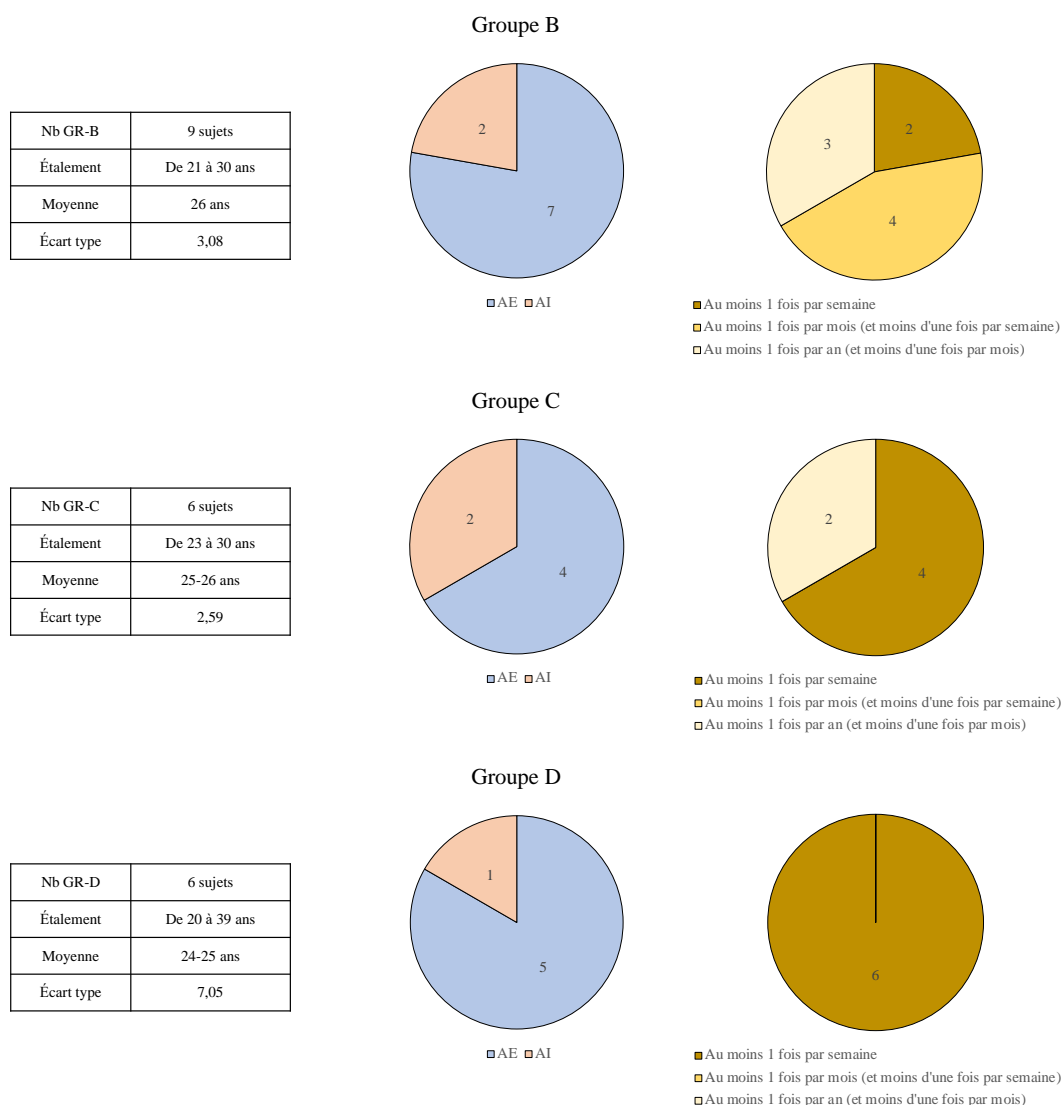


Figure 29 : Tableaux et graphiques en camembert représentant les caractéristiques des sujets par groupe, selon l'âge, la fréquentation des salles de cinéma et la condition socio-professionnelle

7.2 – Méthode d'analyse retenue

Nous avons choisi pour l'analyse de rassembler les résultats de l'ensemble des sujets par question. Cinquante-six valeurs par format ont été comptabilisées étant donné que chaque sujet a écouté deux fois l'ensemble des formats de diffusion. Dans un premier temps, les différences entre les extraits ne seront pas prises en compte.

Nous allons analyser les résultats par question en utilisant un graphique de type Box plot, ou « boîte à moustaches » car il nous permet de mettre en valeur la répartition des résultats d'une même plage de donnée de manière graphique ainsi que de potentielles valeurs aberrantes (Figure 30). Rappelons que six des sept questions proposées utilisaient comme forme de réponse une échelle de Likert 1-7.

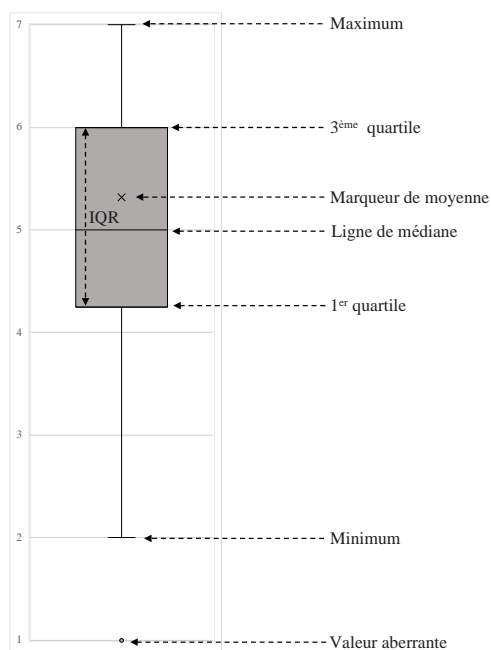


Figure 30 : Description des éléments constitutifs du Box plot

Le Box plot permet de résumer les principaux critères d'analyse descriptive sous forme d'une boîte encadrée par deux moustaches. Au-delà de la présence de la moyenne et de la médiane, la longueur de la boîte correspond à l'écart interquartile (IQR), c'est-à-dire que 50% de données recueillies se trouvent à l'intérieur de la boîte. Les valeurs minimum et maximum des moustaches peuvent s'étendre jusqu'à $1,5 * IQR$ ou sont délimitées en amont par les potentielles valeurs extrêmes de la plage de données analysée. Si une valeur est trop éloignée, elle est dite aberrante et se trouve matérialisée par un point en dehors de l'encadrement des moustaches.

En complément, nous avons procédé pour chacune des questions à un test de significativité en réalisant une analyse de variance (ANOVA). Nous considérons que si notre valeur de probabilité *p-value* obtenue est inférieure à 0,05 nous pouvons rejeter l'hypothèse nulle et accepter qu'il existe des différences statistiquement significatives dans les résultats obtenus.

La méthode d'analyse de Q2 est particulière. Nous souhaitons à travers cette question que le sujet puisse exprimer graphiquement sa sensation d'enveloppement. Deux dessins lui ont été proposés : l'un présentant une tête vue du dessus et l'autre de profil. L'objectif était qu'il représente dans une couronne entourant ces têtes le champ sonore perçu. Nous avons délibérément laissé une libre interprétation à cette question, les sujets pouvaient notamment dessiner une largeur de champ globale ou des sources précises (deux exemples de réponses sont proposés en Annexe G). Nous leur avons cependant indiqué que la densité des traits ne serait pas prise en compte dans l'analyse. Il était donc inutile de noircir certaines zones de manière plus prononcées que d'autres.

Pour l'analyse de cette question, nous avons choisi de réaliser une heatmap (ou carte thermique) radiale. Il s'agit dans notre cas d'un graphique prenant la forme d'une couronne similaire à la question, qui traduit par des nuances de couleurs l'intensité d'une variable. Cette dernière

correspond au nombre de fois où une zone de la couronne aura été noircie. Nous avons pour cela divisé chaque couronne en quatre-vingt-seize cases (Annexe G) et avons compté informatiquement le nombre de fois où les sujets avaient dessiné par-dessus. Le nuancier de couleurs utilisé pour la heatmap est présenté en Figure 31.

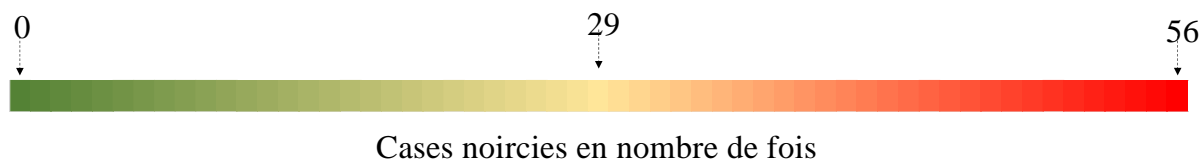


Figure 31 : Nuancier de couleurs utilisé pour la création des heatmaps de Q2

7.3 – Analyse par question

7.3.1 – Q1 : l'enveloppement (réponse par échelle de Likert)

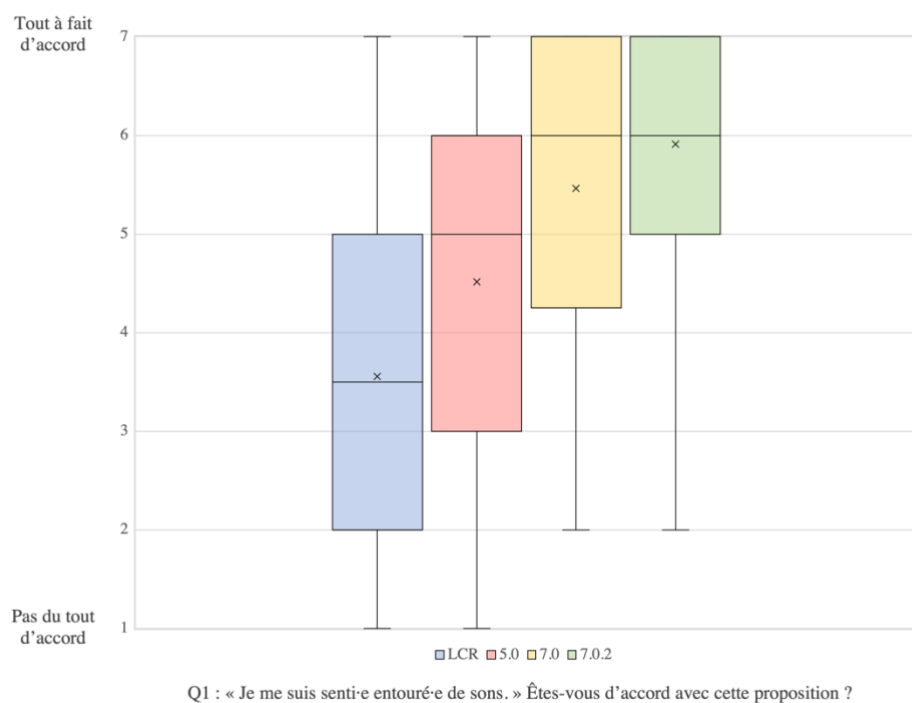


Figure 32 : Box plot de Q1 par format

On observe dans la Figure 32 une notation plus grande de la sensation d'enveloppement en fonction de l'augmentation du nombre de canaux. La moyenne la plus basse est de 3,55 pour le format LCR et celle la plus haute est de 5,9 pour le format 7.0.2. Entre ce dernier format et le format 7.0 les médianes sont identiques mais l'IQR est plus resserré pour le format 7.0.2. Cela traduit un plus grand nombre de réponses condensées dans les notes les plus élevées. L'analyse de variance ($p\text{-value} = 1,53e^{-12} < 0,05$) indique qu'il existe des différences nettement

significatives entre ces résultats, au-delà du niveau de chance. On en déduit que l'augmentation des canaux autour du spectateur est corrélée à une sensation d'enveloppement plus élevée.

7.3.2 – Q2 : l'enveloppement (réponse par dessin)

L'ensemble des heatmaps correspondant à Q2 sont présentées en Figure 33. Nous observons sur la première, correspondant à l'axe de diffusion horizontale du format LCR (en haut à gauche), que les représentations des sujets mettent en valeur un fort effet de frontalité. En effet, les cases dans l'axe de vision ont été noircies jusqu'à cinquante-cinq fois. Plus de la moitié des réponses concernant l'enveloppement ont été matérialisées dans un angle de $\pm 70^\circ$ autour du sujet pour ce format.

En augmentant le nombre de canaux autour du spectateur, cet angle global s'élargit. Il est environ égal à $\pm 100^\circ$ pour le format 5.0 et à $\pm 140^\circ$ pour le format 7.0. Nous constatons pour le dernier format 7.0.2 une répartition uniforme des réponses tout autour du sujet et sans discontinuité. Les réponses individuelles sont alors plus variables, ce qui explique une teinte de rouge moins marquée. En réalisant une moyenne, chaque case a été noircie quarante-deux fois. La différence entre ce graphique et celui correspondant au format 7.0.2 vient de la présence des canaux d'élévation. On constate que leur utilisation tend à rendre plus homogène le champ sonore autour des sujets.

Nous retrouvons dans le graphique représentant l'axe vertical de diffusion (à droite) du format LCR ce même effet frontal. Plus de la moitié des réponses inscrites par les sujets sont comprises dans un angle de 60° au-dessus du sujet par rapport à l'axe horizontal, correspondant à l'emplacement des enceintes derrière l'écran. Cet angle est égal à 55° pour le format 5.0 et à 80° pour le format 7.0. En théorie, nous pouvons questionner la capacité de ces formats n'utilisant pas de canaux d'élévation à procurer une telle sensation de hauteur dans le champ sonore. Nous expliquons ces résultats par le petit volume de l'auditorium dans lequel s'est déroulée l'expérience. Une conséquence directe est la proximité entre les sujets et les enceintes positionnées en hauteur. Sur les graphiques, cela se traduit par une plus grande uniformisation au fur et à mesure que l'on augmente le nombre de canaux. Pour le format 7.0.2, on remarque à l'arrière de la tête virtuelle une zone plus orangée que l'on pourrait ainsi attribuer au placement des enceintes physiques au plafond.

Nous pouvons à travers ces graphiques comprendre de manière plus imagée la sensation d'enveloppement chez les sujets. Celle-ci est plus intense quand les formats de diffusion ont un nombre de canaux élevés. La présence et l'utilisation des enceintes zénithales est un indice potentiel d'homogénéisation du champ sonore autour des sujets, relativement à la nature des extraits présentés.

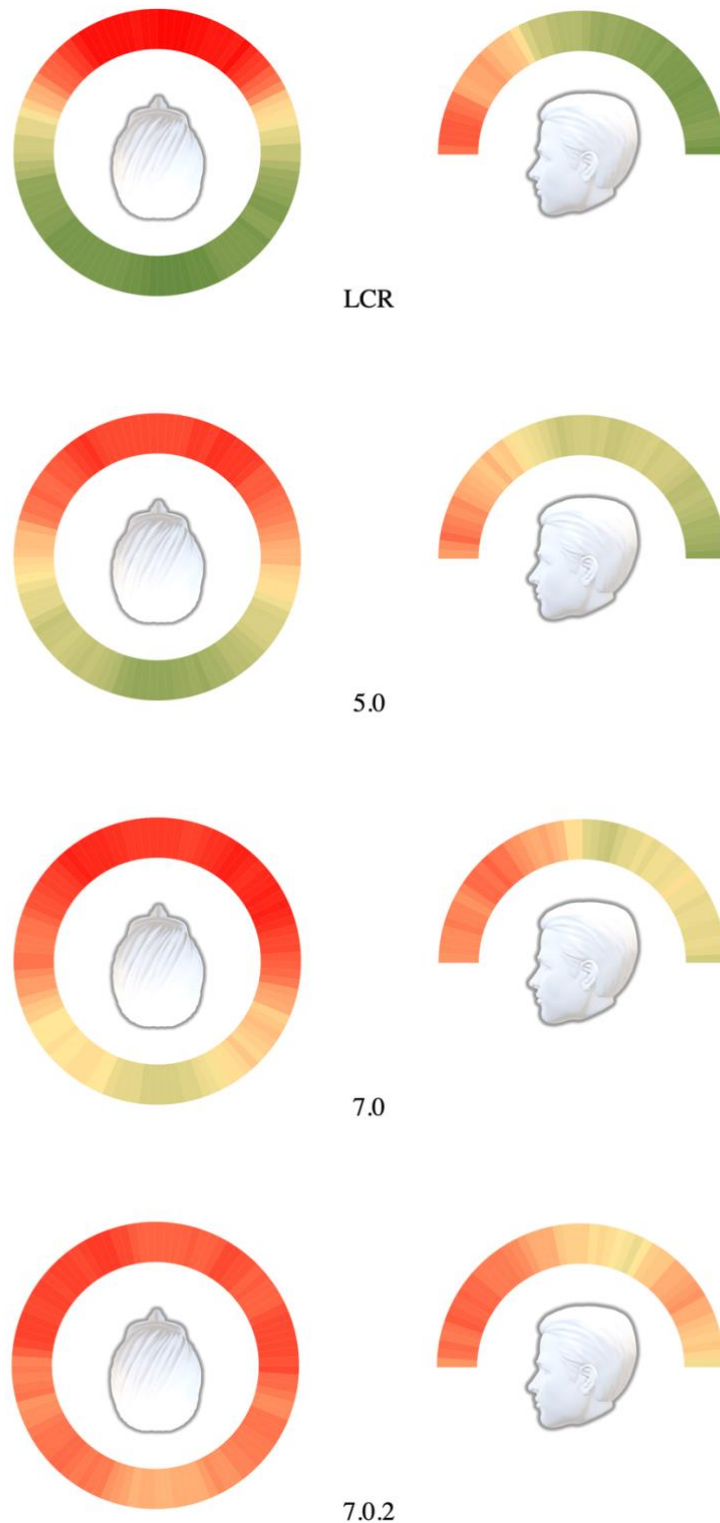


Figure 33 : Heatmaps de Q2 par format

7.3.3 – Q3 : la localisation

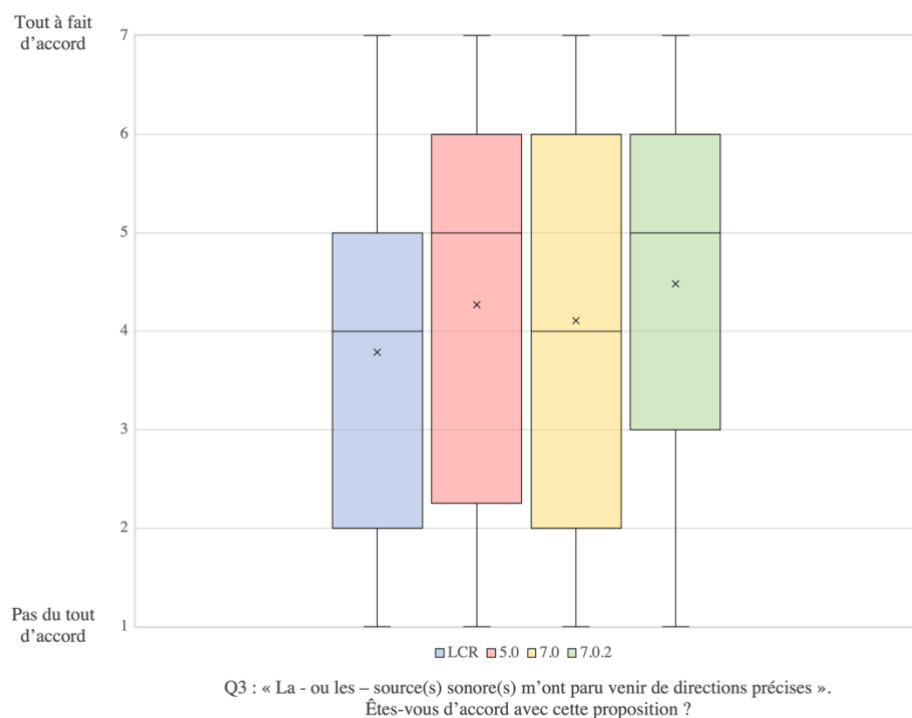


Figure 34 : Box plot de Q3 par format

Nous n'observons pas de différence notable entre les formats au regard du critère de localisation. Si la moyenne la plus basse reste toujours celle du format LCR (3,79) et la plus haute pour le format 7.0.2 (4,48), on constate une grande répartition des résultats quel que soit le format. Les médianes sont fluctuantes et les IQR larges.

Notre ANOVA ($p\text{-value} = 0,27 > 0,05$) ne permet pas de rejeter l'hypothèse nulle ; il n'existe donc pas de différences significatives entre ces formats. Dans notre expérience, le critère de la localisation des sources n'est donc pas lié à la configuration du format d'écoute. Nous expliquons ces résultats par la nature des extraits présentés qui, pour la plupart, ne comportaient pas ou peu de sources ponctuelles que le spectateur pouvait détacher du fond ambiant.

7.3.4 – Q4 : la présence

Nous observons pour le critère de présence une nette différence entre les paires de formats LCR, 5.0 et 7.0, 7.0.2. Pour ces deux derniers formats, plus de la moitié des sujets ont attribué une note de 5 ou plus à la sensation d'être projeté dans le lieu présenté à l'écran. À

l'inverse, les résultats pour les formats LCR et 5.0 sont similaires. Les longueurs de boîtes sont larges (IQR = 3) ce qui traduit une grande variation dans les réponses données. Nous précisons pour le format 7.0 que la médiane est de 5 et se confond donc avec le 1^{er} quartile. Une ANOVA ($p\text{-value} = 0,024 < 0,05$) indique une différence statistiquement significative dans les résultats obtenus. Avec cette analyse descriptive et de variance, nous pouvons déduire que la présence, sensation perçue lors d'une immersion sensorielle dans un espace projeté, est accentuée lors d'une diffusion 7.0 et 7.0.2.

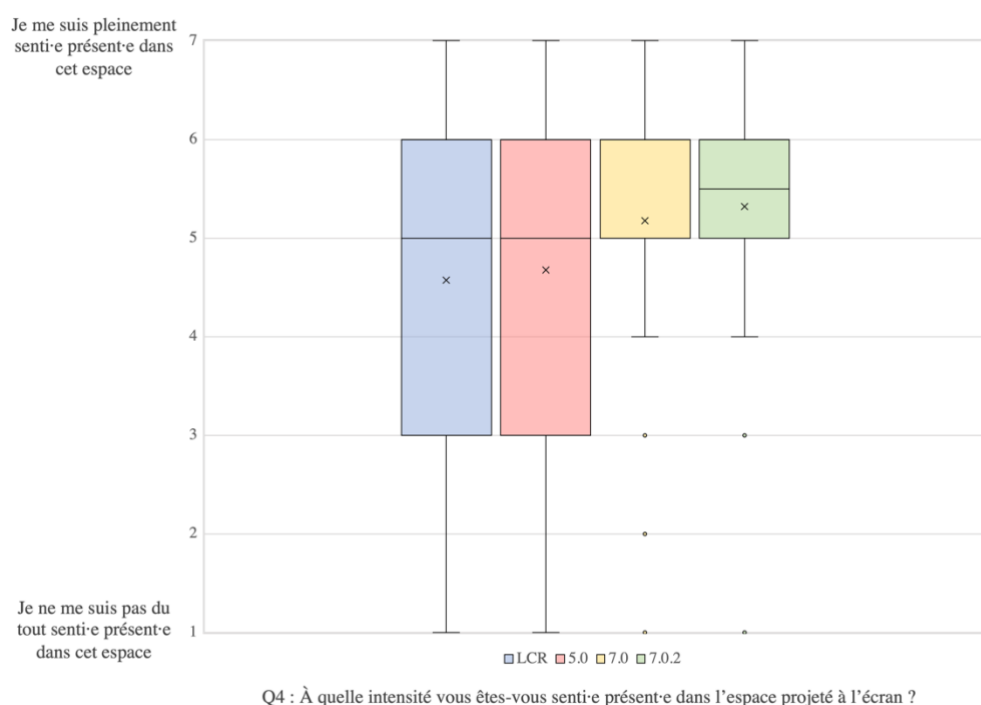


Figure 35 : Box plot de Q4 par format

7.3.5 – Synthèse de Q5, Q6 et Q7

Ces trois questions ont été réunies car il est possible d'en faire une analyse commune. Q5 portait sur la sensation de se sentir captivé, Q6 sur la cohérence image-son et Q7 sur l'imagination du hors-champ visuel et sonore. Les résultats obtenus sont résumés dans les Figure 36, Figure 37 et Figure 38.

De manière générale, nous n'observons pas de différence significative dans les réponses des sujets entre les différents formats. Les moyennes sont resserrées (écart de moyenne maximal de 0,68 pour la Q7 entre le format LCR et 7.0.2) et la répartition des données plutôt similaire, surtout pour Q6 où les IQR pour chaque format sont égaux.

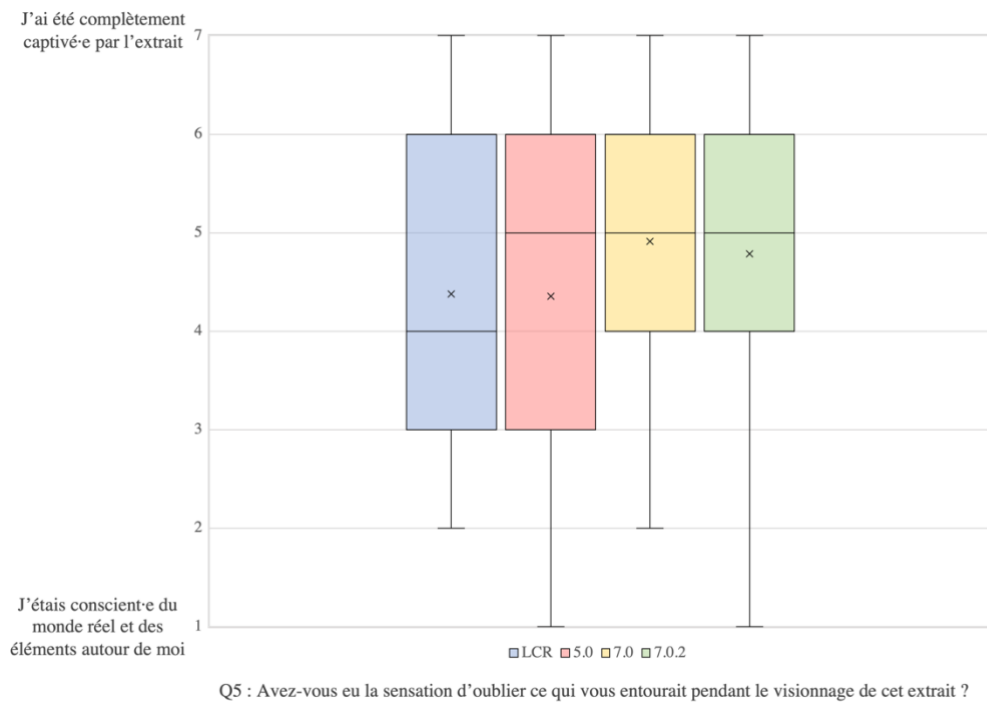


Figure 36 : Box plot de Q5 par format

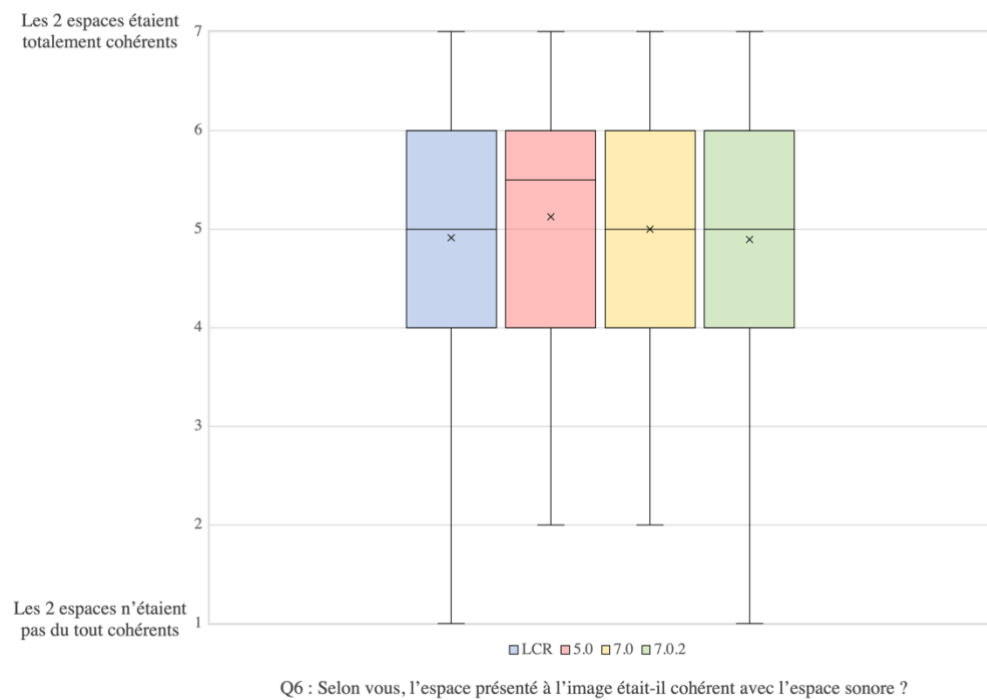


Figure 37 : Box plot de Q6 par format

Pour Q7, nous observons graphiquement une légère tendance des sujets à s'imaginer le lieu présenté à l'écran comme plus étendu au fur et à mesure que l'on augmente le nombre de canaux. Les moyennes sont croissantes et on observe pour les formats 7.0 et 7.0.2 un plus grand nombre de réponses élevées. Les résultats des trois ANOVA ne permettent pas de rejeter l'hypothèse nulle : Q5 : $p\text{-value} = 0,15 > 0,05$; Q6 : $p\text{-value} = 0,87 > 0,05$; Q7 : $p\text{-value} = 0,10 > 0,05$. Nous ne pouvons pas affirmer qu'il existe des différences entre les formats pour ces critères étudiés.

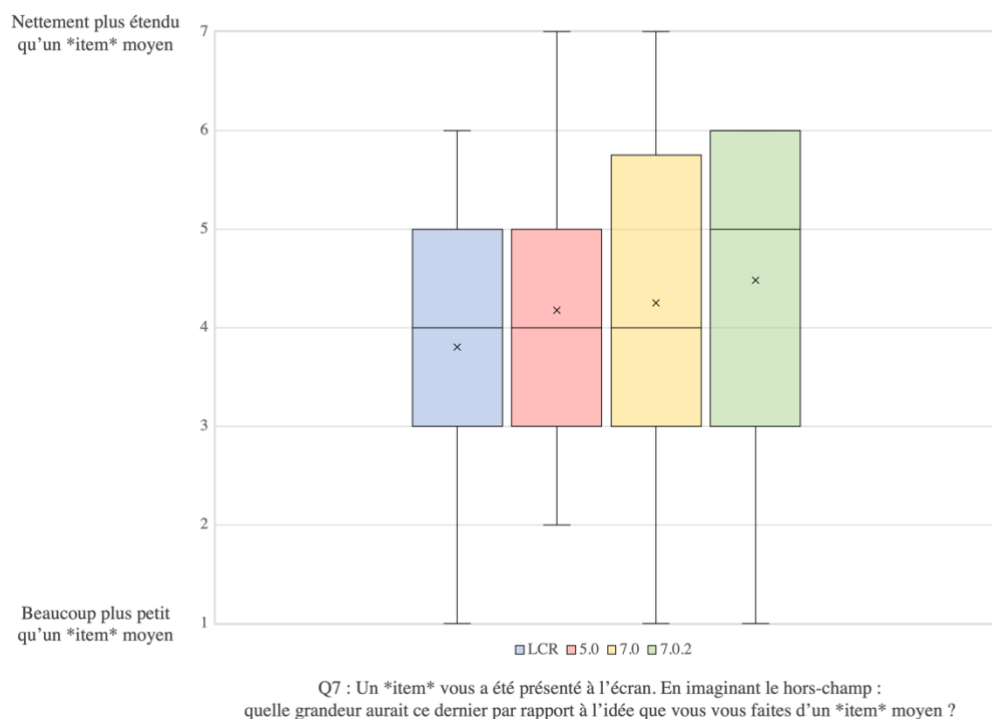


Figure 38 : Box plot de Q7 par format

7.4 – Analyse par question et par extrait

Dans cette seconde analyse, nous allons prendre en compte les potentielles différences entre les extraits qui ont été présentés. Ces derniers sont variés, par la nature de l'espace enregistré (naturel, urbain...), les sources sonores entendues (oiseaux, trafic...) ou encore leur synchronisation avec l'image (vagues qui s'échouent, enfants qui jouent...).

En nous rapportant aux travaux de Hamasaki *et al.* (2006), ainsi que de Oramus et Neubauer (2019), nous émettons l'hypothèse que la nature de ces stimuli a influencé certaines

réponses des sujets. Nous allons donc analyser un ensemble de résultats, par extrait, pour essayer de mettre en valeur les biais visuels qui auraient pu entrer en jeu durant l'expérience. Les trois premières questions ne seront pas traitées dans cette analyse car les deux premières portent spécifiquement sur l'enveloppement, un critère perceptif sonore, et nous avons estimé que la troisième, concernant la localisation, n'était pas pertinente au regard de la nature des extraits étudiés.

7.4.1 – Q4 et Q5 : la présence et la sensation de se sentir captivé

Sur les Figure 39 et Figure 40 sont regroupés les histogrammes réalisés à partir des résultats de Q4 et Q5, par format et par extrait. Pour chacun, une moyenne des réponses prenant en compte l'ensemble des formats est matérialisée par un point gris.

Les courbes de moyennes entre les deux histogrammes sont similaires. Les quatre extraits ayant les moyennes les plus élevées sont les enregistrements d'un champ, d'une école maternelle, d'un jardin public et d'une plage. Ces extraits ont comme point commun qu'ils présentent dans leur enregistrement des sources visuelles et sonores synchrones. Nous voyons les enfants jouer dans la cour de récréation, les vagues s'échouer sur la plage et des passants marcher et discuter dans le jardin public. L'enregistrement de champ (E8) présente lui une église sonnante l'angelus que l'on devine au loin. À l'inverse, les extraits E1, E2, E3 et E6 ne laissent pas paraître à l'image les sources que l'on entend.

Les sujets ont attribué au champ et au jardin public des notes élevées pour le format 7.0.2. Ces extraits présentent plusieurs plans sonores : les oiseaux proches, le vent, puis l'église pour E8 ; les passants qui marchent, des jeunes qui parlent fort, les oiseaux, puis la rumeur de ville et les passages de voitures au loin pour E7.

Une hypothèse d'explication possible concernant les différences entre ces extraits aurait pu être la présence de source sonores ponctuelles, pas nécessairement présentes à l'image. Mais nous en retrouvons dans la totalité des extraits, également dans ceux présentant des moyennes les plus basses : E1 présente différents bruits d'outils de travaux et de sirènes, dans E3 des voisins rentrent dans un appartement et des passants discutent distinctement, enfin, dans E6, une sirène de pompier retentit. Au contraire, la plage océane (E4) est particulièrement monotone, les vagues présentes à l'image sont impossibles à deviner uniquement avec le son (effet psycho-acoustique). Les sujets se sont tout de même sentis plus présents et captivés par ce dernier extrait, ce qui nous fait penser que la synchronisation, même illusoire, entre une

source visuelle et sonore est importante pour obtenir un effet d'immersion. Le sujet focaliserait alors son regard et son ouïe sur ces éléments qui se détachent du fond ambiant et deviendraient des sujets d'attention particuliers.

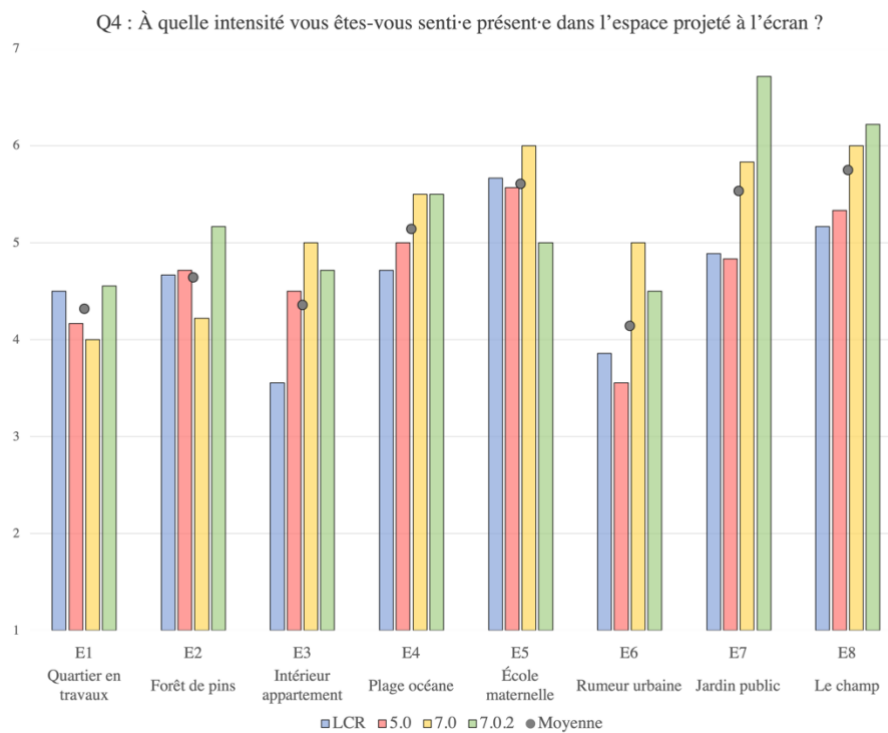


Figure 39 : Histogramme de Q4 par format et par extrait

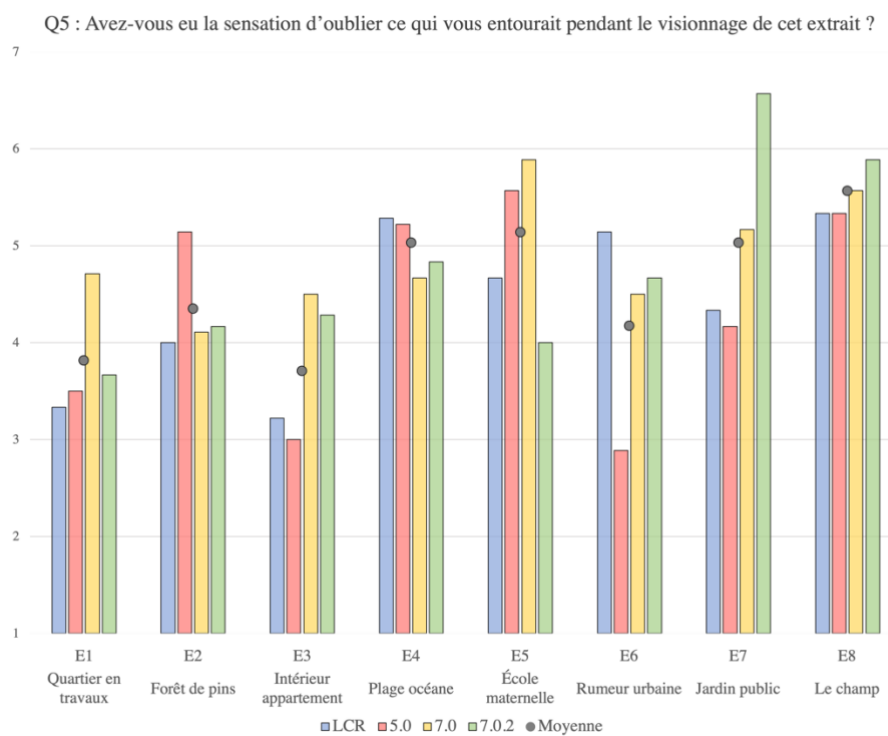


Figure 40 : Histogramme de Q5 par format et par extrait

7.4.2 – Q6 : la cohérence image-son

Q6 : Selon vous, l'espace présenté à l'image était-il cohérent avec l'espace sonore ?

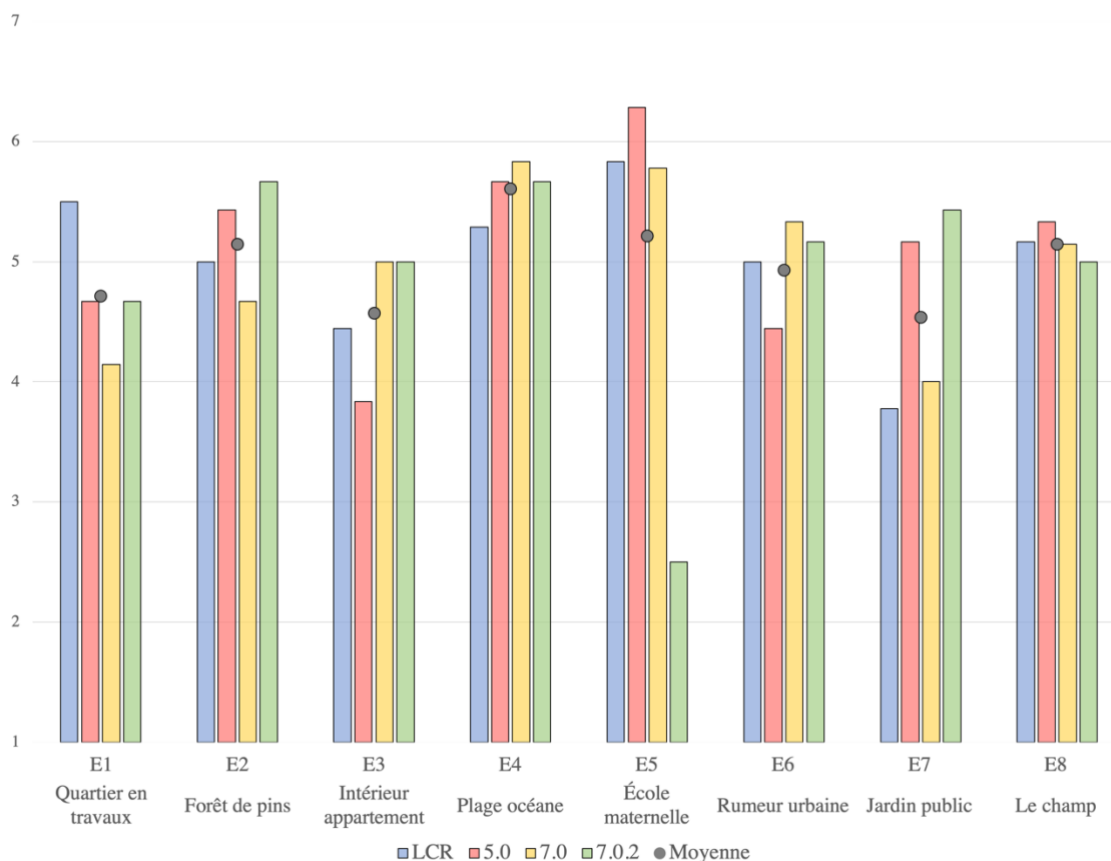


Figure 41 : Histogramme de Q6 par format et par extrait

Lors de l'analyse par format page 80 de Q6 sur la cohérence entre l'espace présenté à l'image et l'espace sonore, aucune différence n'était observable entre les quatre formats. Ici, les résultats détaillés par extrait en Figure 41 ne permettent pas d'avantage de constater de potentielles différences. Les moyennes sont relativement resserrées (écart de moyenne maximum de 1,06 entre E4 et E7) et peu de variations existent entre les formats.

L'extrait E5 présentant des enfants jouant dans une école maternelle est un cas particulier. La moyenne des résultats obtenus pour le format 7.0.2 est de 2,5, un chiffre bas comparé aux autres moyennes. Nous expliquons cela par une corrélation trop importante entre les microphones des canaux d'élévation et ceux de la couche 7.0 par rapport à l'ambiance enregistrée. Lors de l'expérience, plusieurs sujets nous ont fait remarquer qu'ils entendaient trop les cris des enfants ainsi que des frappes de ballons comme des sons provenant du plafond. Ils ont reporté dans leur réponse cette expérience qui n'était pas du tout cohérente avec ce qui

se déroulait à l'image. En effet, le cadre était très large et les enfants, situés au centre de l'écran, jouaient loin.

7.4.3 – Q7 : l'imagination de l'espace

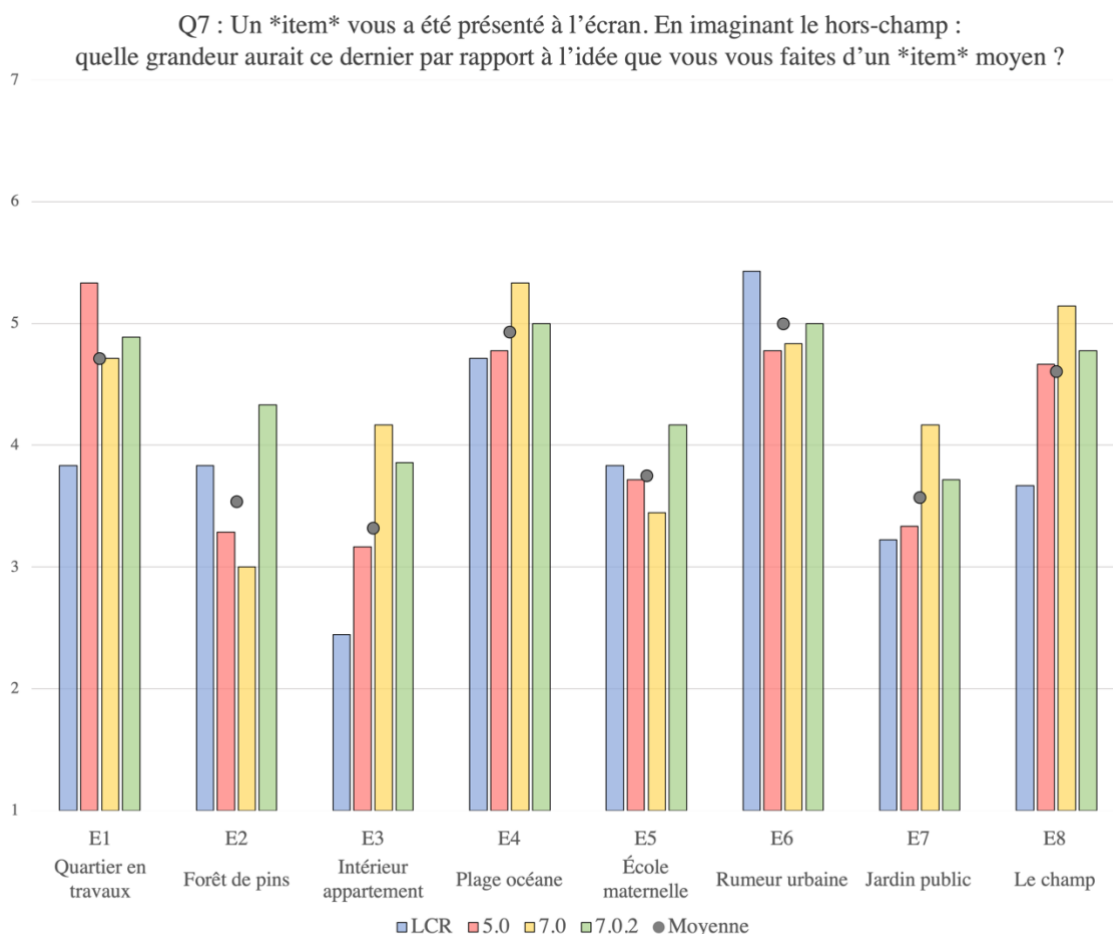


Figure 42 : Histogramme de Q7 par format et par extrait

Pour Q7, nous souhaitons déterminer si un format d'écoute pouvait influencer la perception de grandeur de l'espace qui était projeté à l'image. Nous observons en Figure 42 que malgré des différences de moyenne entre les extraits, peu de variations existent entre ces formats. Nous pouvons cependant noter que pour le troisième extrait présentant un fond d'air d'appartement, les sujets semblaient percevoir l'espace plus grand lors d'une diffusion sonore au format 7.0 et 7.0.2 que pour les deux formats en dessous.

Nous pouvons observer que les deux extraits présentant les moyennes les plus hautes sont E6 (ville vue des toits accompagnée d'une rumeur dense) et E4 (grande plage où les vagues s'échouent). Ce sont des espaces naturellement admis et pensés comme particulièrement vastes.

Les extraits présentant un intérieur d'appartement, un jardin public et une école maternelle, dont les moyennes sont basses par rapport à l'ensemble des extraits, sont des espaces clos ou délimités.

Nous pouvons déduire de l'analyse de cette question que la perception de l'espace est d'avantage guidée par l'image que par un format de diffusion sonore. Une expérience spécifique pourrait être menée sur les lieux intérieurs pour déterminer si, comme dans nos résultats de E3, un lien peut être observé entre le format et la sensation d'espace.

7.5 – Discussion autour de l'expérience

Cette analyse a permis de mettre en évidence les liens entre la configuration d'un format de diffusion et les sensations d'enveloppement et de présence. Ces dernières sont accentuées lorsque les formats comportent un nombre plus important de canaux autour du sujet. Nous rappelons que ces critères sont présentés et adoptés dans la littérature scientifique comme indices potentiels d'immersion (cf. Eaton et Lee, 2019, ainsi que Rigby *et al.*, 2019).

On peut en conclure que le degré de sophistication d'un système de diffusion, en termes de nombre de canaux et de configuration, a une influence directe sur le ressenti d'immersion du spectateur.

Ces résultats sont similaires à ceux analysés par Hamazaki *et al.* (2006) à l'exception du critère de localisation. Celui-ci, évalué en Q3 et proposé comme indice d'immersion par Eaton et Lee (2019) n'était finalement pas pertinent au regard du caractère homogène et continu des ambiances. D'ailleurs, certains sujets ont exprimé leur difficulté à répondre à cette question. Pour une mesure valide de ce critère, il serait alors nécessaire de choisir également des stimuli présentant des sources plus ponctuelles et/ou mouvantes, comme l'utilisation de voix ou d'effets sonores dans le cadre d'une projection cinématographique.

La nature de l'image influe également sur la sensation d'immersion. La représentation de sources sonores à l'écran et leur synchronisation tendent davantage à instaurer une sensation de présence et les sujets se sont sentis plus captivés.

La perception de l'espace est également changeante mais semble dépendre uniquement de la nature visuelle des extraits et de la représentation que les sujets s'en font. La composition de notre corpus a été étudiée pour rassembler des ambiances caractéristiques retrouvées au

cinéma sans prendre en compte la taille ou la densité des espaces comme caractère d'étude. Les résultats présents ne nous permettent donc pas de proposer une analyse valable sur ce critère. De plus, la rédaction de Q7 qui évalue ce dernier était peu intelligible pour les sujets. Nous avons pris le temps de l'expliquer plus en détails au début de chaque session.

Nous avons enfin identifié plusieurs biais dans le protocole de l'expérience qui auraient pu influencer nos analyses.

Premièrement, la répartition des sujets en quatre groupes de visionnage a été nécessaire pour que chacun puisse expérimenter des écoutes sous plusieurs formats sans voir deux fois un même extrait visuel. Cette distribution exigeait alors un nombre suffisant de sujets dans chaque groupe. Nous en avons eu entre six et neuf lors de l'expérience. Ces petits chiffres induisent de grandes variations dans les résultats et complexifient leur interprétation lors d'une analyse par extrait.

L'expérience s'est déroulée dans un auditorium à petit volume de salle. Chaque canal était matérialisé par une seule enceinte ; certaines d'entre elles étaient situées à moins de trois mètres des sujets. Nous pensons que cela a notamment pu accentuer les différences entre les formats et l'effet d'enveloppement. Une expérience similaire dans un environnement grand volume pourrait être menée pour vérifier l'obtention de résultats similaires ou non.

Enfin, certains sujets ont souligné l'importance de diffuser un extrait-exemple pour se familiariser au système d'écoute. D'autres ont qualifié l'expérience de relaxante et ont révélé qu'ils se sentaient plus captivés au fur et à mesure de l'expérience. C'est une tendance que nous observons dans la Figure 39 malgré les variations inter-formats.

Conclusion

Notre travail de recherche avait pour premier objectif de proposer aux lecteurs une réflexion sur l'immersion cinématographique au regard de l'utilisation des systèmes de diffusion sonore immersifs actuellement déployés. À travers l'analyse de différentes études existantes ainsi que par notre propre expérience, nous avons souhaité déterminer l'influence d'un format de diffusion sonore, par son nombre de canaux et leur configuration, sur différents critères perceptifs et cognitifs. Nous avons également souhaité nous inscrire dans un cadre professionnel d'actualité en présentant les systèmes d'enregistrement et de diffusion multicanaux principaux utilisés aujourd'hui, en rencontrant des professionnels au fait de ces enjeux et en développant notre propre dispositif de prise de son pour l'enregistrement d'ambiances multicanales.

Nous avons ainsi défini l'immersion comme un terme polysémique qui se manifeste d'abord lorsque l'attention du spectateur est captée par la présence de situations et de personnages à l'écran auxquels il s'identifie de manière cognitive et émotionnelle. Le spectateur se retrouve alors impliqué dans une narration. Une immersion, cette fois-ci sensorielle, désigne aussi le fait que les spectateurs se sentent submergés par un ensemble de stimulations visuelles et sonores et se retrouvent projetés dans les espaces virtuels du film, au moyen, à titre d'exemples, d'une taille d'écran importante ou d'un dispositif d'enceintes disposées autour d'eux.

Nous avons focalisé notre étude sur le statut particulier que constitue l'ambiance au sein de la bande son. Celle-ci possède des usages différents et permet de faire entendre au spectateur l'environnement sonore, plus ou moins réaliste, dans lequel les situations se déroulent.

Notre expérience, comparant plusieurs formats de diffusion et basée sur la projection audiovisuelle de ce type de sources, a mis en valeur une corrélation entre le nombre de canaux autour du spectateur et les sensations d'enveloppement et de présence. D'après les résultats nous avons aussi conclu que la présence de l'image, la cohérence spatiale visuelle et sonore, ainsi que les points de synchronisation potentiels entre ces deux champs pouvaient influencer la sensation globale d'immersion.

Au fur et à mesure que nous conduisons la réflexion, d'autres questionnements émergent, que pour des impératifs de temps nous n'avons pas poursuivis. En effet, nous

aurions apprécié effectuer des comparatifs d'écoute avec notre dispositif d'enregistrement pour déterminer avec plus de précision la position souhaitée des microphones d'élévation. Nous aurions également aimé rendre celui-ci plus ergonomique en construisant une armature pour réduire son temps de mise en place. Lors de la période d'enregistrement, cela nous aurait permis d'être plus mobile et de capter davantage de sources.

Concernant l'expérience en auditorium et le protocole d'analyse, un grand nombre de sujets nous ont livré leurs impressions, pendant et après les écoutes, à travers des commentaires oraux. Certains rejoignaient les conclusions de l'analyse statistique que nous avons présentée en deuxième partie et d'autres mettaient en valeur des facteurs d'influence, en argumentant notamment leurs différentes réponses selon les extraits. Il aurait été intéressant d'analyser ces commentaires au regard des données récoltées par les questionnaires.

Quelques limitations sont enfin à prendre en compte pour une possible généralisation de nos résultats.

Notre expérience s'est déroulée dans un auditorium professionnel en activité dont l'équipe a eu la gentillesse de nous accueillir pendant deux journées. L'éloignement fait que peu de sujets ont participé à l'expérience. Nous avons vu que ce faible nombre crée des variations importantes dans les résultats, notamment lors d'une analyse par extrait. Pour celle-ci, il nous a été impossible de comparer les résultats entre auditeurs expérimentés et inexpérimentés car cela aurait réduit le nombre de résultats à un chiffre presque nul.

En choisissant de créer un corpus de différentes ambiances sonores nous avons aussi souhaité pouvoir apporter une analyse générale au regard de la sensation d'immersion pour ce type de source. Les stimuli présentés incluaient en conséquence seulement des sons d'ambiance. Cependant c'est bien une bande son complète qu'écoute le spectateur en se rendant en salle ; cette bande comprend la plupart du temps des paroles, des musiques et des effets. L'ambiance est alors intégrée et mixée à un niveau relativement plus faible que celui proposé dans l'expérience. Le spectateur aura donc une stratégie d'écoute différente de celle déterminée par notre contexte expérimental. De plus, nous avons vu que l'ambiance sonore peut avoir une large gamme de significations possibles selon son interaction avec l'image et son utilisation narrative. Ainsi, notre expérience ne prend ni en compte l'ensemble des sources d'une bande son cinématographique ni l'implication affective que peut avoir le spectateur dans un film.

Nos résultats pourraient alors servir d'appui pour l'élaboration d'une expérience similaire où les phénomènes d'engagement et d'absorption d'un sujet dans la narration d'un

film seraient mesurés. Avec toujours comme variable des formats de diffusion, des stimuli audiovisuels intégrant une forme de récit pourraient être utilisés. Ce type de recherches pourrait s'appuyer sur les quatre thèmes déterminés par Rigby *et al.* (2019) et s'inspirer du questionnaire présenté dans leurs travaux.

Il serait également intéressant de mettre en relation le format de diffusion sonore et la focale image, avec pour objectif d'étudier la cohérence spatiale entre le champ visuel et le champ sonore. En effet, lors de notre expérience, quelques sujets ont manifesté leur envie de resserrer ou d'agrandir le champ de l'image en fonction de ce qu'ils écoutaient et des sources visuelles correspondantes. Dans ce même registre, le format de l'image (16:9, 4:3, 1:1, etc.) pourrait constituer aussi une variable à analyser.

Enfin, pour notre expérience, nous avons souhaité présenter des sons d'ambiance comme stimuli car leur statut territoire et narratif dans la bande sonore posait question quant à leur restitution multicanale. Le paramètre de localisation, proposé comme critère d'immersion dans plusieurs études, n'était pas pertinent ici de par la nature majoritairement diffuse des sources. Dans une prochaine recherche, la localisation d'effets sonores ponctuels pourrait alors être étudiée à travers des systèmes de diffusion immersifs et notamment par l'approche du mixage orienté objet. En effet, ces sources constitutives de la bande sonore cinématographique, fixes comme mouvantes, ont parfois la fonction d'ouvrir le cadre de l'image pour proposer des actions se déroulant en hors-champ. Un dispositif de prise de son multicanal permettant une localisation précise et adapté dans sa conception au caractère mobile de certaines de ces sources pourrait être par exemple développé. L'apport éventuel d'une approche orientée objet pour la spatialisation de ces effets sonores en termes de localisation et de présence pourrait ainsi être discuté.

Bibliographie

Adjiman, R., Denizart, J.-M. (2015, 25 mai). *Pour une sémiotique située. Variations sur le thème des intentions d'écoute portées sur les ambiances sonores* [Communication dans un congrès]. Sémiotique appliquée, sémiotique applicable : nouvelles méthodes, Rimouski, Canada. <https://hal-amu.archives-ouvertes.fr/hal-01626782>

Adjiman, R. (2015, 18-20 novembre). *Qualifier et catégoriser les sons d'ambiance pour les indexer dans une sonothèque : axes de réflexion* [Communication dans un congrès]. Documents et dispositifs à l'ère post-numérique : actes du 18e Colloque international sur le document numérique (CIDE18), Montpellier, France.

Adjiman, R. (2018). Les usages des ambiances sonores dans les films de fiction. *Communications*, 2018/1(102), 137-151. <https://doi.org/10.3917/commu.102.0137>

Agrawal, S., Simon, A., Bech, S., Bærentsen, K., Forchhammer S. (2020). Defining Immersion: Literature Review and Implications for Research on Audiovisual Experiences. *Journal of the Audio Engineering Society*, 68(6), 404-417. <https://doi.org/10.17743/jaes.2020.0039>

Association Française du Son à l'Image. (2020). *Dolby Atmos - Atelier Son immersif au cinéma*. <https://www.afsi.eu/articles/58014-atelier-son-immersif-cinema-part-1>

Auro Technologies. (2015). *Auro-3D® home theater setup*. <https://www.auro-3d.com>

Bailblé, C. (2004). Problèmes de spatialisation en son multicanal. *Cahier Louis-Lumière*, (2), 50-61. <https://doi.org/10.3406/cllum.2004.862>

Biocca, F., Delaney, B. (1995). Immersive virtual reality technology. Dans F. Biocca & M. R. Levy (dirs.), *Communication in the age of virtual reality* (pp. 57–124). Routledge.

Bouchara, T., Okuya, Y. (2020). *Livrable 1.3.4 du projet UnitedVR: état de l'art sur la présence, l'immersion, la co-présence et l'incarnation en jeu vidéo et réalité mixte*. Conservatoire National des Arts et Métiers (CNAM) ; CEDRIC laboratory. <https://hal.science/hal-03158860>

Centre National du Cinéma et de l'image animée. (2021). *Bilan 2021 du CNC*. https://www.cnc.fr/professionnels/etudes-et-rapports/bilans/bilan-2021-du-cnc_1689889

Chailloleau, P. (2016). *Caractérisation de la profondeur dans l'espace sonore cinématographique* [Mémoire de master, ENS Louis-Lumière].

Chion, M. (2017). *L'audio-vision - Son et image au cinéma* (4e édition). Armand Colin.

Commission Supérieure Technique de l'image et du son. (s. d.). *Label Excellence / Immersion*. <https://cst.fr/labels-excellence-et-immersion>

Denizart, J.-M. (2017). *L'émergence des significations chez le monteur son, au cours de la recherche et de la sélection des sons : une approche communicationnelle et cognitive* [Thèse de doctorat, Aix-Marseille Université]. Theses.fr. <https://www.theses.fr/2017AIXM0303>

Denizot, F. (2016). *L'approche orientée objet au cinéma : entre surenchère technologique et outil de création sonore* [Mémoire de master, ENS Louis-Lumière].

Dolby Laboratories, Inc. (2011). *Dolby® Surround 7.1 Technical Information for Theaters*.

Dolby Laboratories, Inc. (2012). *Dolby® Atmos™ Un son de prochaine génération pour le cinéma*

Eaton, C., Lee, H. (2019, 27-29 mars). *Quantifying Factors of Auditory Immersion in Virtual Reality* [Communication dans un congrès]. AES International Conference on Immersive and Interactive Audio, York, Royaume-Uni.

Hamasaki, K., Nishiguchi, T., Hiyama, K., Okumura, R. (2006, 20-23 mai). *Effectiveness of height information for reproducing presence and reality in multichannel audio system*. [Communication dans un congrès]. 120th AES Convention, Paris, France.

Hanelt, M., Ehret A. (2017, 20-23 mai). *Ambience recording for 3D audio* [Communication dans un congrès]. 142nd AES Convention, Berlin, Allemagne.

International Telecommunication Union. (2022). *ITU-R BS.775-4 Multichannel stereophonic sound system with and without accompanying picture*.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9), 641-661. <https://doi.org/10.1016/j.ijhcs.2008.04.004>

Kamekawa, T., Marui, A. (2019, 20-23 mars). *Comparison of recording techniques for 3D audio due to difference between listening positions and microphone arrays* [Communication dans un congrès]. 146th AES Convention, Dublin, Irlande.

Lambinet, S. (2022). Le multicanal étendu au cinéma : entretien avec Cyril Holtz, mixeur cinéma, le 29 novembre 2021. *Cahier Louis-Lumière*, (15), 135-143.

Lee, H. (2021). Multichannel 3D Microphone Arrays: A Review. *Journal of the Audio Engineering Society*, 69(1/2), pp. 5-26. <https://doi.org/10.17743/jaes.2020.0069>

Lerouge, C. (1996). *Sur 100 années, le cinéma sonore*. Dujarric.

Lindau, A. (2015). *Spatial Audio Quality Inventory (SAQI). Test Manual*. Audio communication Group, TU Berlin. <https://doi.org/10.14279/depositonce-1.2>

Lombard, M., Reich, R., Grabe, M., Bracken, C., & Ditton, T. (2000). Presence and television : The role of screen size. *Human Communication Research*, 26(1), 75-98. <https://doi.org/10.1111/j.1468-2958.2000.tb00750.x>

Mac, P. (2021, 9 juillet). *Behind the sound: How UEFA has gone large with audio for Euro 2020*. Svgeurope. Consulté le 2 mars 2023 sur <https://www.svgeurope.org/blog/headlines/behind-the-scenes-sound-how-uefa-has-gone-large-with-audio-for-euro-2020/>

Matsui, K. (2015). *22.2 Multichannel Sound Reproduction System for Home Use*. NHK. Consulté le 20 mai 2023 sur <https://www.nhk.or.jp/strl/english/publica/bt/59/3.htm>

de Nanteuil, P.-L. (2008). *Dictionnaire encyclopédique du son*. Dunod

Noly, S. (2013). La prise de son stéréophonique. Dans D. Mercier (dir), *Le livre des techniques du son, tome 3 : l'exploitation* (4 e éd.) (pp. 1-103). Dunod.

Odin, R. (2000). *De la fiction*. De Boeck Supérieur. <https://doi.org/10.3917/dbu.odin.2000.01>

Ono, K., Nishiguchi, T., Matsui, K., Hamasaki, K. (2013, 17-20 octobre). *Portable Spherical Microphone for Super Hi-Vision 22.2 Multichannel Audio*. [Communication dans un congrès]. 135th AES Convention, New York, USA.

Oramus, T., Neubauer, P. (2019, 16-19 octobre). *Comparison Study of Listeners' Perception of 5.1 and Dolby Atmos* [Communication dans un congrès]. 147th AES Convention, New York, USA.

Rigby, J. M., Brumby, D. P., Gould, S. J. J., Cox, A. L. (2019). *Development of a Questionnaire to Measure Immersion in Video Media: The Film IEQ* [Communication dans un congrès]. ACM International Conference on Interactive Experiences for TV and Online Video (TVX '19) New York, NY, USA.

Robert, N. (2021). Le son immersif, partie 2 : Dolby Atmos. *La Lettre de la CST*, (176), 32- 37.

Roginska, A., & Geluso, P. (dirs.). (2018). *Immersive sound: The art and science of binaural and multi-channel audio*. Routledge, Taylor & Francis Group.

Slater, M. (2003). *A Note on Presence Terminology*. https://www.researchgate.net/publication/242608507_A_Note_on_Presence_Terminology

Society of Motion Picture and Television Engineers. (2018). *D-Cinema Immersive Audio Channels and Soundfield Groups* (SMPTE ST 2098-5:2018). <https://doi.org/10.5594/SMPTE.ST2098-5.2018>

Williams, J., Shepstone, S., Murphy, D. (2022, 15-17 août). *Understanding Immersion in the Context of Films with Spatial Audio* [Communication dans un congrès]. International Conference on Audio for Virtual and Augmented Reality, Redmond, WA, USA.

Williams, M. (1984, 27-30 mars). *The Stereophonic Zoom - A Practical Approach to determining the characteristics*. [Communication dans un congrès]. 75th AES Convention, Paris, France.

Williams, M. (2002). *The stereophonic zoom*. Rycote. Consulté le 11 mai 2023 sur https://mymic.rycote.com/uploadresources/images/The_Stereophonic_Zoom.pdf

Williams, M. (2004, 28-31 octobre). *Multichannel Sound Recording using 3,4 and 5 channel arrays for front Sound Stage Coverage*. [Communication dans un congrès]. 117th AES Convention, San Francisco, USA.

Williams, M. (2008, 17-20 mai). *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* [Communication dans un congrès]. 124th AES Convention, Amsterdam, Pays-Bas.

Witmer, B. G., Singer, J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence*, 7(3) 225–240.

Wittek, H., Theile, G. (2017, 18-21 octobre). *Development and application of a stereophonic multichannel recording technique for 3D Audio and VR*. [Communication dans un congrès]. 143th AES Convention, New York, USA.

Zhang, C., Perkis, A., Arndt S. (2017, 30 mai – 2 juin). *Spatial immersion versus emotional immersion, which is more immersive?* [Communication dans un congrès]. Ninth International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, Allemagne.

Annexes

- **Annexe A** : *Immersive Experience Questionnaire for Film and Television* (Rigby *et al.*, 2019)

- **Annexe B** : Abaques stéréophoniques de Williams (2002) pour capsules cardioïdes et supercardioïdes

- **Annexe C** : Liste des équipements utilisés pour les enregistrements d'ambiances multicanales

- **Annexe D** : Photographies du dispositif d'enregistrement multicanal 7.0.2

- **Annexe E** : Photogrammes des extraits audiovisuels projetés pour l'expérience

- **Annexe F** : Questionnaire de l'expérience

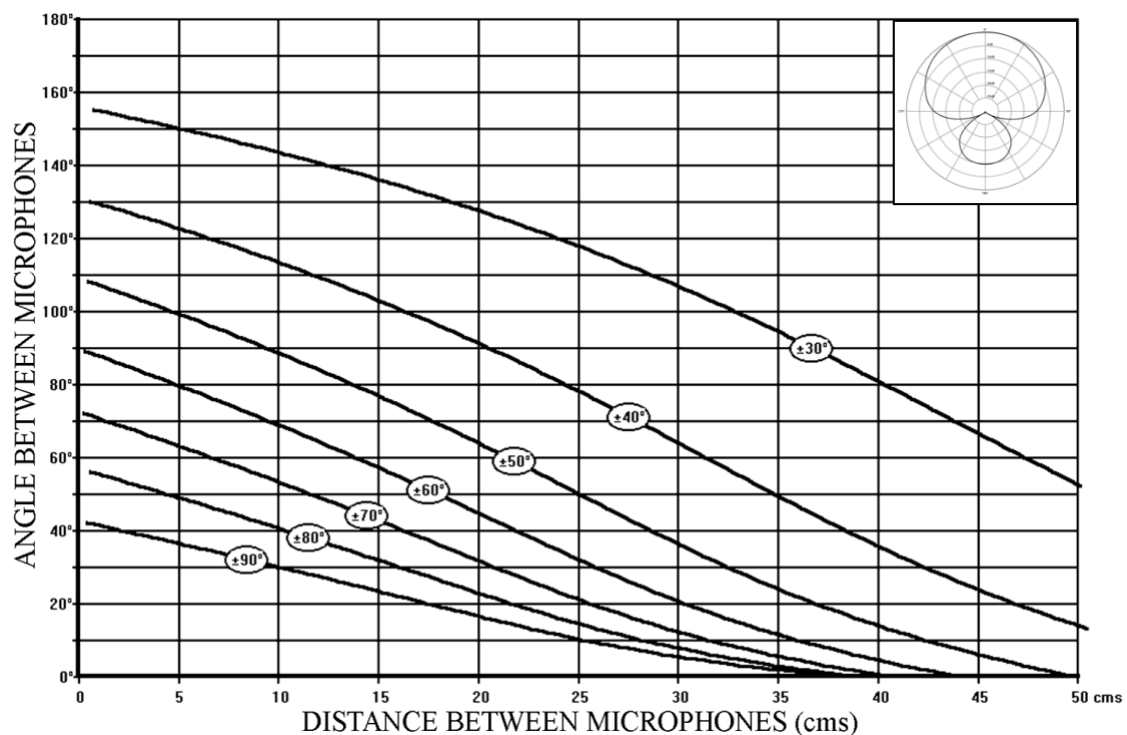
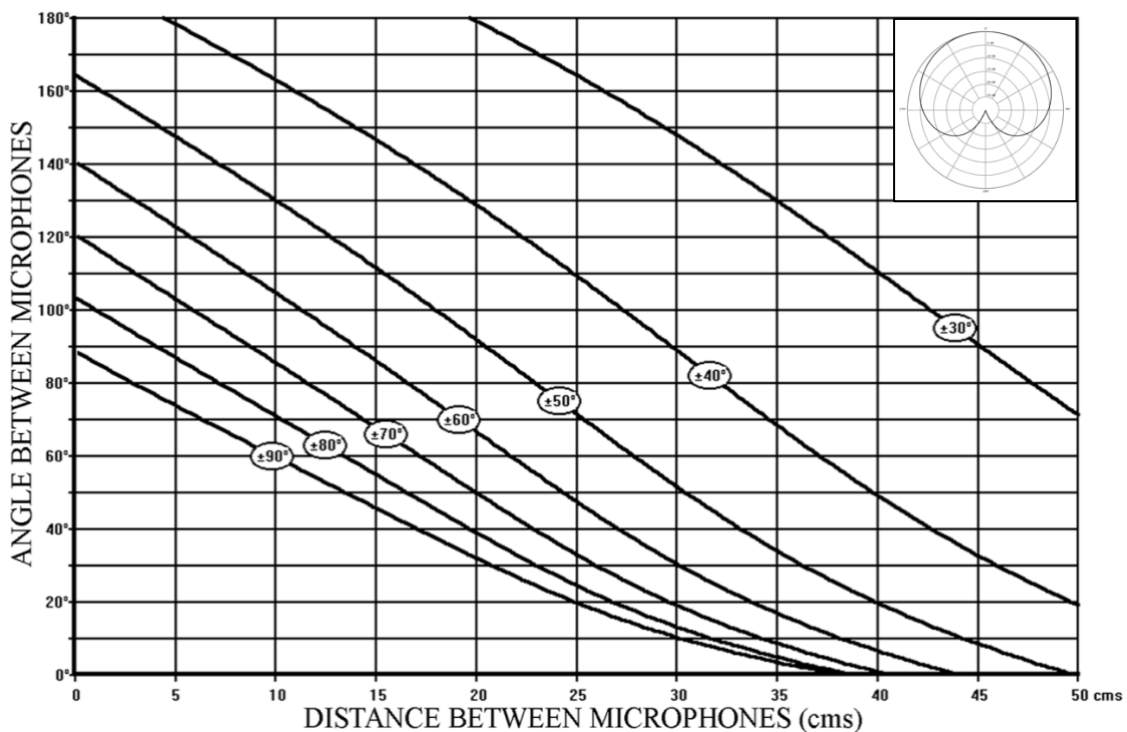
- **Annexe G** : Exemple de réponses à Q2 et graphique en couronne divisé en quatre-vingt-seize cases

- **Annexe H** : Notice d'information et consentement éclairé remise aux sujets

Annexe A : Immersive Experience Questionnaire for Film and Television (Rigby et al., 2019)

1. To what extent did the movie, TV show, or clip hold your attention?
2. To what extent did you feel you were focused on the movie, TV show, or clip?
3. How much effort did you put into watching the movie, TV show, or clip?
4. Did you feel that you were trying your best to follow the events of the movie, TV show, or clip?
5. To what extent did you lose track of time?
6. To what extent did you feel consciously aware of being in the real world whilst watching?
7. To what extent did you forget about your everyday concerns?
8. To what extent were you aware of yourself in your surroundings?
9. To what extent did you notice events taking place around you?
10. Did you feel the urge at any point to stop watching and see what was happening around you?
11. To what extent could you picture yourself in the scene of the events shown in the movie, TV show, or clip?
12. To what extent did you feel as though you were separated from your real-world environment?
13. To what extent did you feel that the movie, TV show, or clip was something you were experiencing, rather than something you were just watching?
14. To what extent was your sense of being in the environment shown in the movie, TV show, or clip stronger than your sense of being in the real world?
15. While watching the movie, TV show, or clip, could you easily picture the events in it taking place?
16. To what extent did you find yourself thinking of ways the story could have turned out differently?
17. To what extent did you find the concepts and themes of the movie, TV show, or clip challenging?
18. Were there any times when you just wanted to give up watching?
19. To what extent did you feel motivated while watching?
20. To what extent did you find the concepts and themes of the movie, TV show, or clip easy to understand?
21. To what extent did you feel like you were making progress towards understanding what was happening during the movie, TV show, or clip, and what you thought might happen at the end?
22. How well do you think you understood what happened in the movie, TV show, or clip?
23. To what extent did you feel emotionally attached to the movie, TV show, or clip?
24. To what extent were you interested in seeing how the events shown in the movie, TV show, or clip would progress?
25. How much did you want the events in the movie, TV show, or clip to unfold successfully for the main characters involved?
26. Were you in suspense about how the events would unfold in the movie, TV show, or clip?
27. At any point did you find yourself become so involved that you wanted to speak to the movie, TV show, or clip directly?
28. To what extent did you enjoy the graphics and the imagery?
29. How much would you say you enjoyed watching the movie, TV show, or clip?
30. When interrupted, were you disappointed that you had to stop watching?
31. Would you like to watch more of this in the future?

Annexe B : Abaques stéréophoniques de Williams (2002) pour capsules cardioïdes et supercardioïdes



Annexe C : Liste des équipements utilisés pour les enregistrements d'ambiances multicanales

<i>DÉNOMINATION</i>	<i>Qt.</i>
MICROPHONES ET ACCESSOIRES	
Microphone Schoeps MK4	9
Bonnette Cinela Cosi-L	5
Bonnette mousse pour micro	9
Bonnette W20 Schoeps	4
Windjammer R1 pour W20 Schoeps	4
Suspension Rycote Invision	9
Barre de couple K&M Large	4
ENREGISTREMENT - MONITORING	
Enregistreur 32 pistes Dante Sound Devices Scorpio	1
Sacoche pour Sound Devices Scorpio	1
Harnais	1
Casque Sennheiser HD25	1
ÉNERGIE - CÂBLES	
Batterie lithium Data 98 W/h Hawk-Woods NP98D	3
Boîtier NP98 avec sortie câble TA4 Hawk-Woods NPC-TA4	1
Chargeur 4 voies rapide HAWK-WOODS ATOM4	1
Câble XLRM-XLRF 5m	10
Bretelle TA5F / 2 CF	4
PIEDS - TRANSPORT	
Pied de micro	4
Pied giraphe	1
Valise pelicase	1
SYNCHRONISATION	
Clap	1
Boîtier Timecode Tentacle sync E	1
C02 Mini-Jack -> Léo	1
C06 Mini-Jack / BNC	1
CAMÉRA	
Caméra Z280	1
SxS-1 64GB	3
Lecteur carte SxS	1
BP-U60	2
BP-U30	1
Chargeur 2 voies BP-U / Alimentation secteur	1
Ensemble tête HiDV	1

Annexe D : Photographies du dispositif d'enregistrement multicanal 7.0.2



Jardin public



Champ



Plage océane



Quartier en travaux

Annexe E : Photogrammes des extraits audiovisuels projetés pour l'expérience



Extrait n°1 : Quartier en travaux



Extrait n°2 : Forêt de pins



Extrait n°3 : Intérieur appartement



Extrait n°4 : Plage océane



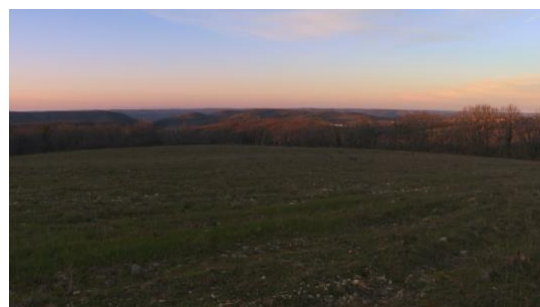
Extrait n°5 : École maternelle



Extrait n°6 : Rumeur urbaine



Extrait n°7 : Jardin public



Extrait n°8 : Champ

Annexe F : Questionnaire de l'expérience

Exemple de questionnaire pour le 1^{er} extrait, suivis des différentes versions de Q7 puis du questionnaire de fin.

Date :	/ 04 / 23	Session n°	Groupe	ID°
--------	-----------	------------	--------	-----

Extrait n°1 : Le quartier en travaux

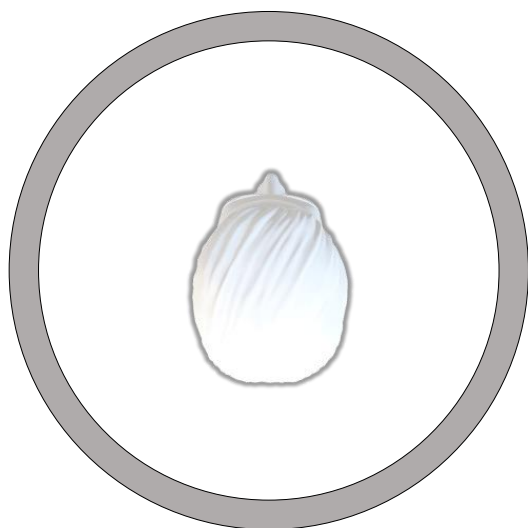
Q1 : « Je me suis senti·e entouré·e de sons. » Êtes-vous d'accord avec cette proposition ?

1-----2-----3-----4-----5-----6-----7

Pas du tout d'accord

Tout à fait d'accord

Q2 : Les dessins suivants représentent une silhouette de tête, du dessus (1) et de profil (2). Indiquez, en noircissant l'espace grisé, de quelle(s) direction(s) semblaient provenir les sons que vous avez entendus :



(1)



(2)

Q3 : « La - ou les - source(s) sonore(s) m'ont paru venir de directions précises ». Êtes-vous d'accord avec cette proposition ?

1-----2-----3-----4-----5-----6-----7

Pas du tout d'accord

Tout à fait d'accord

Q4 : À quelle intensité vous êtes-vous senti·e présent·e dans l'espace projeté à l'écran ?

1-----2-----3-----4-----5-----6-----7

*Je ne me suis pas du tout senti·e
présent·e dans cet espace*

*Je me suis pleinement senti·e
présent·e dans cet espace*

Q5 : Avez-vous eu la sensation d'oublier ce qui vous entourait pendant le visionnage de cet extrait ?

1-----2-----3-----4-----5-----6-----7

*J'étais conscient·e du monde
réel et des éléments autour
de moi*

*J'ai été complètement
captivé·e par l'extrait*

Q6 : Selon vous, l'espace présenté à l'image était-il cohérent avec l'espace sonore ?

1-----2-----3-----4-----5-----6-----7

*Les 2 espaces n'étaient pas
du tout cohérents*

*Les 2 espaces étaient
totalement cohérents*

Q7.1 : Un quartier en travaux vous a été présenté à l'écran. En imaginant le hors-champ : quelle grandeur aurait ce dernier par rapport à l'idée que vous vous faites d'un quartier moyen ?

1-----2-----3-----4-----5-----6-----7

*Beaucoup plus petit qu'un
quartier moyen*

*Nettement plus étendu qu'un
quartier moyen*

Q7.2 : Une forêt vous a été présentée à l'écran. En imaginant le hors-champ : quelle grandeur aurait ce dernier par rapport à l'idée que vous vous faites d'une forêt moyenne ?

1-----2-----3-----4-----5-----6-----7

*Beaucoup plus petit qu'une
forêt moyenne*

*Nettement plus étendu qu'une
forêt moyenne*

Q7.3 : Un intérieur d'appartement vous a été présenté à l'écran. En imaginant le hors-champ : quelle grandeur aurait ce dernier par rapport à l'idée que vous vous faites d'un appartement moyen ?

1-----2-----3-----4-----5-----6-----7

*Beaucoup plus petit qu'un
appartement moyen*

*Nettement plus grand qu'un
appartement moyen*

Q7.4 : Une plage vous a été présentée à l'écran. En imaginant le hors-champ : quelle grandeur aurait ce dernier par rapport à l'idée que vous vous faites d'une plage moyenne ?

1-----2-----3-----4-----5-----6-----7

*Beaucoup plus petit qu'une
plage moyenne*

*Nettement plus étendu qu'une
plage moyenne*

Q7.5 : Une école vous a été présentée à l'écran avec un parc en premier plan. En imaginant le hors-champ : quelle grandeur aurait ce dernier par rapport à l'idée que vous vous faites d'un parc moyen ?

1-----2-----3-----4-----5-----6-----7

*Beaucoup plus petit qu'un
parc moyen*

*Nettement plus étendu qu'un
parc moyen*

Q7.6 : Les toits d'une ville vous ont été présentés à l'écran. En imaginant le hors-champ : quelle grandeur aurait ce dernier par rapport à l'idée que vous vous faites d'une ville moyenne ?

1-----2-----3-----4-----5-----6-----7

*Beaucoup plus petit qu'une
ville moyenne*

*Nettement plus étendu qu'une
ville moyenne*

Q7.7 : Un jardin public vous a été présenté à l'écran. En imaginant le hors-champ : quelle grandeur aurait ce dernier par rapport à l'idée que vous vous faites d'un jardin moyen ?

1-----2-----3-----4-----5-----6-----7

*Beaucoup plus petit qu'un
jardin moyen*

*Nettement plus étendu qu'un
jardin moyen*

Q7.8 : Un champ vous a été présenté à l'écran. En imaginant le hors-champ : quelle grandeur aurait ce dernier par rapport à l'idée que vous vous faites d'un champ moyen ?

1-----2-----3-----4-----5-----6-----7

Beaucoup plus petit qu'un champ moyen

Nettement plus étendu qu'un champ moyen

Questionnaire de fin

- Quel âge avez-vous ?
- Êtes-vous un·e professionnel·le du son, enseignant·e, chercheur·se, artiste, étudiant·e ou ancien·ne étudiant·e dans ce même domaine (rayer la mention inutile) : OUI / NON
- A quelle fréquence allez-vous au cinéma :
 - Au moins 1 fois par semaine
 - Au moins 1 fois par mois (et moins d'une fois par semaine)
 - Au moins 1 fois par an (et moins d'une fois par mois)
- Avez-vous un commentaire à faire sur l'ensemble du test que vous venez de réaliser ?

.....

.....

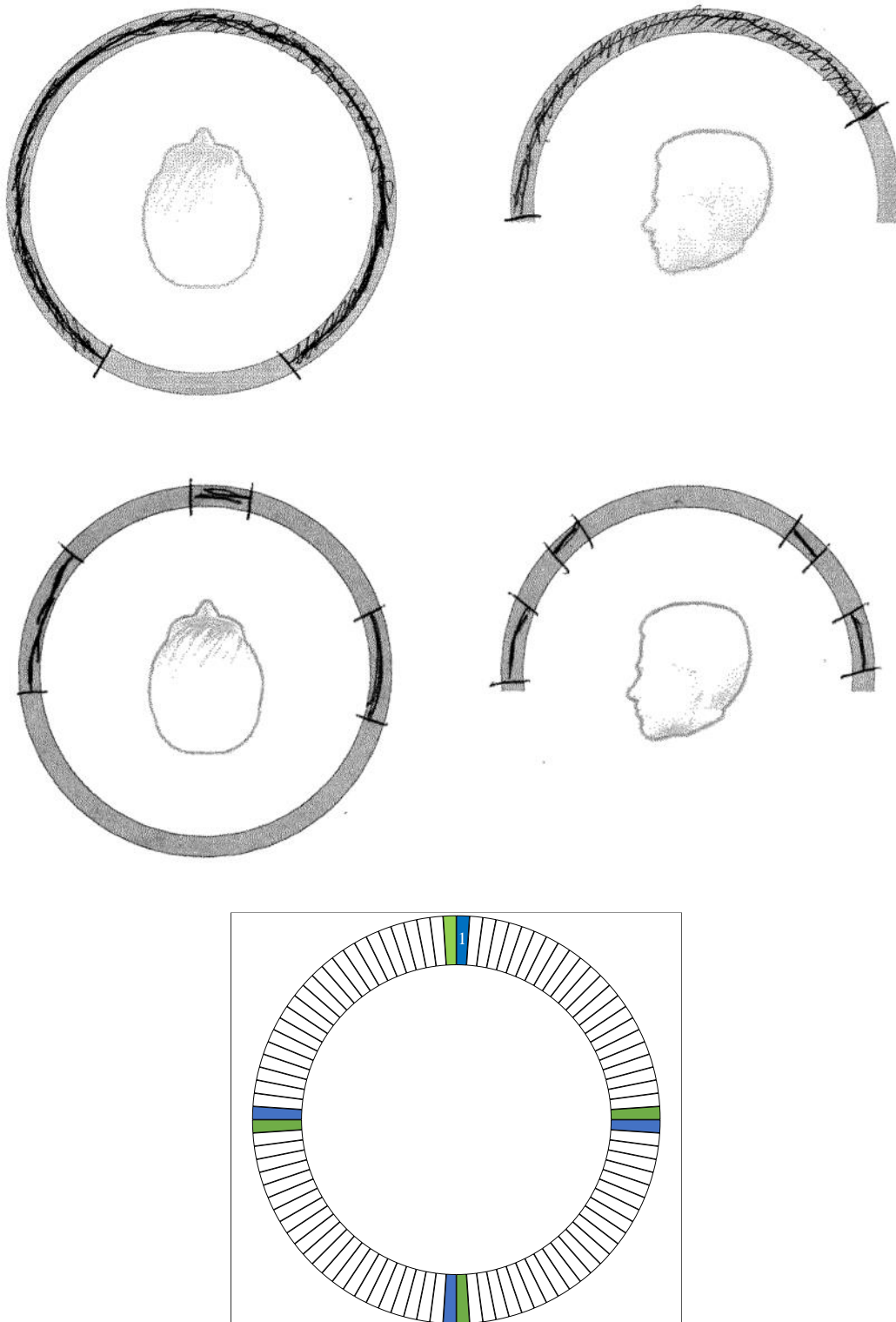
.....

.....

.....

.....

Annexe G : Exemple de réponses à Q2 et graphique en couronne divisé en quatre-vingt-seize cases



Repères de couleur inscrits dans le sens horaire : cases n°1, 24, 25, 48, 49, 72, 73 et 96.

Annexe H : Notice d'information et consentement éclairé remise aux sujets

NOTICE D'INFORMATION ET CONSENTEMENT ECLAIRE

Titre du projet : Partie pratique de mémoire de Master 2 de Son à l'ENS Louis-Lumière. Titre : « Caractériser l'immersion cinématographique : application au cas des ambiances multicanales »

Étudiant responsable du projet : Thibaud Carcy, étudiant en Master 2 de l'ENS Louis-Lumière, t.carcy@ens-louis-lumiere.fr - École nationale supérieure Louis-Lumière, La Cité du Cinéma, 20 Rue Ampère, 93200 Saint-Denis

Lieu de la recherche : Poly Son Post-Production, 49 rue de Bagnolet, 75020 Paris

But du projet de recherche : L'objectif principal de la présente recherche est de mettre en évidence de potentiels liens entre l'utilisation de formats de diffusion sonores multicanaux exploités en salles de cinéma et le ressenti d'immersion spatiale du spectateur.

Ce que l'on attend du participant : Si vous acceptez de participer à cette étude, vous visionnerez dans le cadre de l'expérience plusieurs extraits audiovisuels, constitués de plans fixes de courte durée, sans action apparente, représentant des espaces extérieurs et intérieurs. Ces extraits seront accompagnés d'une bande son dont les caractéristiques varieront au cours de l'expérience. Pour ces visionnages, vous serez installés par petits groupes au centre d'une salle de travail de post-production sonore. Vous serez invités, après chaque extrait, à remplir un questionnaire qui vous sera remis au début de la session. La session complète, incluant l'accueil des participants, les explications et le temps effectif du test aura une durée approximative d'une heure. Les différentes sessions se dérouleront sur la période du 10 au 23 avril 2023, majoritairement en soirée et lors des week-ends. La participation à l'étude étant volontaire, aucune indemnisation n'est prévue pour y participer.

Vos droits de vous retirer de la recherche à tout moment : Accepter de participer à cette étude est totalement volontaire. Vous pourrez décider d'annuler votre consentement et de vous retirer du protocole à tout moment sans donner de raison. Si vous décidez de vous retirer, il vous sera demandé si vous acceptez ou non l'exploitation des données déjà collectées.

Vos droits à la confidentialité et au respect de la vie privée : Les données obtenues seront traitées de manière totalement confidentielle. Dans ce but, nous masquerons votre identité à l'aide d'un numéro qui sera affecté au début de l'étude. Chacune des données collectées (réponses aux questionnaires) sera affectée à votre numéro. Le lien entre la personne et ses données ne sera jamais stocké de façon numérique sur aucun serveur, mais noté sur papier et conservé par l'expérimentateur. Toutes les données collectées seront archivées sur un disque dur de l'expérimentateur ; seul celui-ci y aura accès sans limite de durée. Dans le cas d'une diffusion de l'analyse des résultats de l'étude, orale ou écrite, les données relatives aux résultats des participants resteront anonymes. Enfin, en cas de demande de votre part et

conformément aux dispositions de la loi Informatique et Libertés, vous pourrez exercer vos droits d'accès et de rectification ou suppression des données collectées auprès du responsable du projet.

Risques possibles de l'étude : Les participants sont soumis à des immersions dans des environnements audiovisuels aux niveaux sonores semblables à ceux rencontrés en salle de cinéma. Il est possible que certains participants puissent ressentir une gêne auditive à l'écoute de certaines stimulations sonores.

Moyens de prévention : Chaque séance se fait sous la supervision de l'étudiant responsable du projet, pour s'assurer de son bon déroulement, répondre aux questions des participants si besoin, ou interrompre la session si nécessaire. Le système de diffusion sonore du lieu de l'expérience est calibré selon les normes en vigueur et sera testé avant toutes les sessions pour s'assurer de son bon fonctionnement. Les programmes diffusés dans le lieu de l'expérience ont été préalablement mixés dans un auditorium calibré et leur niveau sonore a été adapté lors d'une séance de pré-test.

Vos droits de poser des questions : Vous pouvez poser des questions au sujet de la recherche à tout moment (avant, pendant et après votre participation) en communiquant avec l'étudiant responsable du projet directement, ou par e-mail à t.carcy@ens-louis-lumiere.fr.

Consentement à la participation : En signant le formulaire de consentement, vous certifiez que vous avez lu et compris les renseignements ci-dessus, qu'il vous a été précisé que vous étiez libre d'annuler votre consentement ou de vous retirer de cette recherche à tout moment, sans préjudice.

A remplir par le participant :

Date, Nom, Prénom :

J'ai lu et compris les renseignements ci-dessus et j'accepte de plein gré participer à cette recherche.
Signature :

A remplir par l'expérimentateur,

10 avril 2023, CARCY, Thibaud

Un exemplaire de ce document vous est remis, un autre exemplaire est conservé par l'étudiant responsable du projet.