Aspects perceptifs de la restitution sonore

Vincent Koehl

DOCTORAT BRETAGNE LOIRE MATHSTIC

Université de Bretagne Occidentale

HABILITATION À DIRIGER DES RECHERCHES

DE L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Acoustique*

Par
Vincent Koehl

Aspects perceptifs de la restitution sonore

Habilitation présentée et soutenue à Brest, le mardi 8 février 2022
Unité de recherche : Laboratoire des Sciences et Technologies de l'Information, de la Communication et de la Connaissance (Lab-STICC UMR CNRS 6285)

Composition du Jury :

Gilles	Coppin	Professeur de l'Institut Mines-Télécom (IMT Atlantique)	Président
Wolfgang	Ellermeier	Professeur des Universités (TU Darmstadt)	Rapporteur
Sabine	Meunier	Chargée de Recherche (CNRS Marseille)	Rapportrice
Rozenn	Nicol	Ingénieure de Recherche (Orange Labs Lannion)	Rapportrice
Mathieu	Paquier	Professeur des Universités (UBO Brest)	Examinateur
Etienne	Parizet	Professeur des Universités (INSA Lyon)	Directeur de recherche
Emanuel	Radoi	Professeur des Universités (UBO Brest)	Examinateur

Table des matières

Remerciements

En tout premier lieu, je tiens à remercier Wolfgang Ellermeier, Sabine Meunier et Rozenn Nicol d'avoir accepté de rapporter le manuscrit et de participer au jury de cette habilitation à diriger des recherches. Je remercie également Mathieu Paquier et Emanuel Radoi d'y avoir participé, ainsi que Gilles Coppin de l'avoir présidé. Merci enfin à Etienne Parizet de m'avoir accompagné dès le début de cette démarche. Merci à vous tous pour vos précieux conseils, questionnements et encouragements.

Merci à tous les collègues du Lab-STICC, permanents ou de passage, et spécialement ceux de l'équipe Perception Sonore, sans qui les travaux de recherche décrits ci-après n'auraient pu être accomplis (ou bien plus laborieusement), pêle-mêle et sans ordre de préférence : Mathieu Paquier, Etienne Hendrickx, Gauthier Berthomieu, Nicolas Côté, Julian Palacino, Bruno Ganzengel et tous ceux que j'oublie mais qui sauront se reconnaître.

Merci à toute l'équipe pédagogique de la filière Image & Son Brest qui m'a chaleureusement accueilli à l'UBO il y a de cela une bonne quinzaine d'années déjà et à ceux qui l'ont rejointe depuis : Karine Peron, Luc Pennamen, François Barvec, Erwan Le Morvan, Nicolas Leborgne, Vincent Mazo, Pascal Olivard, avec une pensée toute particulière pour Georges Tymen.

Merci à Anne Le Roux et à Michèle Kerleroux pour leur soutien dans les démarches administratives. Merci à Vincent Choqueuse pour le coup de pouce à la mise en forme de ce manuscrit.

Pour finir, un grand merci à Simon, Martin et Emilie dont le soutien sans faille m'aura permis de tenir le choc durant quelques longues soirées de rédaction et de finalement voir le bout de cette aventure.

Acronymes anglophones

2I2AFC 2-Interval 2-Alternative Forced Choice. 72
3I3AFC 3-Interval 3-Alternative Forced Choice. 45, 46

ASW Apparent Source Width. 84

BRIR Binaural Room Impulse Response. 55

EEG ElectroEncephaloGraphy. 85
ERB Equivalent Rectangular Bandwidth. 75, 76

FEC Free-air Equivalent Coupling to the ear. 44

HMD Head-Mounted Display. 78, 79
HOA Higher Order Ambisonics. 29, 30, 37
HpTF Headphone Transfer Function. 44
HRTF Head-Related Transfer Function. 35, 78, 85

ILD Interaural Level Difference. 25, 74
IPD Interaural Phase Difference. 72
ITD Interaural Time Difference. 25, 72–76, 83

LF$_E$ Early Lateral Fraction. 84

MOS Mean Opinion Score. 51, 52

PSE Point of Subjective Equality. 72, 74–76

WFS Wave Field Synthesis. 25, 53, 58, 63, 67, 68

Première partie

Curriculum vitae

Chapitre 1

Informations générales

1.1 État Civil

— **Prénom, Nom :** Vincent Koehl.

— **Date et Lieu de Naissance :** né le 27 septembre 1978 à Créhange (57).

— **Situation Familiale :** marié, 2 enfants.

— **Situation Professionnelle :** Maître de Conférences en section 60 (Mécanique, Génie Mécanique, Génie Civil) à l'Université de Bretagne Occidentale (UBO), Faculté des Sciences et Techniques, Laboratoire des Sciences et Techniques de l'Information, de la Communication et de la Connaissance (Lab-STICC UMR CNRS 6285).

1.2 Formation

— **1996 :** Diplôme du Baccalauréat franco-allemand (obtenu avec mention assez bien), Deutsch-Französisches Gymnasium Freiburg im Breisgau (Allemagne).

— **2001 :** Diplôme d'Ingénieur en Génie Mécanique Construction, Institut National des Sciences Appliquées de Lyon.

— **2002 :** Diplôme d'Études Approfondies en Acoustique, Institut National des Sciences Appliquées de Lyon.

— **2005 :** Diplôme de Doctorat en Acoustique, Institut National des Sciences Appliquées de Lyon, école doctorale MÉGA (ED 162 : Mécanique, Énergétique, Génie civil, Acoustique). "Influence des dispersions de structure sur la perception sonore", présentée le 6 décembre 2005 devant le jury composé de :

  — Antoine Chaigne, Professeur des Universités (ENSTA Paris), rapporteur ;

  — Nacer Hamzaoui, Professeur des Universités (INSA Lyon), président du jury ;

  — Catherine Marquis-Favre, Chargée de Recherche (ENTPE Lyon), invitée ;

  — Sabine Meunier, Chargée de Recherche (CNRS Marseille), examinatrice ;

  — Etienne Parizet, Professeur des Universités (INSA Lyon), directeur de thèse ;

  — Patrick Susini, Chargé de Recherche (IRCAM Paris), examinateur ;

  — Reinhard Weber, Professeur des Universités (CvO Uni. Oldenburg), rapporteur.

1.3 Expériences de l'enseignement supérieur et de la recherche

— **2002–2005 :** Moniteur de l'Enseignement Supérieur – Allocataire de Recherche, Institut National des Sciences Appliquées de Lyon. Enseignement de la Mécanique en 1er cycle Ingénieur, thèse de doctorat au Laboratoire Vibrations Acoustique (LVA EA 677).

— **2005–2006 :** Attaché Temporaire d'Enseignement et de Recherche, Institut National des Sciences Appliquées de Lyon. Enseignement de la Mécanique et des Mathématiques en 1er cycle Ingénieur, recherche au Laboratoire Vibrations Acoustique (LVA EA 677).

— **2006–2021 :** Maître de Conférences à l'Université de Bretagne Occidentale (UBO), Brest. Enseignement de l'Acoustique au département de Physique (Licence et Master Image et Son) de la Faculté des Sciences et Techniques, recherche au Laboratoire d'Informatique des Systèmes Complexes (LISyC EA 3883) de 2006 à 2012 puis au Laboratoire des Sciences et Techniques de l'Information, de la Communication et de la Connaissance (Lab-STICC UMR CNRS 6285) à partir de 2012.

1.4 Enseignement et responsabilités pédagogiques

— Enseignement principalement dispensé dans la filière Image & Son Brest (Faculté des Sciences et Techniques, Université de Bretagne Occidentale) :

  — Licence 3 "Sciences Pour l'Ingénieur" parcours "Image et Son" ;

  — Master "Ingénierie de l'image, Ingénierie du son".

  Matières enseignées : acoustique physique, outils mathématiques, électroacoustique, acoustique des salles, psychoacoustique.

— Co-responsable de la filière Image & Son Brest : élaboration des programmes et des dossiers d'accréditation, collaborations académiques et industrielles, dossiers de financement et cahiers des charges de gros équipements (studios d'enregistrements, plateaux techniques). Actuellement :

  — Président du jury de Master 1 "Ingénierie de l'image, Ingénierie du son'' ;

  — Vice-président du jury de Master 2 "Ingénierie de l'image, Ingénierie du son".

— Élu au conseil du département de Physique de la Faculté des Sciences et Techniques, représentant à la commission Bâtiments – Environnement de Travail.

— Représentant de la Faculté des Sciences et Techniques à la commission Arts – Études, délivrant le statut "étudiant-artiste de haut niveau".

— Responsable de l'UE "Physique générale et Acoustique" en Licence 1 Orthophonie (Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale).

1.5 Expertise scientifique

— Conseiller élu au Groupe Perception Sonore (GPS) de la Société Française d'Acoustique (SFA) depuis 2011, responsable scientifique du groupe de janvier 2017 à décembre 2018.

— Membre suppléant de la 60ème section du Conseil National des Universités (CNU), nommé par arrêté ministériel du 13 décembre 2019.

— Membre du comité éditorial de la revue *Acoustique & Techniques*, publiée conjointement par le Centre d'information sur le Bruit (CidB) et la SFA.

— Relecture régulière d'articles pour les revues :
  — *Acta Acustica* (précédemment *Acta Acustica united with Acustica*) ;
  — *Applied Acoustics* ;
  — *Frontiers in Psychology – Auditory Cognitive Neuroscience* ;
  — *Journal of the Audio Engineering Society.*
— Expert mandaté par le Ministère de l'Enseignement Supérieur pour évaluer la candidature de l'École Supérieure de Réalisation Audiovisuelle (ESRA) de Rennes à l'habilitation à délivrer un diplôme reconnu par l'état (2011).

1.6 Activité musicale et associative

Membre de l'Orchestre Universitaire de Brest depuis 2007 (violon), actuellement élu au conseil d'administration de l'association (précédemment président de 2011 à 2014 puis secrétaire de 2014 à 2017).

Chapitre 2

Activités de recherche

2.1 Organisation de congrès scientifiques

— **2005 :** contribution à l'organisation du congrès NOise and Vibrations Emerging Methods (NOVEM 2005) à Saint-Raphaël, organisé par le Laboratoire Vibrations Acoustique (LVA EA 677) de l'INSA de Lyon.

`http://www.novem2005.com/`

— **2006 :** membre du comité d'organisation des 3$^{\text{èmes}}$ Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal Audio (JJCAAS 2006) à Lyon, organisées par les doctorants lyonnais du domaine de la psychoacoustique (INSA, ENTPE, Université de Lyon) sous l'égide de la Société Française d'Acoustique (SFA).

`http://www.jjcaas.org/`

— **2008 :** membre du comité d'organisation des journées Ears Wide Open à Rennes, co-organisées par l'Audio Engineering Society (AES), la SFA et Orange Labs.

`http://www.aesfrance.org/SFA/`

— **2012 :** Acoustics 2012 à Nantes, organisé conjointement par la SFA et l'Institute of Acoustics (IoA, UK). Organisation des sessions "Sound Perception".

`http://www.acoustics2012-nantes.org/`

— **2012 :** membre du comité d'organisation des 2$^{\text{èmes}}$ Journées Perception Sonore (JPS 2012) à Marseille, co-organisées par le Laboratoire de Mécanique et d'Acoustique (LMA UPR CNRS 7051) et le Groupe Perception Sonore (GPS) de la SFA.

`http://www.lma.cnrs-mrs.fr/JPS2012/`

— **2014 :** 12$^{\text{ème}}$ Congrès Français d'Acoustique (CFA 2014) de la SFA à Poitiers, contribution à l'organisation des sessions "Perception" et co-responsable de la session structurée "Sonie".

`http://cfa2014.conference.univ-poitiers.fr/`

— **2016 :** 13$^{\text{ème}}$ Congrès Français d'Acoustique (CFA 2016) de la SFA au Mans, contribution à l'organisation des sessions "Perception" et responsable de la session structurée "Sonie".

`http://cfa2016.univ-lemans.fr/`

— **2017 :** organisateur des 3$^{\text{èmes}}$ Journées Perception Sonore (JPS 2017) à Brest, co-organisées par le Lab-STICC UMR CNRS 6285 (pour l'UBO) et le Groupe Perception Sonore (pour la SFA).

`https://www.univ-brest.fr/JPS2017/`

— **2018 :** soutien à l'organisation des 11$^{\text{èmes}}$ Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal Audio (JJCAAS 2018) à Brest, organisées par des doctorants et post-doctorants du Lab-STICC UMR CNRS 6285 en collaboration avec la Société Française d'Acoustique.

`https://www.univ-brest.fr/JJCAAS2018/`

— **2018 :** 14$^{\text{ème}}$ Congrès Français d'Acoustique (CFA 2018) de la SFA au Havre, membre du comité scientifique, organisation des sessions "Perception / Acoustique Environnementale et Urbaine", co-responsable de la session générale "Perception" et de la session structurée "Environnements sonores : auralisation, restitution et perception".

`http://cfa2018-sfa.fr/`

— **2020 :** 9$^{\text{th}}$ Forum Acusticum (FA 2020) à Lyon, organisé par l'European Acoustics Association (EAA). Définition et préparation des sessions "Psychological & Physiological Acoustics", co-responsable des sessions structurées "Loudness" et "Sound localization by humans".

`https://fa2020.universite-lyon.fr/`

— **2022 :** 14$^{\text{ème}}$ Congrès Français d'Acoustique (CFA 2022) de la SFA à Marseille, organisation des sessions "Perception / Acoustique Environnementale et Urbaine", co-responsable de la session générale "Perception" et de la session structurée "Environnements sonores : auralisation, restitution et perception".

`https://cfa2022.sciencesconf.org/`

2.2 Encadrement de recherche

2.2.1 Niveau Master 2

— **8 étudiants par an :** Projets d'initiation à la recherche de Master 2 "Ingénierie de l'image, Ingénierie du son", effectués durant le 2$^{\text{ème}}$ semestre et donnant lieu à la rédaction d'un mémoire. Les sujets proposés relèvent de la psychoacoustique et permettent de faire le lien entre les problématiques de l'ingénierie sonore et les thèmes de recherche du laboratoire, par exemple :

  — largeur de source apparente (flou de localisation lié au champ réverbéré dans une salle) ;

  — diplacousie binaurale dysharmonique (différence de hauteur tonale entre les deux oreilles).

— **Gauthier Berthomieu (avril 2016 – septembre 2016) :** "Influence des différences interaurales de temps sur le niveau sonore perçu". Stage recherche de Master 2 "Ingénierie de l'image, Ingénierie du son".

  — Publication : [CJ3].

2.2.2 Niveau doctoral

— **Gauthier Berthomieu (novembre 2016 – décembre 2019) :** "Influence de la position d'une source sur le niveau sonore perçu". Thèse de l'Université de Bretagne Occidentale, soutenue le 12 décembre 2019, financée par un Contrat Doctoral d'Établissement.

  — Encadrement : Mathieu Paquier (directeur de thèse), Vincent Koehl.

— Publications : [AI1, AI2, CI1, CI3, CI5, CN1, CJ1, CJ2].

— **Etienne Hendrickx (novembre 2012 – décembre 2015) :** “Influence de la stéréoscopie sur la perception du son : cas de mixages sonores pour le cinéma en relief”. Thèse de l’Université de Bretagne Occidentale, soutenue le 4 décembre 2015, financée par le projet européen (Interreg IVa) Cross-Channel Film Lab 2 (CCFL2) et par la Région Bretagne.
  — Encadrement : Gilles Coppin (directeur de thèse), Mathieu Paquier, Vincent Koehl.
  — Publications : [AI7, AI8, AI11, CI8, CI9, CN4, CN6].

— **Simeon Delikaris-Manias (janvier 2011 – novembre 2017) :** “Parametric spatial audio processing utilising compact microphone arrays”. Thèse de l’Aalto University (Espoo, Finlande), soutenue le 17 novembre 2017, accueilli à l’UBO de janvier 2011 à juin 2011 lors d’une mobilité financée par une bourse de l’Université Européenne de Bretagne (UEB).
  — Encadrement : Ville Pulkki (directeur de thèse), Vincent Koehl (lors de sa mobilité).
  — Publications : [CI11, CI14, CI15, CI20].

— **Nicolas Côté (novembre 2005 – juin 2010) :** “Integral and diagnostic intrusive prediction of speech quality”. Thèse de la Technische Universität Berlin, soutenue le 30 juin 2010, financée par Deutsche Telekom Laboratories (T-Labs) en partenariat avec Orange Labs.
  — Encadrement : Sebastian Möller (directeur de thèse), Valérie Gautier-Turbin, Vincent Koehl.
  — Publications : [AI13, CI18, CI19, CI22].

2.2.3 Niveau post-doctoral

— **Julian Palacino (janvier 2015 – décembre 2016) :** “Edition et DIffusion SONore spatialisée en 3 Dimensions (EDISON 3D)”. Projet Agence Nationale de la Recherche (ANR) “contenus et interactions”.
  — Publications : [AI7, CI4, CI6, CI7].

— **Nicolas Côté (septembre 2010 – juillet 2011) :** “Usage et perception du son spatialisé dans un contexte de réalité virtuelle”. Projet Conseil Général du Finistère (CG29) “aide pour l’accueil d’étudiants post-doctorants”.
  — Publications : [AI4, CI10, CI13].

2.2.4 Participation à des jurys de thèses

— **François Salmon :** “Contrôle des impressions spatiales d’un effet de réverbération dans un environnement virtuel”, thèse soutenue le 26 mars 2021 à l’Institut de Recherche Technologique (IRT) b<>com de Rennes devant le jury composé de :
  — Jean-Yves Aubié, Ingénieur de Recherche (IRT b<>com Rennes), invité ;
  — Roland Badeau, Professeur Institut Mines-Télécom (Télécom Paris), rapporteur ;

- Nicolas Epain, Ingénieur de Recherche (IRT b<>com Rennes), co-encadrant ;
- Etienne Hendrickx, Maître de Conférences (UBO Brest), co-encadrant ;
- Vincent Koehl, Maître de Conférences (UBO Brest), invité ;
- Catherine Lavandier, Professeur des Universités (CYU Cergy), examinatrice ;
- Mathieu Lavandier, Chargé de Recherche (ENTPE Lyon), rapporteur ;
- Mathieu Paquier, Professeur des Universités (UBO Brest), directeur de thèse ;
- Olivier Warusfel, Chargé de Recherche (IRCAM Paris), examinateur.

— **Michaël Vannier :** "Sonie de champs acoustiques stationnaires en situation d'écoute dichotique", thèse soutenue le 11 mai 2015 à l'INSA de Lyon devant le jury composé de :

- Wolfgang Ellermeier, Professeur des Universités (TU Darmstadt), rapporteur ;
- Nicolas Grimault, Chargé de Recherche (CNRS Lyon), rapporteur ;
- Vincent Koehl, Maître de Conférences (UBO Brest), examinateur ;
- Sabine Meunier, Chargée de Recherche (CNRS Marseille), examinatrice ;
- Etienne Parizet, Professeur des Universités (INSA Lyon), directeur de thèse ;
- Daniel Pressnitzer, Directeur de Recherche (CNRS Paris), examinateur.

2.3 Publications

2.3.1 Articles dans des revues internationales à comité de lecture reconnues

[AI1] Gauthier Berthomieu, Vincent Koehl et Mathieu Paquier. « Does loudness relate to the strength of the sound produced by the source or received by the ears? A review of how focus affects loudness ». *Frontiers in Psychology – Auditory Cognitive Neuroscience* 12 (2021). doi : 10.3389/fpsyg.2021.583690.

[AI2] Gauthier Berthomieu, Vincent Koehl et Mathieu Paquier. « Directional loudness of low-frequency noises actually presented over loudspeakers and virtually presented over headphones ». *Journal of the Audio Engineering Society* 67:9 (2019), p. 655-665. doi : 10.17743/jaes.2019.0018.

[AI3] Vincent Koehl, Mathieu Paquier et Etienne Hendrickx. « Effect of the interaural time difference on the loudness of pure tones as a function of the frequency ». *Acta Acustica united with Acustica* 103:4 (2017), p. 705-708. doi : 10.3813/AAA.919098.

[AI4] Mathieu Paquier, Nicolas Côté, Frédéric Devillers et Vincent Koehl. « Interaction between auditory and visual perceptions on distance estimations in a virtual environment ». *Applied Acoustics* 105 (2016), p. 186-199. doi : 10.1016/j.apacoust.2015.12.014.

[AI5] Mathieu Paquier, Vincent Koehl et Brice Jantzem. « Effect of headphone position on absolute threshold measurements ». *Applied Acoustics* 105 (2016), p. 179-185. doi : 10.1016/j.apacoust.2015.12.003.

[AI6] Mathieu Paquier, Vincent Koehl et Cédric Moign. « Effect of drone reed material on great highland bagpipe sound ». *Acta Acustica united with Acustica* 102:4 (2016), p. 752-762. doi : 10.3813/AAA.918991.

[AI7] Etienne Hendrickx, Mathieu Paquier et Vincent Koehl. « Audiovisual spatial coherence for 2D and stereoscopic-3D movies ». *Journal of the Audio Engineering Society* 63:11 (2015), p. 889-899. doi : 10.17743/jaes.2015.77.

[AI8] Etienne Hendrickx, Mathieu Paquier, Vincent Koehl et Julian Palacino. « Ventriloquism effect with sound stimuli varying in both azimuth and elevation ». *The Journal of the Acoustical Society of America* 138:6 (2015), p. 3686-3697. doi : 10.1121/1.4937758.

[AI9] Vincent Koehl et Mathieu Paquier. « Loudness of low-frequency pure tones lateralized by interaural time differences ». *The Journal of the Acoustical Society of America* 137:2 (2015), p. 1040-1043. doi : 10.1121/1.4906262.

[AI10] Mathieu Paquier et Vincent Koehl. « Discriminability of the placement of supraaural and circumaural headphones ». *Applied Acoustics* 93 (2015), p. 130-139. doi : 10.1016/j.apacoust.2015.01.023.

[AI11] Etienne Hendrickx, Mathieu Paquier et Vincent Koehl. « The Influence of Stereoscopy on the Sound Mixing of Movies: A Study on the Front/Rear Balance of Ambience ». *Journal of the Audio Engineering Society* 62:11 (2014), p. 723-735. doi : 10.17743/jaes.2014.0044.

[AI12] Vincent Koehl et Mathieu Paquier. « A comparative study on different assessment procedures applied to loudspeaker sound quality ». *Applied acoustics* 74:12 (2013), p. 1448-1457. doi : 10.1016/j.apacoust.2013.06.008.

[AI13] Nicolas CÔTÉ, Vincent KOEHL, Sebastian MÖLLER, Alexander RAAKE, Marcel WÄLTERMANN et Valérie GAUTIER-TURBIN. « Diagnostic instrumental speech quality assessment in a super-wideband context ». *Journal of the Audio Engineering Society* 60:3 (2012), p. 156-164.

[AI14] Etienne PARIZET et Vincent KOEHL. « Application of free sorting tasks to sound quality experiments ». *Applied Acoustics* 73:1 (2012), p. 61-65. DOI : 10.1016/j.apacoust.2011.07.007.

[AI15] Etienne PARIZET et Vincent KOEHL. « Influence of train colour on loudness judgments ». *Acta Acustica united with Acustica* 97:2 (2011), p. 347-349. DOI : 10.3813/AAA.918414.

[AI16] Vincent KOEHL et Etienne PARIZET. « Influence of structural variability upon sound perception : Usefulness of fractional factorial designs ». *Applied Acoustics* 67:3 (2006), p. 249-270. DOI : 10.1016/j.apacoust.2005.06.002.

2.3.2 Articles dans des revues internationales à comité de lecture dans des numéros "special issues"

[AS1] Vincent KOEHL, Mathieu PAQUIER et Etienne HENDRICKX. « Effects of interaural differences on the loudness of low-frequency pure tones ». *Acta Acustica united with Acustica* 101:6 (2015). Special section : Loudness, p. 1168-1173. DOI : 10.3813/AAA.918909.

[AS2] Vincent KOEHL et Etienne PARIZET. « Listening test methods for perceptual assessment of structural uncertainty ». *Noise control engineering journal* 55:1 (2007). Special issue on uncertainty in noise measurement and prediction, p. 55-66. DOI : 10.3397/1.2402313.

2.3.3 Articles dans des revues nationales à comité de lecture

[AN1] Mathieu PAQUIER et Vincent KOEHL. « Perception sonore de la variabilité de positionnement d'un casque audio ». *Acoustique et Techniques* 60 (2010), p. 21-26.

2.3.4 Communications dans des congrès internationaux à comité de lecture et actes publiés

[CI1] Gauthier BERTHOMIEU, Vincent KOEHL et Mathieu PAQUIER. « Loudness of speech pronounced by a visible or hidden speaker located at several distances ». *Proceedings of Forum Acusticum 2020, the 9th European Congress on Acoustics*. Lyon, France, déc. 2020, p. 3407-3408. DOI : 10.48465/fa.2020.0082.

[CI2] Etienne HENDRICKX, Mathieu LAVANDIER, Vincent KOEHL et Mathieu PAQUIER. « Comparing the influence of several trajectories of head-tracked movements on the externalization of speech stimulus using non-individualized binaural synthesis ». *Proceedings of Forum Acusticum 2020, the 9th European Congress on Acoustics*. Lyon, France, déc. 2020, p. 915. DOI : 10.48465/fa.2020.0218.

[CI3] Gauthier BERTHOMIEU, Vincent KOEHL et Mathieu PAQUIER. « Loudness and distance estimates for noise bursts coming from several distances with and without visual cues to their source ». *Proceedings of ICA 2019, the 23rd International Congress on Acoustics integrating 4th EAA Euroregio*. Aachen, Germany, sept. 2019, p. 3897-3904. DOI : 10.18154/RWTH-CONV-239023.

[CI4] Etienne Hendrickx, Julian Palacino, Vincent Koehl, Frédéric Changenet, Etienne Corteel et Mathieu Paquier. « Should sound and image be coherent during live performances? » *Proceedings of the 2018 Audio Engineering Society International Conference on Spatial Reproduction – Aesthetics and Science*. Paper 11-2. Tokyo, Japan, août 2018. doi : 10.17743/aesconf.2018.978-1-942220-22-0.

[CI5] Gauthier Berthomieu, Vincent Koehl et Mathieu Paquier. « Influence of interaural time differences on the loudness of low-frequency pure tones at varying signal and noise levels ». *Proceedings of Acoustics '17, the 3rd joint meeting of the Acoustical Society of America and the European Acoustics Association*. Boston, MA, juin 2017. doi : 10.1121/2.0000553.

[CI6] Julian Palacino, Mathieu Paquier, Vincent Koehl, Frédéric Changenet et Etienne Corteel. « Assessment of the impact of spatial audiovisual coherence on source unmasking – Preliminary discrimination task ». *Proceedings of the 140th Audio Engineering Society Convention*. Paper 9516. Paris, France, mai 2016.

[CI7] Julian Palacino, Mathieu Paquier, Vincent Koehl, Frédéric Changenet et Etienne Corteel. « Impact of spatial audiovisual coherence on source unmasking ». *Proceedings of ICA 2016, the 22nd International Congress on Acoustics*. Buenos Aires, Argentina, sept. 2016. doi : 10.1121/2.0000476.

[CI8] Etienne Hendrickx, Mathieu Paquier et Vincent Koehl. « Does Stereoscopy Change Our Perception of Soundtracks? » *Proceedings of the 57th Audio Engineering Society Conference : The Future of Audio Entertainment Technology – Cinema, Television and the Internet*. Paper 7-1. Los Angeles, CA, mars 2015.

[CI9] Etienne Hendrickx, Mathieu Paquier et Vincent Koehl. « Should a movie have two different soundtracks for its stereoscopic and non-stereoscopic versions? A study on the front/rear balance ». *Proceedings of IC3D 2013, the 5th IEEE International Conference on 3D Imaging*. Liège, Belgium, déc. 2013. doi : https://doi.org/10.1109/IC3D.2013.6732079.

[CI10] Nicolas Côté, Vincent Koehl et Mathieu Paquier. « Ventriloquism effect on distance auditory cues ». *Proceedings of Acoustics 2012 joint congress (11ème Congrès Français d'Acoustique – 2012 Annual IOA Meeting)*. Nantes, France, avr. 2012, p. 1069-1073.

[CI11] Vincent Koehl, Mathieu Paquier et Simeon Delikaris-Manias. « Subjective assessments of spherical microphone arrays – Paired comparisons of two arrays designed using different microphone models ». *Proceedings of Acoustics 2012 joint congress (11ème Congrès Français d'Acoustique – 2012 Annual IOA Meeting)*. Nantes, France, avr. 2012, p. 521-526.

[CI12] Mathieu Paquier, Vincent Koehl et Brice Jantzem. « Influence of headphone position in pure-tone audiometry ». *Proceedings of Acoustics 2012 joint congress (11ème Congrès Français d'Acoustique – 2012 Annual IOA Meeting)*. Nantes, France, avr. 2012, p. 3931-3936.

[CI13] Nicolas Côté, Vincent Koehl, Mathieu Paquier et Frédéric Devillers. « Interaction between auditory and visual distance cues in virtual reality applications ». *Proceedings of Forum Acusticum 2011, the 6th European Congress on Acoustics*. Aalborg, Denmark, juin 2011, p. 1275-1280.

[CI14] Simeon Delikaris-Manias, Vincent Koehl, Mathieu Paquier, Rozenn Nicol et Jérôme Daniel. « A comparative study of spherical microphone arrays based on subjective assessment of recordings reproduced over different audio systems ». *Proceedings of Forum Acusticum 2011, the 6th European Congress on Acoustics*. Aalborg, Denmark, juin 2011, p. 2227-2230.

[CI15] Vincent Koehl, Mathieu Paquier et Simeon Delikaris-Manias. « Comparison of subjective assessments obtained from listening tests through headphones and loudspeaker setups ». *Proceedings of the 131st Audio Engineering Society Convention*. Paper 8560. New York City, NY, oct. 2011.

[CI16] Mathieu Paquier, Vincent Koehl et Brice Jantzem. « Effects of headphone transfer function scattering on sound perception ». *Proceedings of WASPAA 2011, the 9th Workshop on Applications of Signal Processing to Audio and Acoustics*. New Paltz, NY, oct. 2011, p. 181-184. doi : 10.1109/ASPAA.2011.6082317.

[CI17] Mathieu Paquier, Vincent Koehl, Rozenn Nicol et Jérôme Daniel. « Subjective assessment of microphone arrays for spatial audio recording ». *Proceedings of Forum Acusticum 2011, the 6th European Congress on Acoustics*. Aalborg, Denmark, juin 2011, p. 2737-2742.

[CI18] Nicolas Côté, Vincent Koehl, Valérie Gautier-Turbin, Alexander Raake et Sebastian Möller. « An intrusive super-wideband speech quality model : DIAL ». *Proceedings of Interspeech 2010, the 13th International Conference on Spoken Language Processing*. Makuhari, Japan, sept. 2010, p. 1317-1320.

[CI19] Nicolas Côté, Vincent Koehl, Sebastian Möller, Alexander Raake, Marcel Wältermann et Valérie Gautier-Turbin. « Diagnostic instrumental speech quality assessment in a super-wideband context ». *Proceedings of PQS 2010, the 3rd International Workshop on Perceptual Quality of Systems*. Bautzen, Germany, sept. 2010, p. 65-70. doi : 10.21437/PQS.2010-12.

[CI20] Simeon Delikaris-Manias, Vincent Koehl, Mathieu Paquier, Rozenn Nicol et Jérôme Daniel. « Does capsule quality matter? A comparison study between spherical microphone arrays using different types of omnidirectional capsules ». *Proceedings of Ambisonics Symposium 2010, the 2nd International Symposium on Ambisonics and Spherical Acoustics*. Paper O9. Paris, France, mai 2010.

[CI21] Mathieu Paquier et Vincent Koehl. « Audibility of headphone positioning variability ». *Proceedings of the 128th Audio Engineering Society Convention*. Paper 8147. London, United Kingdom, mai 2010.

[CI22] Nicolas Côté, Vincent Koehl, Valérie Gautier-Turbin, Alexander Raake et Sebastian Möller. « Reference units for the comparison of speech quality test results ». *Proceedings of the 126th Audio Engineering Society Convention*. Paper 7784. Munich, Germany, mai 2009.

[CI23] Vincent Koehl et Mathieu Paquier. « Influence of level setting on loudspeaker preference ratings ». *Proceedings of the 126th Audio Engineering Society Convention*. Paper 7782. Munich, Germany, mai 2009.

[CI24] Vincent Koehl et Mathieu Paquier. « Loudspeaker sound quality : comparison of assessment procedures ». *Proceedings of Acoustics '08, 2nd joint meeting of the Acoustical Society of America and the European Acoustics Association*. Paris, France, juin 2008, p. 2073-2078. doi : 10.1121/1.2933708.

[CI25] Vincent Koehl et Etienne Parizet. « Auditory perception of structural uncertainty ». *Proceedings of Euronoise 2006, the 6th European Conference on Noise Control*. Paper S22–20. Tampere, Finland, mai 2006.

[CI26] Etienne Parizet et Vincent Koehl. « Categorisation: a useful tool for applied perceptive studies ». *Proceedings of Euronoise 2006, the 6th European Conference on Noise Control*. Paper SS04–107. Tampere, Finland, mai 2006.

[CI27] Vincent Koehl et Etienne Parizet. « Auditory assessment of structural uncertainties ». *Proceedings of Novem 2005, the 2nd International Conference on NOise and Vibration Emerging Methods*. Paper 49. Saint-Raphaël, France, avr. 2005.

[CI28] Vincent Koehl et Etienne Parizet. « Perceptual assessment of sounds emitted by a system subject to structural uncertainties ». *Proceedings of the International Symposium on Managing Uncertainties in Noise Measurements and Predictions*. Le Mans, France, juin 2005.

[CI29] Vincent Koehl et Etienne Parizet. « Perceptual consequences of structural uncertainties ». *Proceedings of Forum Acusticum 2005, the 4th European Congress on Acoustics*. Budapest, Hungary, août 2005, p. 1743-1746.

[CI30] Vincent Koehl et Etienne Parizet. « Evaluation of the influence of various dispersions on acoustical perception using experiment designs ». *Proceedings of CFA/DAGA '04 joint congress (7ème Congrès Français d'Acoustique – 30. Deutsche Jahrestagung für Akustik)*. Strasbourg, France, mars 2004, p. 1107-1108.

2.3.5 Communications dans des congrès nationaux à comité de lecture et actes publiés

[CN1] Gauthier Berthomieu, Vincent Koehl et Mathieu Paquier. « Sonie directionnelle de bruits en basses fréquences : influence du mode de restitution ». *Actes du CFA 2018, 14ème Congrès Français d'Acoustique*. Le Havre, France, avr. 2018, p. 103-109.

[CN2] Mathieu Paquier, Clément Garapon, Tristan-Gael Bara, Grégory Mignot, Nathalie Le Bigot, Gauthier Berthomieu, Etienne Hendrickx et Vincent Koehl. « Perception de la distance de sources sonores se rapprochant vs. s'éloignant de lauditeur ». *Actes du CFA 2018, 14ème Congrès Français d'Acoustique*. Le Havre, France, avr. 2018, p. 369-370.

[CN3] Mathieu Paquier, Sabine Meunier, Olivier Macherey, Arnaud Noreña, Vincent Koehl et Etienne Hendrickx. « Limitation des niveaux sonores pour les lieux diffusant de la musique amplifiée : données expérimentales sur le danger des niveaux élevés, réglementation et problèmes engendrés en sonorisation ». *Actes du CFA 2018, 14ème Congrès Français d'Acoustique*. Le Havre, France, avr. 2018, p. 501-502.

[CN4] Etienne Hendrickx, Mathieu Paquier, Vincent Koehl et Julian Palacino. « Effet ventriloque pour des sources sonores variant simultanément en azimut et elévation ». *Actes du CFA 2016, 13ème Congrès Français d'Acoustique*. Le Mans, France, avr. 2016, p. 2313-2319.

[CN5] Vincent Koehl, Mathieu Paquier et Etienne Hendrickx. « Effet de différences interaurales de temps sur la sonie de sons purs en fonction de la fréquence ». *Actes du CFA 2016, 13ème Congrès Français d'Acoustique*. Le Mans, France, avr. 2016, p. 1161-1165.

[CN6] Etienne Hendrickx, Mathieu Paquier et Vincent Koehl. « Influence de la stéréoscopie sur le mixage des ambiances "surround" au cinéma ». *Actes du CFA 2014, 12ème Congrès Français d'Acoustique*. Poitiers, France, avr. 2014, p. 1769-1775.

[CN7] Vincent Koehl et Mathieu Paquier. « Influence de différences interaurales de temps sur la sonie de sons purs en basse fréquence ». *Actes du CFA 2014, 12ème Congrès Français d'Acoustique*. Poitiers, France, avr. 2014, p. 1945-1951.

[CN8] Mathieu Paquier et Vincent Koehl. « Perception sonore de la variabilité de positionnement d'un casque audio ». *Actes du CFA 2010, 10ème Congrès Français d'Acoustique*. Lyon, France, avr. 2010.

2.3.6 Communications dans des journées nationales

[CJ1] Gauthier Berthomieu, Vincent Koehl et Mathieu Paquier. *Influence du mode de restitution sur la sonie directionnelle de bruits en basses fréquences*. JJCAAS 2018, 11èmes Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal audio. Brest, France, juin 2018.

[CJ2] Gauthier Berthomieu, Vincent Koehl et Mathieu Paquier. *Influence des différences interaurales de temps sur la sonie de sons purs en basses fréquences à différents niveaux de signal et de bruit*. JPS 2017, 3èmes Journées Perception Sonore. Brest, France, juin 2017.

[CJ3] Gauthier Berthomieu et Vincent Koehl. *Influence des différences interaurales de temps sur le niveau sonore perçu*. JJCAAS 2016, 10èmes Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal audio. Paris, France, nov. 2016.

[CJ4] Vincent Koehl et Mathieu Paquier. *Sonie directionnelle en basse fréquence - Effet sur des sons purs latéralisés uniquement par des différences de temps*. JPS 2012, 2èmes Journées Perception Sonore. Marseille, France, déc. 2012.

[CJ5] Vincent Koehl et Etienne Parizet. *Caractérisation perceptive de dispersion de structures*. 3èmes Journées du GDR 2493 "Bruit des transports". Bron, France, jan. 2005.

[CJ6] Vincent Koehl et Etienne Parizet. *Influence de dispersions structurales sur la perception du son*. JJCAAS 2005, 2èmes Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal audio. Marseille, France, mars 2005.

[CJ7] Vincent Koehl et Etienne Parizet. *Évaluation de l'influence de défauts mécaniques sur la perception sonore d'un objet*. JJCAAS 2003, 1ères Journées Jeunes Chercheurs en Audition, Acoustique musicale et Signal audio. Paris, France, oct. 2003.

2.3.7 Thèse de doctorat

[TD1] Vincent Koehl. « Influence des dispersions de structure sur la perception sonore ». Thèse de doctorat. Lyon : Institut National des Sciences Appppliquées, déc. 2005.

Deuxième partie

Mémoire de recherche

Liste des figures

Introduction

Mes activités de recherche relèvent de la psychoacoustique et portent principalement sur la perception du son dans des contextes où celui-ci est restitué, tels que la réalité virtuelle, le cinéma, la diffusion musicale (avec ou sans image associée) et les télécommunications.

Ces recherches sont menées à l'Université de Bretagne Occidentale, précédemment dans le cadre du Laboratoire d'Informatique des Systèmes Complexes (LISyC EA 3883, de 2006 à 2012) et actuellement dans le cadre du Laboratoire des Sciences et Techniques de l'Information, de la Communication et de la Connaissance (Lab-STICC UMR CNRS 6285, depuis 2012). La thématique de recherche "Perception Sonore", historiquement liée aux équipes de recherche du LISyC puis du Lab-STICC spécialisées dans la réalité virtuelle, est désormais traitée par une équipe dédiée (Responsable : Mathieu Paquier) dans la nouvelle organisation du laboratoire. La plupart des études décrites ci-dessous ont d'ailleurs été effectuées avec mes deux collègues actuels de l'équipe Perception Sonore : Etienne Hendrickx et Mathieu Paquier. Ce mémoire traite donc des recherches dont j'ai été l'instigateur, que j'ai encadrées ou auxquelles j'ai participé activement dans le cadre de notre équipe de recherche.

Mes travaux se sont déroulés dans le cadre de projets nationaux (pôles de compétitivité, conseil départemental du Finistère, région Bretagne, ANR) et européen (Interreg IVa). Ils ont permis de développer des relations industrielles sous forme de contrats (Orange Labs, Deutsche Telekom, Canon Research France) ou de partenariats (Cabasse, Schoeps Mikrofone). Ils ont également donné lieu à des collaborations académiques internationales (Technische Universität Berlin, Aalto University Helsinki).

Ces recherches portent sur des aspects fondamentaux (la localisation, la sonie) comme appliqués (la qualité perçue, la spatialisation adéquate) de la perception sonore. Elles sont ici organisées selon 5 axes représentatifs :

1. Évaluation perceptive des systèmes de captation sonore spatialisée ;
2. Évaluation perceptive des systèmes de restitution sonore ;
3. Modélisation de la qualité vocale en téléphonie mobile ;
4. Interactions audiovisuelles ;
5. Sonie en fonction de la localisation sonore.

Ce document prend le parti de présenter de manière synthétique l'ensemble des études réalisées en consacrant un chapitre à chacun de ces axes. Ceux-ci peuvent parfois comporter des problématiques scientifiques et des méthodologies expérimentales communes. Ces travaux ainsi que leurs principaux résultats sont décrits ci-après, après les avoir brièvement replacés dans leurs contextes en termes de projets nationaux ou internationaux, de contrats industriels et d'encadrements de thèses. Le lecteur pourra se référer aux principales publications produites

pour le détail des protocoles expérimentaux, des analyses statistiques, des discussions et conclusions complètes sur ces résultats.

Enfin, un sixième et dernier chapitre sera consacré aux perspectives ouvertes par ces résultats et détaillera le prolongement envisagé pour ces travaux ainsi que le développement de nouveaux axes de recherche.

Chapitre 1

Évaluation perceptive des systèmes de captation sonore spatialisée

1.1 Contexte

Les travaux sur l'évaluation des systèmes de captation sonore en vue d'une restitution spatialisée se sont d'abord déroulés dans le cadre d'un contrat de recherche avec Orange Labs (Lannion, interlocuteurs : Rozenn Nicol, Jérôme Daniel). Il s'agissait dans un premier temps d'évaluer perceptivement des dispositifs multimicrophoniques basés sur différentes technologies de son spatialisé : différences d'intensité et de temps, ambisonique à différents ordres [CI17]. La seconde partie de cette collaboration a été consacrée à la sélection des transducteurs les plus adaptés à une captation sonore spatialisée par réseau microphonique sphérique [CI20]. Un partenariat avec Schoeps Mikrofone (Karlsruhe – Allemagne, interlocuteur : Helmut Wittek) a permis de concevoir et tester des réseaux incluant différents types de microphones (ingénieur de recherche : Simeon Delikaris-Manias). Les contenus audio captés par ces différents systèmes ont fait l'objet d'une évaluation perceptive lors de restitutions sur dispositifs constitués d'enceintes [CI14, CI11] ou au casque audio [CI15].

Cet axe de recherche s'est poursuivi lors d'une collaboration avec le Laboratory of Acoustics and Audio Signal Processing de l'Aalto University (Espoo – Finlande, interlocuteur : Ville Pullki) où Simeon Delikaris-Manias a ensuite effectué une thèse de doctorat [1] dans la continuité des études déjà effectuées sur les réseaux microphoniques. Ce travail de thèse a pu être entamé dans notre laboratoire dans le cadre d'une mobilité financée par une bourse de l'Université Européenne de Bretagne [2].

Les résultats de ces différents travaux ont fait l'objet d'une mise en oeuvre concrète dans le cadre du projet "pôles de compétitivité" MARVEST (MARitime Virtual rEality and Simulator Technologies) labellisé par le pôle mer Bretagne Atlantique et le pôle mer Méditerranée. Le but était de créer un simulateur de pilotage à vocation navale utilisant des technologies permettant d'améliorer le réalisme sonore et visuel (autres membres du consortium : OPTIS, ECA-Faros, ECA-Sindel, Genesis, Clarté). Ainsi les réseaux microphoniques développés, basés sur la technologie Higher Order Ambisonics (HOA), ont été utilisés pour une captation sonore spatialisée à l'intérieur d'un navire remorqueur. Ces enregistrements ont ensuite été déployés et validés dans le simulateur de démonstration.

1.2 Arbres et réseaux microphoniques

La captation sonore en multicanal est souvent effectuée à l'aide d'arbres microphoniques où chaque microphone est destiné à capter un signal assigné à l'une des cinq enceintes acoustiques composant le système de restitution :

— C ($Center$) pour l'enceinte centrale ;

— L ($Left$) pour l'enceinte gauche ;

— R ($Right$) pour l'enceinte droite ;

— L_S ($Left_{Surround}$) pour l'enceinte gauche arrière ;

— R_S ($Right_{Surround}$) pour l'enceinte droite arrière ;

selon la terminologie de l'Union Internationale des Télécommunications (désignée ci-après ITU pour International Telecommunication Union). Ces enceintes sont disposées autour de l'auditeur selon la recommandation ITU-R BS.775–3 [3]. Un exemple de dispositif multicanal "ITU" est représenté ci-après en Figure 1.2.

Les arbres microphoniques destinés à enregistrer les signaux alimentant respectivement ces enceintes sont composés de microphones captant les sources sonores avec des différences d'intensité et de temps selon leurs directivités et espacements respectifs, permettant de spatialiser des sources virtuelles entre les hauts-parleurs sur le même principe que la stéréophonie [4]. Le but de cette expérience était de comparer ces dispositifs aux principes empiriques, largement utilisés par les ingénieurs du son, à des réseaux microphoniques basés sur la technique ambisonique [5], permettant une décomposition du champ sonore à différents ordres sur les harmoniques d'une sphère (celle-ci définissant l'ordre 0).

Cette étude visait à comparer 4 dispositifs de captation sonore spatialisée destinée à un rendu multicanal, dont 2 arbres microphoniques :

— Wide Cardioid Surround Array (WCSA) [6], composé de 5 microphones hypocardioïdes (modèle DPA 4015, présentant 3 dB d'atténuation à 90°) montés sur un support dédié (modèle DPA S5) ;

— Optimized Cardioid Triangle with Surround (OCT-Surround) [7], composé d'un microphone cardioïde (modèle Schoeps CCM 4, présentant 6 dB d'atténuation à 90°) et de deux microphones supercardioïdes (modèle Schoeps CCM 41, présentant 9 dB d'atténuation à 90°) pour le triangle OCT ainsi que deux microphones cardioïdes (modèle Schoeps CCM 4 à nouveau) pour le "Surround" ;

et 2 réseaux microphoniques ambisoniques :

— un microphone ambisonique d'ordre 1 (modèle Soundfield ST 350) composé de 4 capsules cardioïdes disposées en tétraèdre [8] ;

— un microphone ambisonique d'ordre 4 (HOA pour Higher Order Ambisonics) composé de 32 microphones omnidirectionnels (modèle DPA 4060) répartis sur la surface d'une sphère de diamètre 7 cm [9].

Ces quatre systèmes de prise de son ont été déployés simultanément dans la salle "Le Tambour" de l'Université Rennes 2 pour effectuer des enregistrements lors des journées Ears Wide Open en mars 2008 (voir section 2.1 de la partie I). Le Tambour est une salle de spectacle polyvalente, à l'acoustique modulable. Les quatre systèmes ont été placés sur l'avant de la scène, superposés de manière à ce que leurs centres puissent être considérés comme coïncidents dans

le plan horizontal (Figure 1.1(a)). Deux formations musicales ont été enregistrées dans le but de capter des scènes sonores différant notamment en termes de largeur et de profondeur :

— un quatuor de guitares sur une seule rangée courte (Figure 1.1(b)) ;

— un big band de jazz composé d'une vingtaine de musiciens répartis sur plusieurs rangées larges (Figure 1.1(c)).

(a) (b) (c)

FIGURE 1.1 – Dispositifs microphoniques positionnés dans la salle du tambour (a) afin de capter un quatuor de guitares (b) et un big band de jazz (c).

De courts extraits (environ 5 s) de ces prises de son ont ensuite été restitués sur un système multicanal composés de 5 enceintes PSI A25M disposées selon la recommandation de l'ITU [3] dans une salle peu réverbérante : le studio d'enregistrement de la formation "Image & Son" de l'Université de Bretagne Occidentale (Figure 1.2). L'évolution du temps de réverbération (TR) de cette salle en fonction de la fréquence est en accord avec les préconisations de l'International Electrotechnical Commission relatives aux tests d'écoute sur haut-parleurs (IEC 60268–13 [10]). Ces stimuli ont été évalués par 18 étudiants en Master ingénierie du son et 20 autres étudiants, respectivement considérés auditeurs "experts" et "naïfs" [11]. Tous ont déclaré ne pas avoir de problème d'audition.

Les sujets devaient initialement indiquer leurs jugements de préférence lors de comparaison par paires (transformés ensuite en score de préférence). Dans une deuxième session, les auditeurs devaient comparer ces enregistrements en termes d'enveloppement, de naturel, de localisation et de profondeur. Ces termes correspondent aux attributs perceptifs qui sous-tendent généralement les jugements de préférence [12].

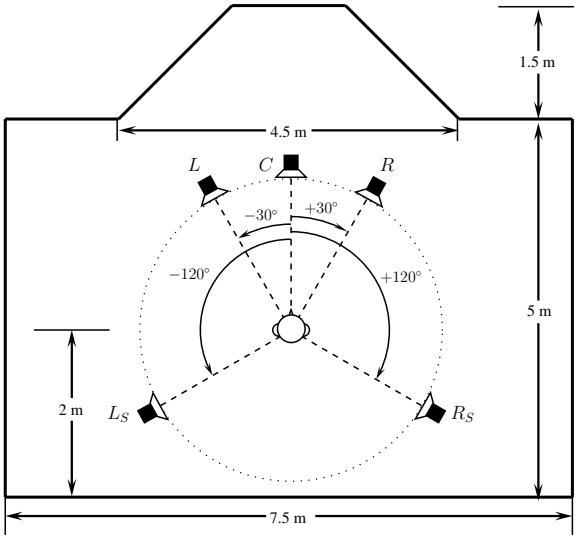
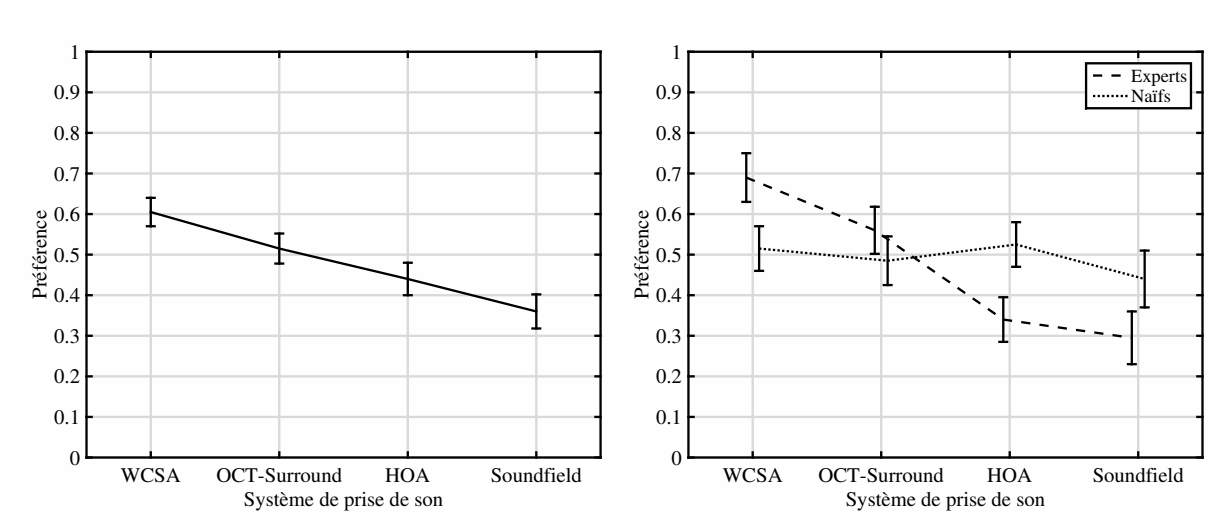


FIGURE 1.2 – Représentation schématique de la salle d'écoute dans laquelle sont installés le système multicanal restituant les enregistrements effectués par les différents dispositifs microphoniques et un auditeur procédant à leur évaluation subjective.

Les résultats ont montré [CI17] une variation significative du score de préférence selon le système de prise de son, indiquant une préférence pour les arbres microphoniques comparativement aux systèmes ambisoniques (Figure 1.3(a)). Ces préférences sont particulièrement marquées pour les auditeurs experts (Figure 1.3(b)) et principalement corrélées aux comparaisons en termes de naturel.



(a) (b)

FIGURE 1.3 – Score de préférence moyen (dans son intervalle de confiance à 95 %) en fonction du système de prise de son, tous auditeurs confondus (a) et en séparant les deux groupes d'auditeurs (b) : experts (trait tireté) et naïfs (trait pointillé).

1.3 Influence de la qualité des transducteurs dans un réseau microphonique

Les réseaux microphoniques ambisoniques apparaissent ainsi significativement moins bien noté que les arbres microphoniques. Le fait que ces scores de préférence soient liés au naturel de la restitution a permis de formuler une hypothèse selon laquelle une piste d'amélioration réside dans le choix des modèles de microphones utilisés dans ces réseaux sphériques.

Par exemple, le réseau microphonique ambisonique d'ordre 4 développé par Orange Labs était celui muni des plus petits microphones (DPA 4060). Ces transducteurs miniatures présentent l'avantage de pouvoir être disposés en grand nombre dans un faible encombrement (32 microphones dans une sphère de diamètre 7 cm) [9] mais leurs caractéristiques intrinsèques (réponse en fréquence et rapport signal sur bruit) peuvent souffrir de la comparaison avec des modèles de microphones possédant une plus large membrane. Ainsi la qualité et le naturel d'un tel réseau microphonique pourraient être significativement améliorés par l'utilisation de microphones de qualité supérieure.

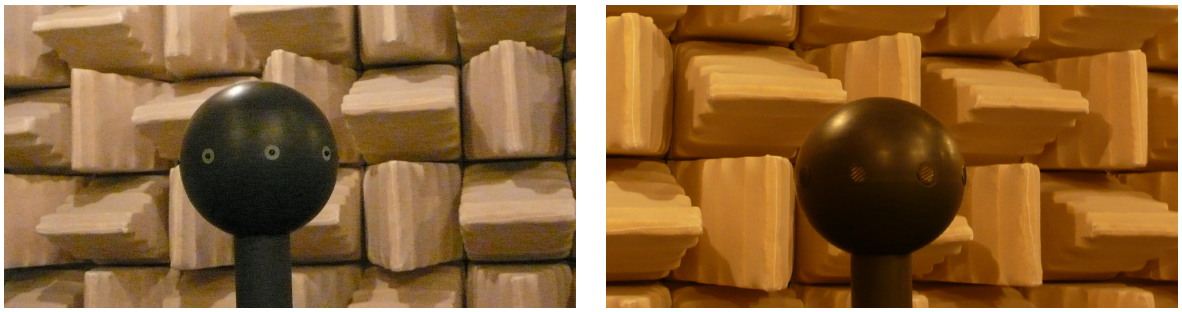
Le but de cette étude était donc d'évaluer l'apport de la qualité des transducteurs à la qualité globale d'un réseau ambisonique en comparant des réseaux ne différant que par le modèle de transducteur utilisé. Une collaboration avec Schoeps Mikrofone a permis d'obtenir le prêt de 8 microphones omnidirectionnels compacts Schoeps CCM 2 possèdant une réponse en fréquence et un rapport signal sur bruit significativement différents des microphones miniatures DPA 4060 [CI20].

Ainsi, deux réseaux microphoniques sphériques ne différant que par le modèle des microphones omnidirectionnels utilisés ont été conçus à base d'une sphère rigide de diamètre 15 cm :

— A : réseau sphérique composé de 8 microphones DPA 4060 (Figure 1.4(a)) ;

— B : réseau sphérique composé de 8 microphones Schoeps CCM 2 (Figure 1.4(b)).

Ces deux réseaux permettent une décomposition du champ sonore à l'ordre 3 dans le plan horizontal.



(a) (b)

FIGURE 1.4 – Sphère rigide incluant 8 microphones omnidirectionnels de modèle DPA 4060 (a) ou de modèle Schoeps CCM 2 (b).

Ces deux sphères ont été utilisées pour effectuer des prises de son d'une formation musicale : un quatuor composé d'une flûte, d'une clarinette, d'une contrebasse et d'un hautbois. Ces captations ont été réalisées dans le studio d'enregistrement de la formation "Image & Son". Les deux systèmes ont été placés devant les musiciens, superposés de manière à ce que leurs centres puissent être considérés comme coïncidents dans le plan horizontal (Figure 1.5).

FIGURE 1.5 – Réseaux microphoniques sphériques disposés devant les musiciens du quatuor.

De courts extraits (environ 5 s) de ces prises de son ont ensuite été décodés [CI20] de manière à être restitués dans les mêmes conditions que celles décrites en section 1.2 (Figure 1.2). Le système multicanal (L, C, R, R_S, L_S) était donc composé de 5 enceintes PSI A25M disposées selon la recommandation de l'ITU [3] dans une salle d'écoute vérifiant les préconisations de l'IEC [10]. Des versions monophoniques (décodage à l'ordre 0 restitué uniquement sur l'enceinte C) et stéréophoniques (décodage à l'ordre 1 restitué sur les enceintes L et R) de ces extraits ont également été proposées car elles permettent a priori une meilleure discrimination entre les systèmes à comparer [13, 14].

Ces stimuli ont été évalués dans un premier temps par 13 auditeurs "naïfs" [11] ayant déclaré ne pas avoir de problème d'audition ; puis par 12 étudiants en Master ingénierie du son, considérés comme auditeurs "experts" [11] et possédant des seuils d'audition normaux d'après un audiogramme passé dans le mois précédent le test. Des comparaisons par paire entre les enregistrements issus des systèmes A et B étaient à réaliser sur des échelles de différence (Figure 1.6(a)) et de préférence (Figure 1.6(b)).

Les résultats indiquent [CI11, CI14] que la différence entre les deux systèmes était perçue par les deux groupes d'auditeurs (Figure 1.6(a)) et significativement plus élevée pour les auditeurs experts que pour les auditeurs naïfs. En termes de préférence, le système B était légèrement mais significativement supérieur au système A (Figure 1.6(b)), sans différence significative entre les deux groupes d'auditeurs [CI11, CI14]. Le mode de restitution (monophonique, stéréophonique, multicanal) n'a pas eu d'effet significatif sur la différence perçue ni sur la préférence.

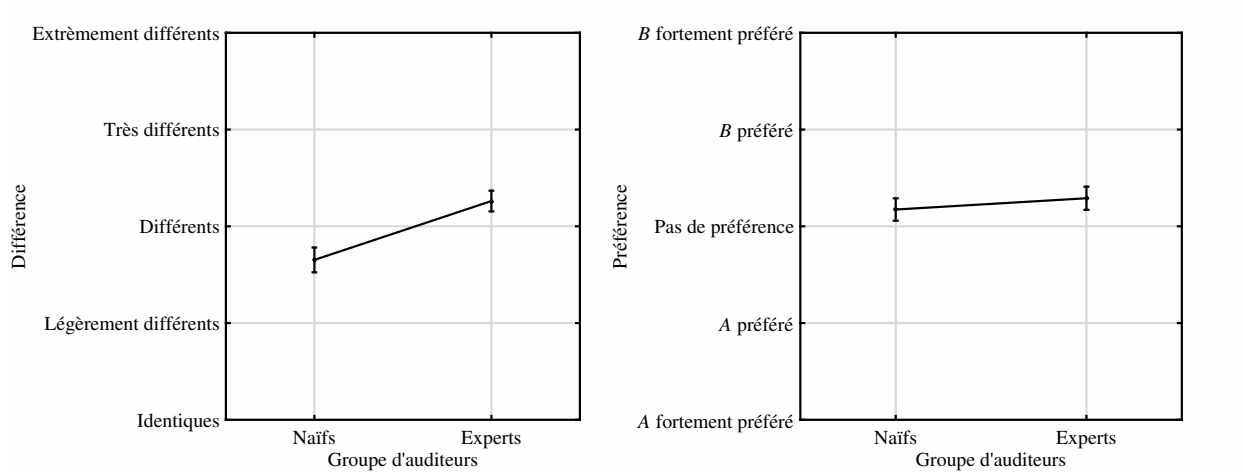


FIGURE 1.6 – Évaluation comparative entre A et B (dans son intervalle de confiance à 95 %) en fonction du groupe d'auditeurs, sur une échelle de différence (a) et de préférence (b).

Ainsi l'utilisation de transducteurs présentant des différences dans leurs caractéristiques et supposément dans leurs qualités perçues donne bien lieu à la perception d'une différence entre les deux systèmes étudiés et d'une légère préférence pour l'un d'eux. Cependant, et bien qu'il s'agisse de jugements relatifs, le dispositif de restitution (les enceintes utilisées, la pièce dans laquelle elles étaient disposées etc.) est toutefois susceptible de masquer des différences ou de niveler des préférences entre les systèmes A et B à comparer. Aussi, une partie des stimuli précédents a été proposée au casque audio (Sennheiser HD 650) afin d'effectuer les comparaisons sur des échelles de différence et de préférence dans un contexte d'écoute différent.

Les signaux initialement décodés pour un rendu sur enceintes ont donc été transformés en signaux binauraux à restituer sur casque à l'aide des fonctions de transfert d'une tête artificielle (modèle KEMAR). Les HRTF (Head-Related Transfer Functions) génériques utilisées correspondaient aux directions des enceintes L, C, R, R_S et L_S (Figure 1.2). Ces stimuli ont ensuite été proposés aux auditeurs experts à fin de comparaison (différence et préférence) entre les systèmes A et B. S'agissant d'écoute au casque, les auditeurs avaient pour consigne de le placer le plus confortablement possible sur leur tête et de ne plus modifier son placement pour toute la durée du test. Le test sur casque a été répété 8 fois afin de prendre en compte la variabilité induite par le positionnement du casque [15].

Une partie des stimuli a donc permis la comparaison des systèmes A et B par les auditeurs experts, à la fois sur enceintes et sur casque audio. Les résultats [CI15] indiquent que les auditeurs experts ont différencié les systèmes A et B de la même manière, qu'ils soient restitués sur enceintes ou au casque (Figure 1.7(a)). Cependant, la préférence vers le système B a été significativement plus marquée lors de la restitution au casque (Figure 1.7(b)).

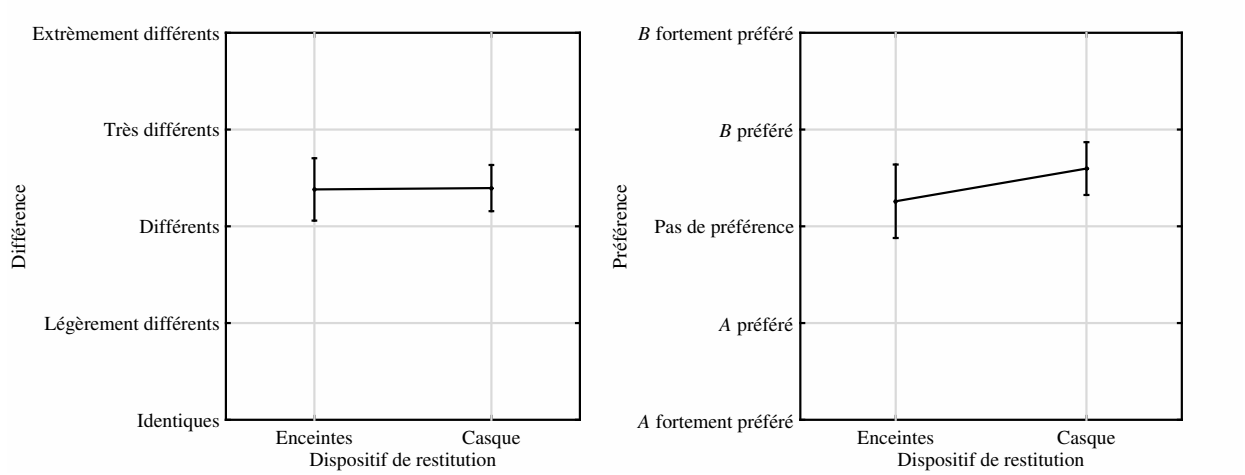


FIGURE 1.7 – Évaluation comparative entre A et B (dans son intervalle de confiance à 95 %) en fonction du dispositif de restitution, sur une échelle de différence (a) et de préférence (b).

Cette préférence confirme et amplifie la tendance déjà observée sur enceintes : la qualité perçue du système B est significativement améliorée par les transducteurs qui le composent.

1.4 Application à un simulateur en réalité virtuelle

Les résultats de ces différents travaux ont fait l'objet d'une mise en oeuvre concrète dans le cadre du projet MARVEST (MARitime Virtual rEality and Simulator Technologies) dont le but était de créer un simulateur de pilotage naval utilisant des technologies destinées à améliorer le réalisme sonore et visuel. Ainsi les réseaux microphoniques développés, basés sur la technologie Higher Order Ambisonics (HOA), ont été utilisés pour une captation sonore spatialisée à l'intérieur d'un navire remorqueur. Des essais préliminaires ont été réalisés en rade de Brest. Le prototype de microphone ambisonique d'ordre 4 [9] a été disposé sur la passerelle d'un remorqueur (Figure 1.8(a)) afin d'effectuer des prises de son spatialisé destinées à être exploitées dans un simulateur de navigation. De telles prises de son présentent l'avantage d'être décodable sur n'importe quel système de restitution sonore (casque ou dispositif multi-enceintes) équipant le simulateur. Les enregistrements finaux, destinés à être restitués dans le simulateur de validation, ont été réalisés sur la passerelle d'un remorqueur de haute mer (Figure 1.8(b)). Le réseau microphonique ambisonique précédemment développé [CI11] y a été placé dans une position représentative de celle du pilote.

(a) (b)

FIGURE 1.8 – Vue de la passerelle du remorqueur Abeille Iroise (a) dans laquelle est disposé un réseau microphonique sphérique [9] et vue extérieure du remorqueur Luberon (b).

Ces enregistrements ont ensuite été restitués et évalués dans le simulateur de démonstration représenté en Figure 1.9 : un environnement virtuel de type CAVE (Cave Automatic Virtual Environment). Le sujet s'y trouve entouré d'écrans permettant un rendu visuel stéréoscopique, la restitution sonore pouvant être effectuée au casque ou par des enceintes placées derrière les écrans. Ce dispositif a permis de valider le réalisme de la mise en situation par un panel de sujets experts (pilotes et instructeurs de pilotage naval).

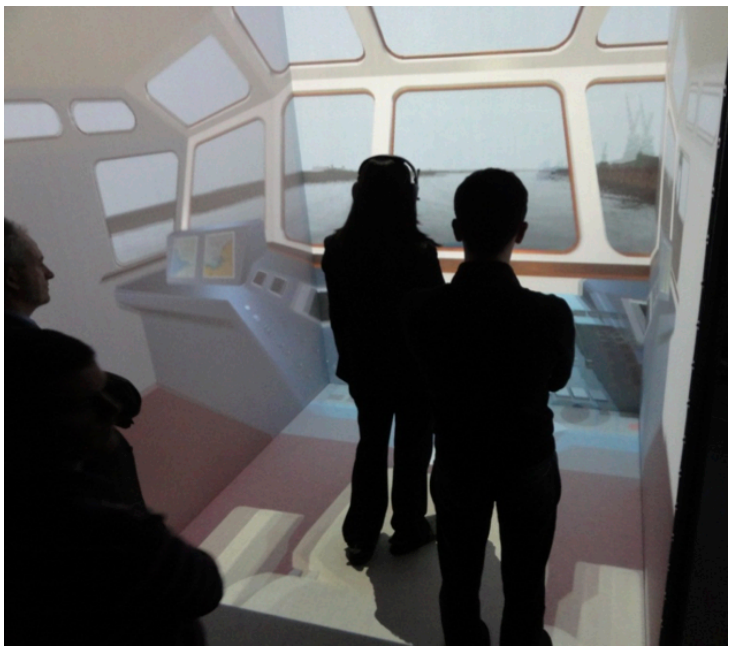


FIGURE 1.9 – Simulateur de démonstration représentant un poste de pilotage naval virtuel.

Chapitre 2

Évaluation perceptive des systèmes de restitution sonore

2.1 Contexte

L'évaluation du son spatialisé sur enceintes ou au casque a permis de soulever la problématique de la qualité intrinsèque d'un système de restitution sonore (spatialisée ou non). Cet axe de recherche a donc eu pour but de définir des protocoles d'évaluation subjective destinés aux enceintes acoustiques ou aux casques audio.

Concernant les enceintes, les résultats expérimentaux ont permis de déterminer des protocoles pour l'évaluation de la qualité perçue [CI24, CI23, AI12]. Ces résultats ont été valorisés lors de contrats de recherche avec Cabasse et Canon Research France (interlocuteurs respectifs : Yvon Kerneis à Brest et Eric N'Guyen à Rennes). Ces contrats consistaient en des études expérimentales et bibliographiques destinées à définir comment évaluer la qualité perçue d'une enceinte, en compensant notamment l'influence de la salle d'écoute dans les basses fréquences. Le prototype d'enceinte évalué dans ce cadre est à ce jour commercialisé : "L'Océan", intégrant un procédé de compensation breveté par Cabasse [16].

Les expériences menées au casque audio ont quant à elles permis de mettre en évidence l'importance de son positionnement sur la tête de l'auditeur dans des applications d'écoute musicale [CN8, AN1, CI16, AI10] comme d'audiométrie [CI12, AI5].

2.2 Évaluation de la qualité perçue d'une enceinte acoustique

L'évaluation de la qualité sonore d'une enceinte acoustique est un processus complexe et les différentes recommandations relatives à l'évaluation de la qualité subjective d'une enceinte – publiées par l'IEC [10], l'ITU [17] ou encore l'AES [18] – ne permettent pas de dégager un consensus sur la procédure à suivre. Elles font cependant émerger un certain nombre de bonnes pratiques à mettre en œuvre pour concevoir un test subjectif fiable et répétable destiné à l'évaluation de la qualité perçue d'enceintes acoustiques à partir d'extraits musicaux :

1. les enceintes à évaluer doivent être cachées pour éviter tout biais visuel ;
2. les extraits musicaux utilisés pour procéder à l'évaluation doivent être diversifiés pour multiplier les contenus temporels et fréquentiels ;
3. le niveau de sortie des enceintes doit être réglé selon les extraits pour proposer des stimuli égalisés en sonie ;
4. les stimuli doivent être relativement courts (pas plus de 10 s par essai) en raison des capacités de la mémoire auditive à court terme ;
5. les jugements de qualité doivent être relatifs plutôt qu'absolus.

En résumé, un exemple de procédure fiable et répétable pour l'évaluation de la qualité subjective d'enceintes acoustiques consisterait à les comparer par paires (5) avec des stimuli consécutifs de 5 s chacun (4) ayant fait l'objet d'une égalisation de sonie (3) propre à chaque extrait utilisé (2) et diffusés par des enceintes cachées (1). Toutes ces préconisations permettent de contrôler les conditions d'écoute du test mais les éloignent du contexte dans lequel peut s'effectuer une écoute musicale sur enceintes, notamment du fait de la durée des stimuli et du niveau d'écoute fixé. L'objectif de cette étude était donc de comparer la procédure décrite plus haut (désignée procédure 1) à des procédures proposant des conditions plus naturelles du point de vue de l'écoute musicale. Ainsi, des extraits plus longs (30 s) ont été proposés aux auditeurs lors d'une procédure 2 en tout point similaire à la procédure 1 hormis le fait que les extraits n'étaient plus écoutés successivement mais alternativement (les auditeurs pouvaient basculer d'une enceinte à l'autre à tout moment). Enfin, une procédure 3, portant également sur les extraits de 30 s en écoute alternative, laissait en plus la possibilité aux auditeurs d'ajuster le niveau d'écoute sur chacune des deux enceintes impliquées dans la comparaison par paire. Les instructions données aux auditeurs lors de cette procédure étaient de procéder au réglage d'un niveau d'écoute confortable sur chacune des deux enceintes de la paire puis d'effectuer la comparaison.

Quatre modèles d'enceintes (désignés A, B, C et D) ont été choisis afin de comparer ces procédures de test. Ces modèles ont été choisis dans la même gamme de prix et en veillant à ce qu'ils possèdent des différences de timbre perceptibles afin de révéler des différences de qualité perçue. Deux de ces enceintes étaient des modèles dits de "monitoring" (de marques PSI et Dynaudio) et deux autres des modèles dits de "haute-fidélité" (de marques Cabasse et Bowers & Wilkins). Les quatre enceintes ont été disposées derrière un écran acoustiquement transparent dans une salle peu réverbérante : le studio d'enregistrement de la formation "Image & Son" de l'Université de Bretagne Occidentale (Figure 2.1). L'évolution du temps de réverbération de cette salle en fonction de la fréquence est en accord avec les préconisations IEC 60268-13 [10] relatives aux tests d'écoute sur haut-parleurs. Toutes les combinaisons d'enceintes sur les quatre positions dans la pièce ont été testées pour compenser les possibles effets de la position. Chaque auditeur testait une de ces 24 combinaisons pour ne pas avoir à modifier les positions au cours du test et chaque combinaison a été testée par deux auditeurs différents soit 48 auditeurs au

total (étudiants en Master ingénierie du son et ingénieurs du son).

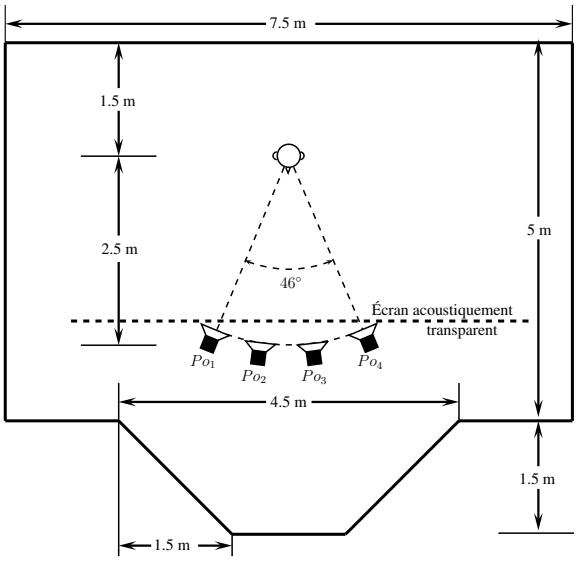


FIGURE 2.1 – Représentation schématique de la salle d'écoute dans laquelle étaient disposés les quatre modèles d'enceintes cachées à l'auditeur par un écran acoustiquement transparent.

Trois extraits musicaux, aux caractéristiques spectrales et dynamiques différentes, ont été utilisés pour comparer ces enceintes :

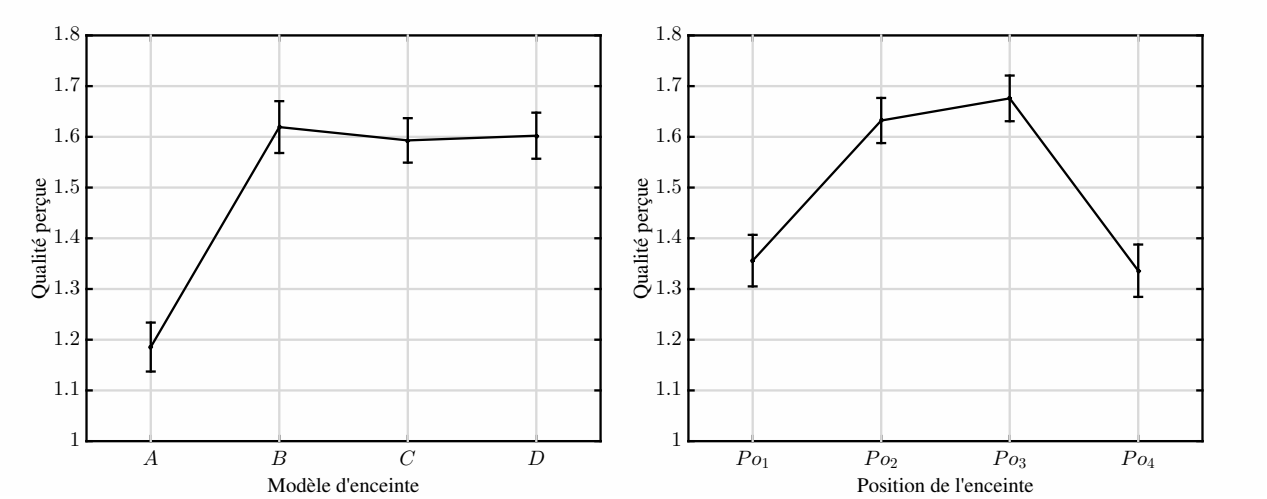
1. orchestre symphonique ;
2. guitare et voix ;
3. piano solo ;

dans des versions courtes (5 s) et longues (30 s) égalisées en sonie par 3 expérimentateurs. Ces extraits ont été utilisés pour évaluer les 4 enceintes (cachées) lors de comparaisons par paire (6 paires par extraits) selon les 3 procédures décrites plus haut et résumées ici :

— Procédure 1 : extraits courts consécutifs, niveau d'écoute égalisé en sonie ;
— Procédure 2 : extraits longs alternatifs, niveau d'écoute égalisé en sonie ;
— Procédure 3 : extraits longs alternatifs, niveau d'écoute libre.

Ces 3 procédures constituaient les 3 sessions, séparées par des pauses et proposées dans un ordre aléatoire, d'un seul et même test durant envion 1h par auditeur. Les résultats des comparaisons par paire ont ensuite été convertis en score de qualité perçue.

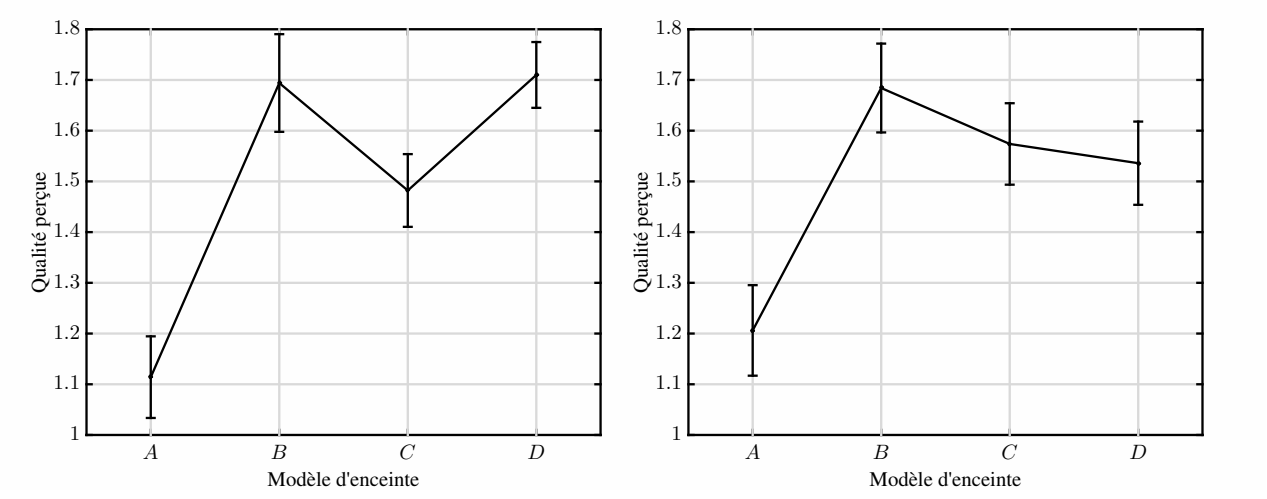
Les résultats ont montré [AI12] que la qualité perçue dépend significativement du modèle d'enceinte ainsi que de sa position dans la pièce. Concernant le modèle d'enceinte, l'un d'eux (A) a systématiquement obtenu une évaluation de qualité significativement plus faible que les trois autres modèles (Figure 2.2(a)), ceux-ci n'étant pas discriminés. Lorsque les enceintes étaient dans les positions excentrées Po_1 et Po_4 telles qu'indiquées sur la Figure 2.1, la qualité perçue a été significativement inférieure à celle obtenue dans les deux positions centrales Po_2 et Po_3 (Figure 2.2(b)).



(a) (b)

FIGURE 2.2 – Qualité perçue moyenne (dans son intervalle de confiance à 95 %) en fonction du modèle d'enceinte (a) et en fonction de la position de l'enceinte.

Certains extraits ou procédures ont cependant permis de révéler des différences de qualité plus marquées entre les enceintes. Ainsi des différences significatives entre les modèles B, C et D ont pu être observées pour l'extrait 3 (Figure 2.3(a)) ou la procédure 3 (Figure 2.2(b)).



(a) (b)

FIGURE 2.3 – Qualité perçue moyenne (dans son intervalle de confiance à 95 %) en fonction du modèle d'enceinte pour l'extrait 3 (a) et pour la procédure 3 (b).

En définitive, la procédure 3 s'est révélée la plus discriminante quel que soit l'extrait, tandis que la procédure 2 n'était plus discriminante que sur 2 extraits parmi 3. La procédure 1 a livré des résultats similaires à ceux observés toutes procédures confondues (Figure 2.2(a)) et n'a donc pas mis en évidence des différences de qualité entre les enceintes B, C et D. Le fait de placer les auditeurs dans des conditions d'écoute plus naturelles du point de vue de l'écoute musicale a donc permis une écoute plus critique.

Une analyse des niveaux a révélé que le réglage dépendait significativement de l'extrait et était systématiquement supérieur au pré-réglage initial. Ainsi le niveau réglé par 3 expérimen-

tateurs ne correspondait au niveau moyen préféré par les auditeurs. En revanche le niveau n'est pas apparu comme dépendant significativement du modèle d'enceinte, indiquant que les quatre modèles ont été écoutés en moyenne au même niveau. Les auditeurs auraient pu avoir tendance à écouter plus fort leurs enceintes préférées et ainsi augmenter encore leurs préférences en comparant des modèles à des niveaux différents, ce qui n'a pas été observé ici.

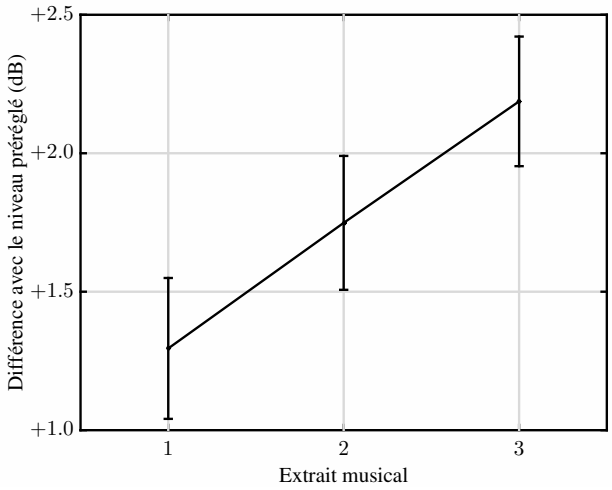


FIGURE 2.4 – Différence moyenne (dans son intervalle de confiance à 95 %) entre le niveau préréglé initialement et le niveau réglé par l'auditeur en fonction de l'extrait musical.

Les conclusions tirées de cette étude sur les procédures d'évaluation ont été mises à profit lors de contrats de recherche avec Canon/Cabasse. Le but n'était alors pas de comparer différents modèles d'enceinte, mais des traitement différents, destinés à compenser la réponse de la salle d'écoute en basse fréquence, dans une zone fréquentielle où la présence de modes propres peut nuire à la qualité perçue. Des tests d'écoute ont été réalisés dans une salle parallélépipédique possédant de nombreux modes pouvant être excités et perçus selon les positions respectives des enceintes et de l'auditeur. Les différents traitements à l'étude ont été comparés en écoutant alternativement des extraits longs (30 s) diffusés par des enceintes cachées. Les résultats de ces tests ont servi de base de travail au développement du dispositif de compensation de salle "Cabasse Room Compensation System" [16] intégré au modèle d'enceinte "L'Océan".

FIGURE 2.5 – Enceintes Cabasse "L'Océan" avec module de compensation de l'acoustique de la salle.

2.3 Audibilité de la variabilité de positionnement d'un casque audio

Les utilisations possibles des casques audio sont nombreuses, allant de l'écoute musicale pour le grand public au monitoring de prise de son et mixage audio pour les ingénieurs du son. Ils permettent également une restitution sonore spatialisée s'appuyant sur la technologie binaurale [19] trouvant ses applications dans la réalité virtuelle et les jeux vidéo. Ils permettent également de conduire des expériences de perception sonore sans qu'il soit nécessaire de disposer d'un environnement à l'acoustique contrôlée.

Cependant, la variabilité inhérente à l'opération de positionnement du casque audio sur la tête de l'auditeur a des conséquences sur la fonction de transfert entre le transducteur et l'oreille, communément désignée HpTF (Headphone Transfer Function). Si les conséquences audibles de cette variabilité ont été peu étudiée, elle peut néanmoins être caractérisée objectivement par la mesure de cette fonction de transfert. Ainsi, des HpTF mesurées successivement présentent une bonne répétabilité en basses fréquences mais des écarts de 10 dB peuvent être relevés pour les fréquences comprises entre 9 et 14 kHz [20]. Tous les modèles de casque n'offrent pas la même répétabilité. Par exemple, les casques circum-auraux (dont les coussins entourent les pavillons de l'auditeur) sont réputés plus répétables dans leur positionnement [21] que les casques supra-auraux (dont les coussins recouvrent les pavillons). Quatre modèles de casque, offrant a priori différents degrés de répétabilité, ont été testés :

— A : Sennheiser HD 497 (supra-aural) ;
— B : Sony MDR–CD580 (circum/supra-aural) ;
— C : Sony MDR–CD2000 (circum/supra-aural) ;
— D : Sennheiser HD 600 (circum-aural).

Les modèles B et C sont constitués de coussins entourant les pavillons de l'auditeur mais revêtus d'un tissu très léger recouvrant les pavillons. Ils ont ainsi été considérés comme des intermédiaires entre circum et supra-aural.

Ces différents modèles de casques ont été repositionnés à 8 reprises sur les oreilles d'une tête artificielle (Neumann KU 100) dont les microphones de mesure sont situés en entrée de conduit auditif bloqué (Figure 2.6). Les casques utilisés vérifiaient tous la condition de couplage équivalent à l'air libre ou Free-air Equivalent Coupling to the ear (FEC), souvent simplement désignés comme casques "ouverts" [22]. Trois signaux sonores ont été diffusés sur chacun des casques et enregistrés par les microphones de la tête artificielle pour chacune des 8 positions :

1. bruit rose ;
2. guitare et voix ;
3. orchestre symphonique.

Les deux extraits musicaux (signaux 2 et 3) ont été choisis parmi ceux qui avaient été utilisés pour l'évaluation de la qualité des enceintes [AI12].

Ces enregistrements, représentatifs des différentes positions du casque sur la tête artificielle, ont par la suite été diffusés sur un seul casque dans le but d'évaluer si ces différents positionnements étaient responsables de différences audibles. Le casque C a été utilisé à cet effet. Les mesures effectuées sur ce modèle ont révélé des HpTF large-bande stables et sa réponse en fréquence a été compensée avant restitution des enregistrements [AI10]. Les stimuli ont été proposés à 10 sujets "experts" [11] selon une procédure à choix forcé présentant 3 alternatives parmi

FIGURE 2.6 – Casque Sennheiser HD 600 (modèle D) positionné sur la tête artificielle Neumann KU 100.

3 intervalles : 3-Interval 3-Alternative Forced Choice (3I3AFC). À chaque essai, 1 stimulus était différent des 2 autres car provenant d'un enregistrement effectué dans une autre position du casque sur la tête. L'auditeur devait indiquer lequel des 3 stimuli était différent des deux autres afin de déterminer si les auditeurs pouvaient détecter les différences dues au positionnement du casque.

Les résultats montrent [AI10] que le taux de détection du stimulus différent varie significativement selon le modèle de casque audio (Figure 2.7(a)) et selon le contenu sonore (Figure 2.7(b)). Ces taux sont par ailleurs tous significativement supérieurs à 33.33 %, le taux qui serait atteint en répondant au hasard dans une tâche de discrimination 3I3AFC.

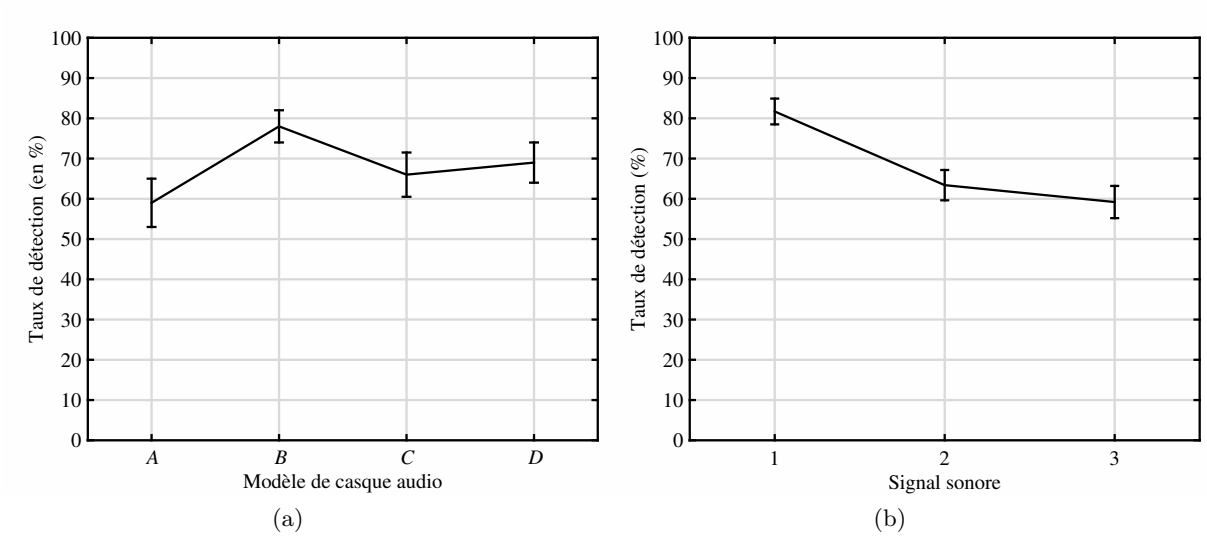
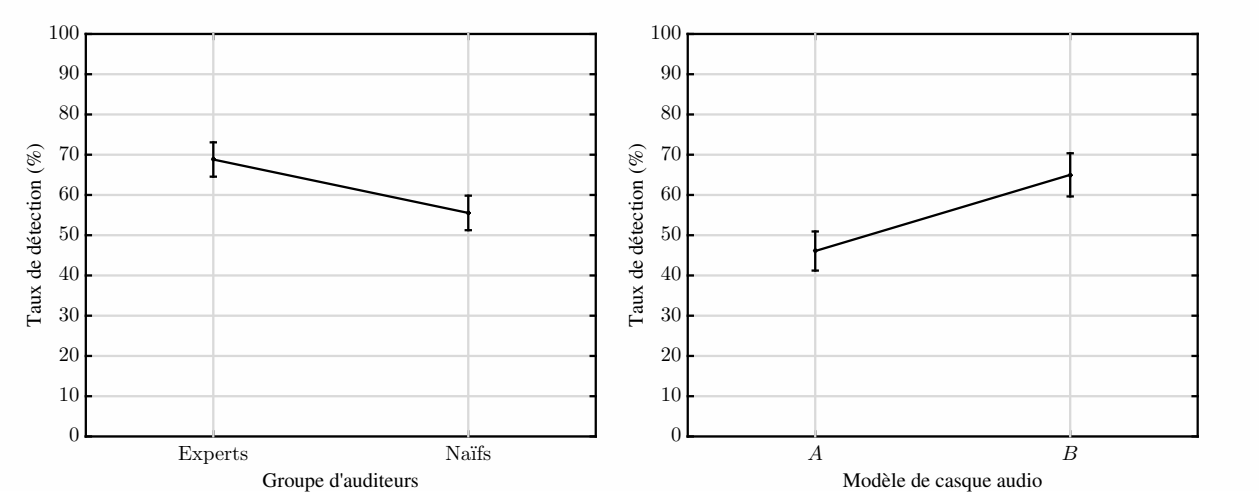


FIGURE 2.7 – Taux de détection moyen (dans son intervalle de confiance à 95 %) en fonction du modèle de casque audio (a) et du signal sonore (b).

Les différences de positionnement du casque sur la tête artificielle ont donc engendré des différences dans les stimuli clairement audibles par les sujets experts. Afin de vérifier si ces différences sont audibles ou non par des sujets moins discriminants, les modèles de casque A et B, offrant respectivement les taux de reconnaissance le plus faible et le plus élevé ont été proposés à 10 sujets "naïfs" [11] selon la même procédure 3I3AFC.



(a) (b)

FIGURE 2.8 – Taux de détection moyen (dans son intervalle de confiance à 95 %) pour les modèles de casque A et B en fonction du groupe d'auditeurs (a) et en fonction du modèle pour les auditeurs naïfs uniquement (b).

Les résultats montrent [AI10] que les auditeurs naïfs ont un taux de détection significativement plus faible que les experts (Figure 2.8(b)) mais tout de même supérieur à 33% pour chacun des deux casques testés (Figure 2.8(b)). Ainsi, la variabilité de positionnement de ces différents modèles de casque audio a généré des différences audibles quel que soit le degré d'expertise des auditeurs. Si cette variabilité porte peu à conséquence lorsque les casques sont utilisés à des fins d'écoute musicale, ses répercussions pourraient être beaucoup plus importantes dans des applications audiométriques telles que la mesure des seuils auditifs.

L'effet de la position du casque sur la mesure du seuil d'audition a donc été étudié pour deux modèles de casque :

— Sennheiser HD 600 (circum-aural), utilisé dans de nombreux tests perceptifs et dont la variabilité de positionnement a des conséquences audibles à niveau supraliminaire [AI10] ;

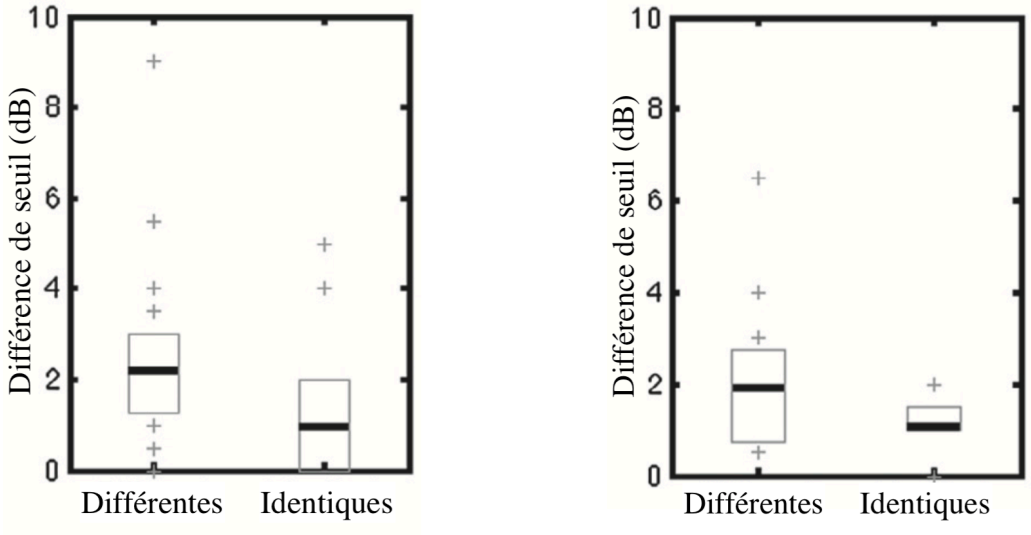
— Telephonics TDH39, couramment utilisé en audiométrie.

Ces deux modèles ont été utilisés pour mesurer les seuils d'audition de 20 sujets naïfs [11] normo-entendants selon la procédure standard ANSI/ASA S3.21 [23] destinée à l'audiométrie tonale. Les seuils ont donc été mesurés pour des sons purs aux fréquences centrales des bandes d'octave de 250 à 8000 Hz classiquement explorées ainsi qu'aux fréquences additionnelles de 125, 6000, 11000 et 14000 Hz.

La mesure des seuils étant une procédure intrinsèquement variable, elle a été conduite deux fois consécutivement en demandant au sujet de conserver le casque sur sa tête afin d'estimer la variabilité de mesure du seuil. Cette mesure a également été conduite deux fois consécutivement

en demandant au sujet d'enlever puis repositionner le casque sur sa tête afin de vérifier si la variabilité de positionnement augmentait significativement la variabilité de mesure du seuil. Les résultats ont ainsi été analysés en termes de différence entre deux mesures de seuil consécutives. Ces mesures consécutives ont été effectuées dans un ordre aléatoire pour deux positions identiques ou différentes du casque.

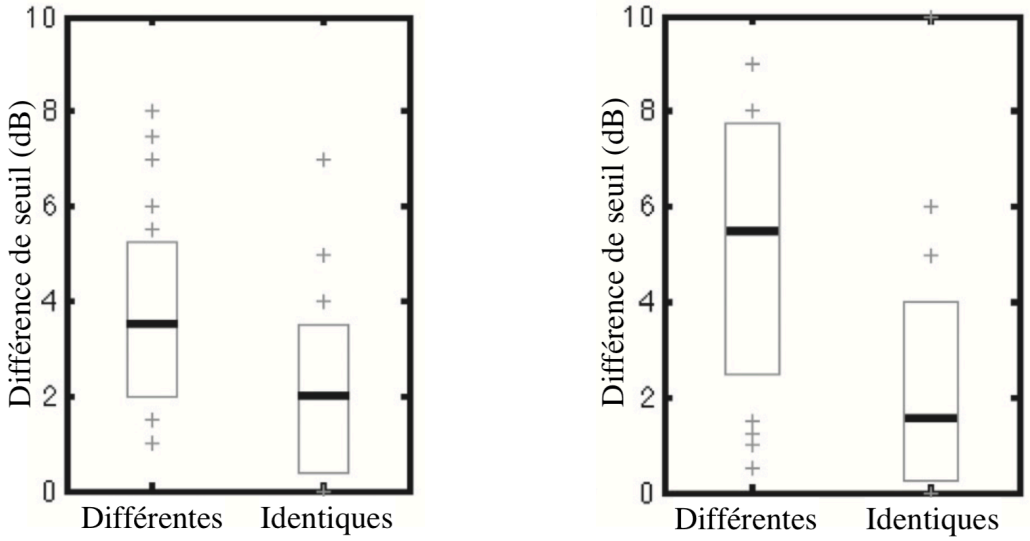
Les résultats indiquent [AI5] que le repositionnement du casque sur la tête de l'auditeur a un effet significatif sur la variabilité (différence entre deux mesures consécutives) du seuil, mais pour certaines fréquences seulement. Concernant le casque HD 600, la différence entre deux positions différentes se révèle ainsi significativement supérieure à la différence entre deux positions identiques à 2000 Hz (Figure 2.9(a)) et 11000 Hz (Figure 2.9(b)).



(a) (b)

FIGURE 2.9 – Différence médiane (dans son diagramme en boîte) entre deux seuils mesurés consécutivement en fonction des deux positions (différentes ou identiques) du casque HD 600, à 2000 Hz (a) et 11000 Hz (b) [AI5].

Les mesures de seuil effectuées sur le casque TDH39 indiquent quant à elles que la différence entre deux positions différentes se révèle significativement supérieure à la différence entre deux positions identiques à 4000 Hz (Figure 2.10(a)) et 6000 Hz (Figure 2.10(b)).



(a) (b)

FIGURE 2.10 – Différence médiane (dans son diagramme en boîte) entre deux seuils mesurés consécutivement en fonction des deux positions (différentes ou identiques) du casque TDH39, à 4000 Hz (a) et 6000 Hz (b) [AI5].

Ainsi, même si seulement 2 fréquences sur 12 testées sont concernées pour chaque casque, ces résultats attestent que le repositionnement d'un casque sur la tête d'un auditeur engendre des différences significatives dans la mesure du seuil d'audition.

Chapitre 3

Modélisation de la qualité vocale en téléphonie mobile

3.1 Contexte

Mes travaux de recherche se sont également orientés vers la perception d'un stimulus sonore aux caractéristiques particulières : la voix humaine ; dans un contexte particulier : la téléphonie mobile. Ces travaux ont fait l'objet d'une collaboration avec la Technische Universität Berlin (interlocuteurs : Sebastian Möller, Alexander Raake) ayant pour contexte la thèse de Nicolas Côté [24], financée par Deutsche Telekom Laboratories (T-Labs) en partenariat avec Orange Labs (Lannion, interlocutrice : Valérie Gauthier-Turbin). Ce travail avait pour but l'élaboration d'un modèle destiné à évaluer la qualité perçue de signaux de parole transmis par voie téléphonique super large bande et s'inscrivait dans le cadre d'un processus de normalisation initié par l'ITU (International Telecommunication Union).

L'objectif était de modéliser la perception subjective de la qualité vocale en se basant sur différentes dimensions perceptives et d'en déduire des outils de prédiction de cette qualité [CI22, CI19, CI18]. Cette approche a permis de développer le modèle DIAL (Diagnostic Instrumental Assessment of Listening quality) qui a fait l'objet d'une publication dans une revue internationale avec comité de lecture [AI13] et qui a servi de base de travail à la norme ITU-T P.863 [25] destinée à l'évaluation de la qualité perçue en téléphonie mobile super large bande.

Le modèle DIAL a également fait l'objet d'un contrat de recherche en 2010 avec Orange Labs (Lannion, interlocutrice : Valérie Gauthier-Turbin) en vue de son déploiement vers les réseaux mobiles Orange.

3.2 Approche multidimensionnelle de la qualité vocale

3.2.1 Architecture du modèle DIAL

Le modèle DIAL (Diagnostic Instrumental Assessment of Listening quality) est destiné à l'estimation de la qualité globale de la voix transmise en téléphonie mobile super large bande (*super-wideband*, dont la bande passante est contenue entre 50 et 14000 Hz). Il s'agit d'un modèle "intrusif" nécessitant la connaissance du signal de parole en entrée $x(k)$ et du signal transmis correspondant $y(k)$. Le fonctionnement de ce modèle est schématiquement représenté en Figure 3.1 et décrit ci-dessous. Un pré-traitement inspiré du modèle PESQ (Perceptual Evaluation of Speech Quality [26]), consistant essentiellement en un alignement temporel, est initialement appliqué pour obtenir les signaux $x'(k)$ et $y'(k)$ qui seront utilisés pour les estimations de qualité.

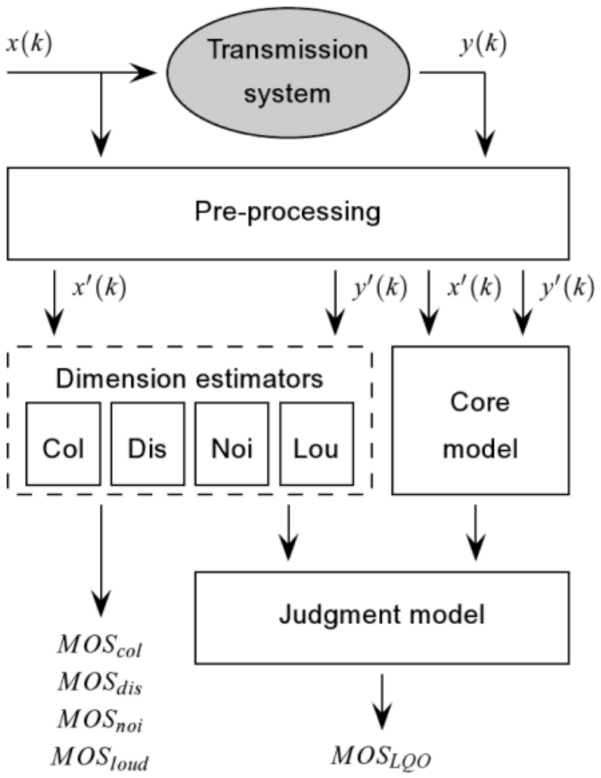


FIGURE 3.1 – Représentation schématique des principaux principes de fonctionnement du modèle DIAL [24].

L'originalité du modèle DIAL réside dans une estimation de qualité selon chacune des dimensions perceptives qui sous-tendent le jugement de qualité globale. Dans le contexte de la téléphonie mobile super large bande, l'espace perceptif de la qualité vocale est composé des 3 dimensions suivantes [27] :

— la dimension "discontinuité" (*discontinuity*) décrit les dégradations dues aux interruptions typiques dans les réseaux mobiles ;

— la dimension "bruyance" (*noisiness*) prend en compte les différents types de bruits venant s'additionner au signal d'entrée (bruit dû à l'environnement du locuteur, bruit analogique dû aux transducteurs, bruit numérique dû aux codecs) ;

— la dimension “coloration” (*coloration*) représente les modifications fréquentielles que subit le signal entre la bouche du locuteur et l'oreille de l'auditeur (essentiellement dues à l'acoustique de la pièce où le signal est émis et à la bande passante du système de transmission).

Le modèle intègre différents estimateurs [28, 29] qui permettant l'estimation de chacune de ces dimensions d'après les signaux pré-traités $x'(k)$ et $y'(k)$.

À ces trois dimensions vient s'ajouter la sonie (*loudness*), qui n'est pas à proprement parler une dimension de l'espace perceptif mais qui affecte néanmoins la qualité globale [30]. La sonie à long terme est estimée par un modèle destiné aux sons non-stationnaires [31].

Le modèle permet donc d'estimer chacune de ces dimensions et de leur d'attribuer un score de qualité MOS (Mean Opinion Score [32]), respectivement : $\mathrm{MOS}_{\mathrm{dis}}$, $\mathrm{MOS}_{\mathrm{noi}}$, $\mathrm{MOS}_{\mathrm{col}}$ et $\mathrm{MOS}_{\mathrm{lou}}$. De manière très schématique, le score de qualité selon chaque paramètre est déterminé par comparaison avec des valeurs de référence issues de tests subjectifs [CI22].

La qualité est également estimée par un modèle “core” utilisé pour estimer l'effet de dégradations non-linéaires typiques liées à l'utilisation de codecs. Ce modèle consiste en une modification du modèle intrusif TOSQA (Telecommunication Objective Speech Quality Assessment [33]) qui estime directement un score de qualité $\mathrm{MOS}_{\mathrm{core}}$ d'après les dégradations non-linéaires observées entre le signal d'entrée $x'(k)$ et le signal de sortie $y'(k)$.

Ces cinq scores MOS sont ensuite agrégés par un modèle cognitif de jugement [CI18] qui détermine l'influence de chacun de ces paramètres perceptifs dans la qualité globale. Le score $\mathrm{MOS}_{\mathrm{LQO}}$ de qualité globale ou de “qualité d'écoute objective” (*Listening Quality Objective*) consiste alors schématiquement en une moyenne pondérée. Le modèle permet en outre d'obtenir des informations de diagnostic de cette qualité globale, par le biais des dimensions qui la composent. Il peut ainsi fournir des pistes d'amélioration en révélant, par exemple, la faiblesse d'un système de transmission selon une de ces dimensions.

3.2.2 Validation du modèle DIAL

La validation du modèle a été effectuée en comparant les scores de qualité “objectifs” estimés par le modèle à des scores de qualité “subjectifs” (*Listening Quality Subjective*) issus de tests perceptifs, respectivement désignés $\mathrm{MOS}_{\mathrm{LQO}}$ et $\mathrm{MOS}_{\mathrm{LQS}}$. Le modèle DIAL a donc permis de calculer les scores $\mathrm{MOS}_{\mathrm{LQO}}$ de signaux provenant de 55 bases de données (représentatives des dégradations typiques, des applications possibles et des langues utilisées dans les réseaux mobiles) pour lesquelles les scores $\mathrm{MOS}_{\mathrm{LQS}}$ étaient disponibles. Les performances prédictives du modèle DIAL ont été estimées d'après deux mesures d'accord entre les scores $\mathrm{MOS}_{\mathrm{LQO}}$ et $\mathrm{MOS}_{\mathrm{LQS}}$:

— le coefficient de corrélation de Pearson ρ ;

— la racine de l'erreur quadratique moyenne σ.

Le modèle DIAL affiche en moyenne une meilleure corrélation $\rho = 0.935$ ($\sigma = 0.17$) que les modèles existants (PESQ [26], Enhanced PESQ [34], TOSQA [33] etc.). La Figure 3.2 indique par exemple la correspondance entre scores subjectifs ($\mathrm{MOS}_{\mathrm{LQS}}$) et objectifs ($\mathrm{MOS}_{\mathrm{LQO}}$) sur une de ces bases de données comprenant un corpus de 68 sons ($\rho = 0.89$, $\sigma = 0.19$).

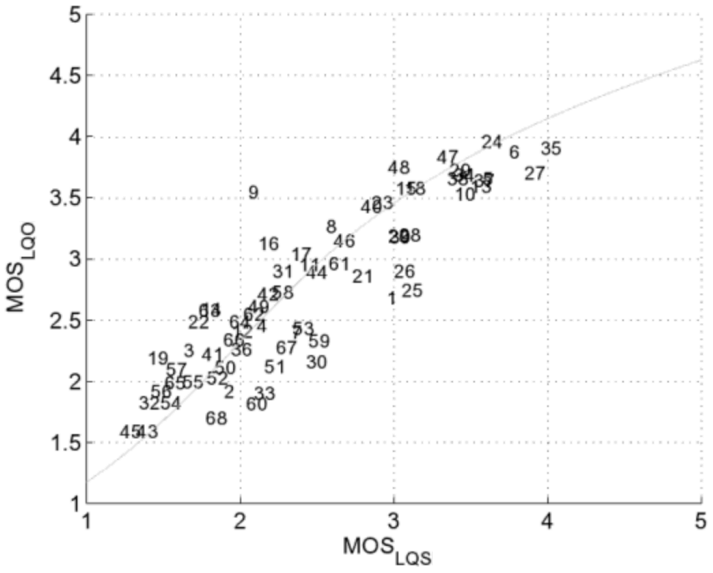


FIGURE 3.2 – Score MOS_{LQO} en fonction du score MOS_{LQS} sur un corpus représentatif de 68 sons et interpolation polynomiale de degré 3 [24].

Par ailleurs, l'approche diagnostique du modèle permettant d'attribuer un score de qualité MOS selon chacune des quatre "dimensions" (MOS_{dis}, MOS_{noi}, MOS_{col} et MOS_{lou}) offre également des pistes d'amélioration à celui-ci. Des tests perceptifs ont été réalisés afin d'obtenir des estimations subjectives directes de ces quatre dimensions [35, AI13] : 76 signaux de parole ont été sélectionnés pour définir un corpus représentatif de la dynamique de ces dimensions. Ces stimuli ont été évalués par 24 auditeurs naïfs sur des échelles continues (allant par exemple de "continue" à "discontinue" sur la dimension *discontinuity*).

Ces mesures subjectives ont à nouveau permis d'estimer les performances prédictives du modèle DIAL (d'après ρ et σ) sur chacune des quatre dimensions perceptives. Les résultats ont révélé [AI13] que le modèle pouvait notamment être amélioré dans l'estimation de la dimension *noisiness* qui affiche la plus faible corrélation avec les données subjectives (Table 3.1).

Dimension	ρ	σ
discontinuity	0.854	0.438
noisiness	0.755	0.606
coloration	0.897	0.356
loudness	0.905	0.253

TABLE 3.1 – Coefficient de corrélation ρ et moyenne de l'erreur quadratique σ entre les scores MOS (MOS_{dis}, MOS_{noi}, MOS_{col} et MOS_{lou}) issus du modèle DIAL et ces dimensions évaluées lors de de tests perceptifs [AI13].

Chapitre 4

Interactions audiovisuelles

4.1 Contexte

Les expériences menées dans des applications de réalité virtuelle (dans le cadre du projet MARVEST notamment) ont fait émerger un autre axe de recherche : les interactions multi-sensorielles, en l'occurrence audiovisuelles. La présence de stimuli visuels peut ainsi modifier la perception sonore et donner lieu, par exemple, à des variations significatives de niveau sonore ressenti [AI15].

Le projet "Usage et perception du son spatialisé dans un contexte de réalité virtuelle" (post-doctorat de Nicolas Côté [36]) a permis de mettre en évidence les interactions entre les modalités auditive et visuelle dans la perception de la distance [CI13, CI10, AI4].

Puis le projet CCFL2 (Cross-Channel Film Lab 2), au cours duquel s'est déroulée la thèse d'Etienne Hendrickx [37], a eu pour but de déterminer comment adapter le son spatialisé au cinéma en "3D relief" (ou 3D stéréoscopique). Les études menées dans le cadre de ce projet ont ainsi pu mettre en évidence le fait que la stéréoscopie induisait chez les sujets de nouvelles attentes en termes d'enveloppement sonore [CI9, CN6, AI11, CI8], la cohérence audiovisuelle étant bénéfique quel que soit le mode de projection (classique ou stéréoscopique) [AI7, AI8, CN4].

Enfin, le projet EDISON 3D (Édition et DIffusion SONore spatialisée en 3 Dimensions), dans le cadre duquel Julian Palacino a effectué son post-doctorat, a exploré des problématiques similaires dans un contexte de concert sonorisé. Les sources sonores étaient spatialisées selon la technique Wave Field Synthesis (WFS) permettant une plus grande cohérence audiovisuelle [CI6, CI7, CI4]. Cette étude a fait l'objet d'une collaboration avec Sonic Emotion Labs pour l'aspect WFS (interlocuteur : Etienne Corteel) et Radio France pour l'aspect sonorisation (interlocuteur : Frédéric Changenet).

4.2 Perception sonore et visuelle de la distance dans un environnement virtuel

Un auditeur peut estimer la distance le séparant d'une source sur la base d'indices auditifs (niveau sonore, atténuation des hautes fréquences, rapport champ direct à champ réverbéré...) et/ou visuels (perspective linéaire et de texture, taille et occlusion des objets, disparité binoculaire...) procurés par celle-ci [36]. Le but de cette étude était d'étudier l'interaction entre ces deux modalités dans un environnement virtuel. Cette expérience s'est déroulée dans une salle d'expérimentation du Centre Européen de Réalité Virtuelle (CERV), une plate-forme expérimentale du Lab-STICC située à Plouzané. Le sujet faisait face à un écran projetant la continuité virtuelle de la salle réelle dans laquelle il était installé (Figure 4.1).

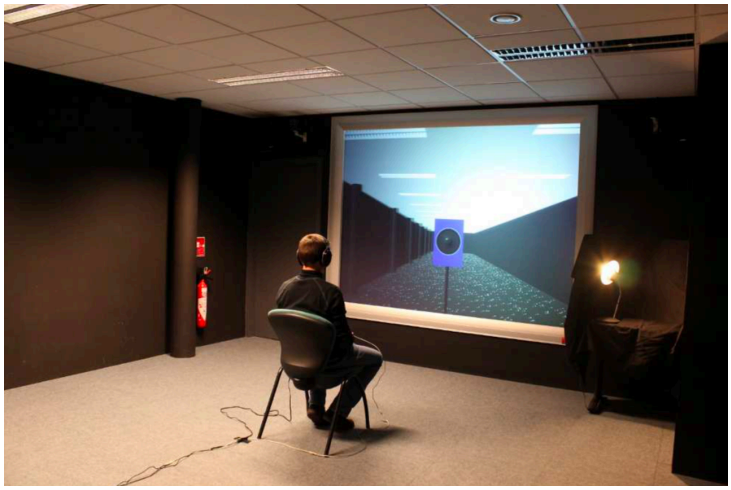


FIGURE 4.1 – Sujet faisant face à l'écran dans la salle expérimentale du CERV.

Cet écran se trouvait donc à l'interface entre l'environnement réel dans lequel se trouvait le sujet et un environnement virtuel permettant de représenter la continuité de la salle expérimentale dans laquelle se trouverait une source (un haut-parleur). Cette source virtuelle pouvait ainsi être placée à différentes distances (allant de 2 à 20 m) du sujet (Figure 4.2). Ce dernier avait alors pour tâche d'estimer la distance de la source d'après les indices auditifs et visuels disponibles.

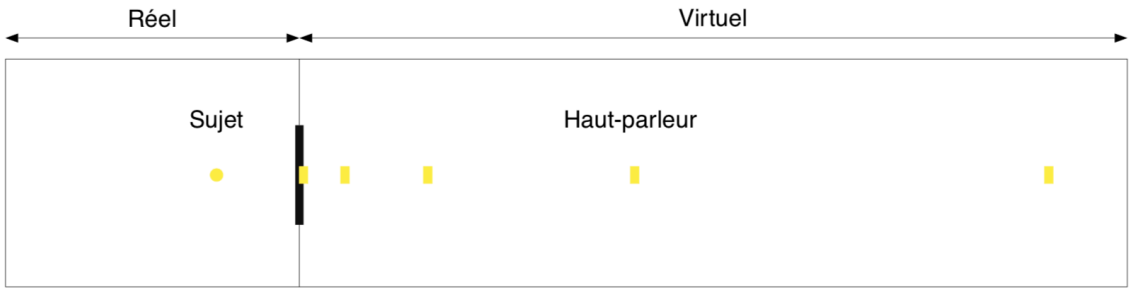


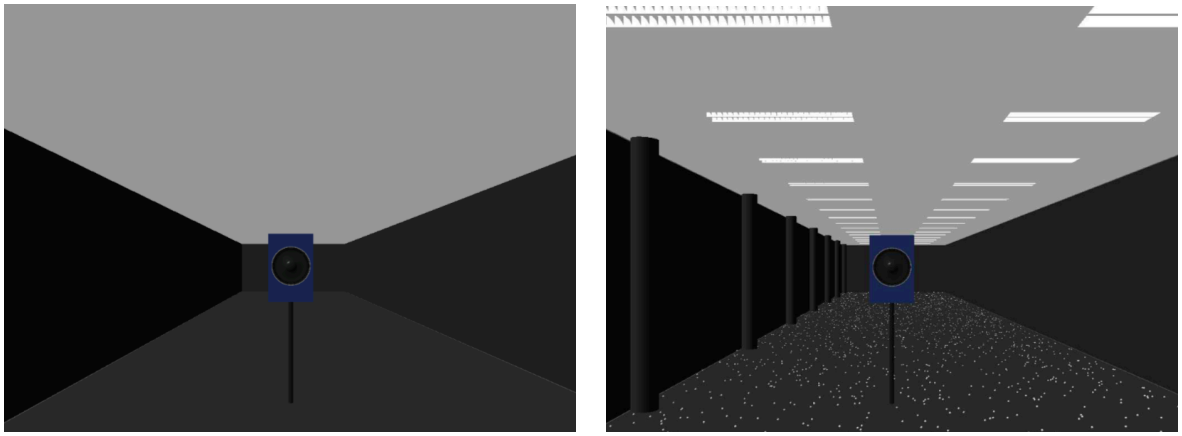
FIGURE 4.2 – Représentation schématique de l'environnement réel dans lequel se trouve le sujet et de son prolongement virtuel dans lequel se trouve le haut-parleur au-delà de l'écran.

La restitution visuelle de la source dans son environnement virtuel était effectuée par deux

vidéo-projecteurs (Barco 3D video package) situés dans un local technique derrière l'écran et permettant la rétro-projection d'une image stéréoscopique. Les vidéo-projecteurs étaient équipés de filtres polarisants, le sujet portant des lunettes passives. Deux configurations d'environnement virtuel visuel ont été testées :

— un environnement "pauvre" en indices visuels comprenant les murs, le plafond et le sol, apportant principalement des informations au niveau de la perspective linéaire (Figure 4.3(a)) ;

— un environnement "riche" en indices visuels comprenant, en plus des indices précédents, une texture appliquée au sol, des colonnes et des néons, apportant des informations au niveau de la perspective de texture et de la taille des objets (Figure 4.3(b)).



(a) (b)

FIGURE 4.3 – Environnement visuel virtuel proposant différents degrés d'informations visuelles : "pauvre" (a) et "riche" (b).

La restitution sonore de la source dans son environnement virtuel était effectuée par un casque audio (Sennheiser HD 650). Ce casque a permis de restituer aux oreilles du sujet des stimuli binauraux correspondant au signal délivré par l'enceinte (de la voix parlée) dans cet environnement virtuel. Le signal initialement anéchoïque a été convolué par les réponses impulsionnelles binaurales de la salle étudiée. Ces BRIR (Binaural Room Impulse Response) ont été synthétisées à partir des réponses impulsionnelles d'une tête artificielle (modèle KEMAR) auxquelles ont été ajoutées des réflexions précoces puis diffuses. Celles-ci ont été respectivement simulées pour les différentes configurations de la source dans la salle représentée au sujet [38] et extraites d'une base de données [39]. Ce procédé a permis, à l'instar de l'environnement visuel, de simuler et tester deux configurations d'environnement sonore virtuel :

— un environnement "mat" possédant un temps de réverbération relativement court (0.370 s), apportant essentiellement des informations de niveau sonore ;

— un environnement "réverbérant" possédant un temps de réverbération relativement long (0.860 s), apportant de surcroît des informations de rapport champ direct à champ diffus.

Les sujets devaient reporter à l'aide d'un pavé numérique la distance perçue ρ_{per} pour des distances cibles ρ_{cib} (2, 3, 5, 10 et 20 m) restituées à partir de stimuli auditifs uniquement, visuels uniquement ou audiovisuels. 24 sujets naïfs [11] ayant déclaré avoir une audition normale et une vue normale ou corrigée ont participé à cette expérience.

En accord avec la littérature sur le sujet [40], les résultats indiquent [AI4] que la distance estimée sur la base d'indices auditifs seuls est quasi-systématiquement sous-estimée (Figure 4.4). Cette sous-estimation augmente avec la distance quel que soit le temps de réverbération. La distance estimée est significativement plus proche de la distance cible pour le temps de réverbération le plus élevé (0.860 s) avec même une légère sur-estimation pour les distances faibles.

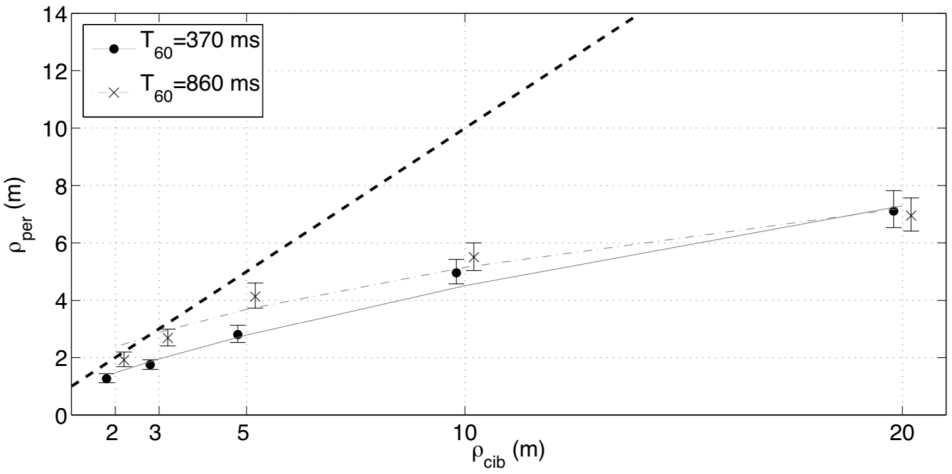


FIGURE 4.4 – Distance perçue moyenne ρ_{per} (dans son intervalle de confiance à 95 %) en fonction de la distance cible ρ_{cib} pour les stimuli auditifs : TR = 370 ms (trait plein) et TR = 860 ms (trait mixte) [36].

En présence d'indices visuels uniquement, la distance perçue est également sous-estimée même si plus proche de la distance cible qu'en présence d'indices auditifs uniquement (Figure 4.5). Cette sous-estimation évolue alors linéairement avec la distance et les deux environnements visuels proposés n'ont pas donné lieu à des différences significatives dans les estimations de distance.

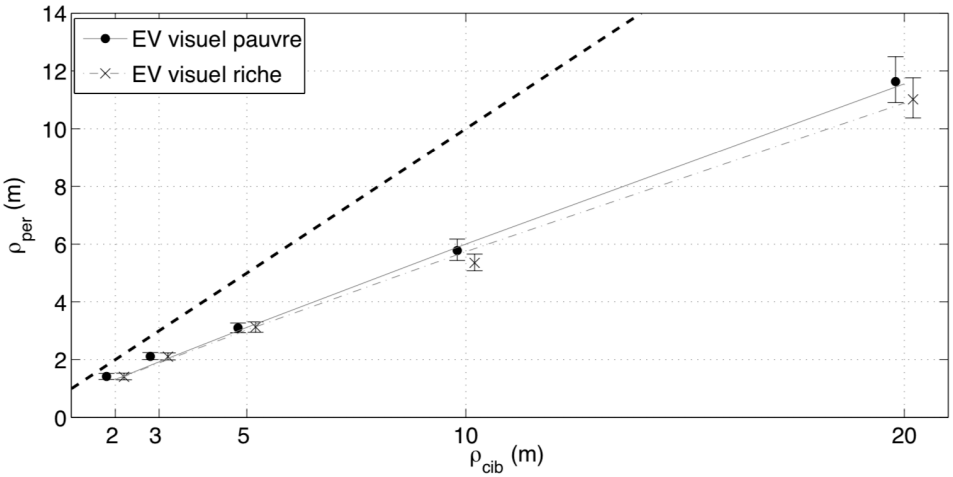


FIGURE 4.5 – Distance perçue moyenne ρ_{per} (dans son intervalle de confiance à 95 %) en fonction de la distance cible ρ_{cib} pour les stimuli visuels : environnement virtuel visuel "pauvre" (trait plein) et "riche" (trait mixte) [36].

Enfin, en présence d'indices auditifs et visuels, la distance perçue est une nouvelle fois sous-estimée, dans toutes les configurations d'environnement visuel et de réverbération (Figure 4.5). Cette sous-estimation évolue à nouveau linéairement avec la distance et est à peine plus importante que celle observée en présence d'indices visuels uniquement, indiquant une prévalence de la modalité visuelle sur la modalité auditive dans l'estimation de la distance. Cette observation se révèle semblable à un effet de "capture visuelle" reporté pour des sources dans un environnement réel [41].

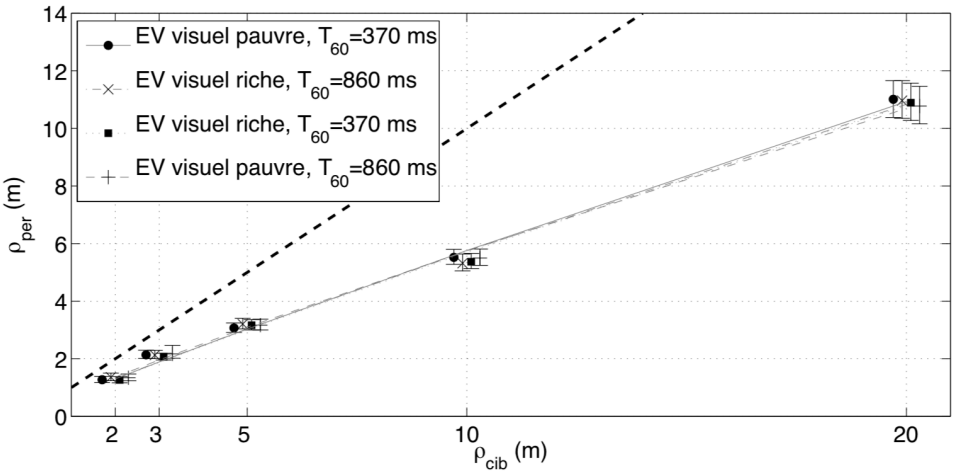


FIGURE 4.6 – Distance perçue moyenne ρ_{per} (dans son intervalle de confiance à 95 %) en fonction de la distance cible ρ_{cib} pour les stimuli audiovisuels : environnement virtuel visuel "pauvre" avec TR = 370 ms (trait plein), visuel "riche" avec TR = 860 ms (trait mixte), visuel "riche" avec TR = 370 ms (trait pointillé) et visuel "pauvre" avec TR = 860 ms (trait tireté) [36].

En définitive, les estimations de distance observées dans un environnement virtuel en présence d'indices auditifs et/ou visuels sont conformes aux observations effectuées en environnement réel [40, 41].

4.3 Perception sonore dans un contexte de cinéma 3D stéréoscopique

La bande-son d'un film destiné au cinéma est souvent réalisée dans un auditorium de mixage équipé d'un système multicanal "5.1" pour la diffusion sonore. Conformément à la recommandation ITU-R BS.775-3 [3], cette bande-son est alors composée de :

— 3 canaux frontaux $L - C - R$ (*Left – Center – Right*) ;

— 2 canaux arrières dits "surround" $L_S - R_S$ ($Left_{Surround}$ – $Right_{Surround}$) ;

— 1 canal LFE (*Low-Frequency Effects*) destiné aux basses fréquences.

Chaque canal est basiquement restitué par une enceinte dédiée. Lors de la diffusion dans un cinéma, les deux canaux "surround" peuvent être restitués respectivement par des réseaux d'enceintes dédiés tandis que plusieurs caissons de graves peuvent être utilisés pour le canal LFE, selon le volume du lieu.

La généralisation de ce dispositif a eu pour conséquence d'uniformiser les pratiques de mixage et de spatialisation du son. Par exemple, les objets sonores (dialogues et effets ponctuels tels que bruits de pas, claquement de porte...) sont conventionnellement diffusés par le canal central C uniquement [42]. Les musiques et les sons d'ambiance sont en général diffusés dans les enceintes avant gauche et avant droite (L et R), ainsi qu'à moindre volume dans les enceintes "surround" [37]. Cette approche du mixage orientée "canal" est désormais remise en cause par de nouveaux procédés tels que la WFS ou le Dolby Atmos qui permettent une approche dite "objet" [43].

L'avènement dans les années 2000 du cinéma 3D stéréoscopique (aussi appelé 3D relief), destiné à offrir une expérience visuelle plus immersive au spectateur, a néanmoins amené les ingénieurs du son à remettre en cause leurs pratiques de mixage pour proposer une expérience sonore elle aussi plus immersive.

Ce travail a donc eu pour objectif de déterminer si les attentes en termes d'enveloppement et de spatialisation des objets sonores étaient les mêmes selon que la bande-son était associée à la projection d'une image classique ou stéréoscopique (respectivement désignées 2D et 3D). Ainsi, l'enveloppement a été étudié du point de vue de l'équilibre avant/arrière des sons d'ambiance (pour des sujets en situation de mixeur mais aussi de spectateur). La spatialisation des objets sonores a quant à elle été étudiée du point de vue de la cohérence audiovisuelle en azimut et en profondeur.

4.3.1 Influence de l'image stéréoscopique sur la perception des sons d'ambiance

Une expérience préliminaire a été mise en place pour vérifier si les stratégies de mixage sonore se révèlent différentes selon que l'image associée est projetée en 2D ou en 3D. Celle-ci a été menée dans l'auditorium de mixage de la formation "Image & Son Brest" de l'UBO. Les 11 sujets de cette expérience étaient des étudiants-ingénieurs du son (de niveau Master) placés dans un rôle de mixeur son pour 11 courtes (environ 30 s) séquences audiovisuelles (image 2D/3D et son multicanal). La plupart de ces séquences ont été spécialement filmées (à l'aide d'une caméra Panasonic AG-3DP1) pour les besoins de cette expérience (exemples en Figure 4.7), dans le cadre du projet CCFL2.

(a) (b)

FIGURE 4.7 – Exemples de séquences audiovisuelles utilisées pour la perception des sons d'ambiance [37].

Les séquences audiovisuelles étaient projetées en version stéréoscopique (3D) et classique (2D), le sujet étant équipé de lunettes 3D actives. Un mixage sonore initial a été réalisé par les expérimentateurs pour chaque séquence et restitué sur un dispositif multicanal classique (3 enceintes frontale et deux enceintes surround [3]). Ce mixage proposait une répartition initiale des sons d'ambiance entre les enceintes avant ($L - R$) et arrière ($L_S - R_S$). La tâche des sujets consistait en l'ajustement de la balance avant/arrière des sons d'ambiance en agissant sur la différence de gain ΔG entre les enceintes avant et arrière à l'aide d'un bouton rotatif infini et sans gradations. Cette expérience a été effectuée deux fois par chaque sujet lors de deux sessions distinctes, séparées par une pause de 15 minutes, afin de permettre aux sujets de s'habituer à un exercice de mixage sonore en présence d'image 3D.

Les résultats montrent [AI11] que lors de la première session, seule la séquence 10 (Figure 4.7(b)) a bénéficié d'un réglage de gain avant/arrière (ΔG) significativement différent selon que les images étaient présentées en 2D ou en 3D. Par ailleurs, et contrairement à l'hypothèse initiale, la répartition des sons d'ambiance était significativement plus frontale en 3D qu'en 2D (Figure 4.8).

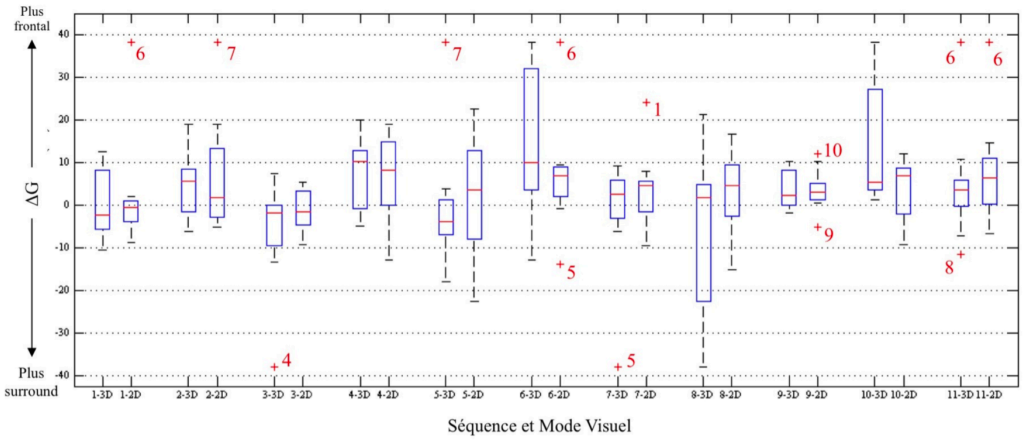


FIGURE 4.8 – Différence de gain avant/arrière ΔG médiane (dans son diagramme en boîte) en fonction de la séquence et du mode visuel associé (3D/2D) pour la session 1 [37].

En revanche, lors de la deuxième session, 3 séquences parmi 11 ont obtenu un réglage ΔG significativement différent selon que les images étaient présentées en 2D ou en 3D. Les sons d'ambiance des séquences 2, 4 (Figure 4.7(a)) et 9 ont ainsi été répartis significativement plus "surround" en 3D qu'en 2D (Figure 4.9).

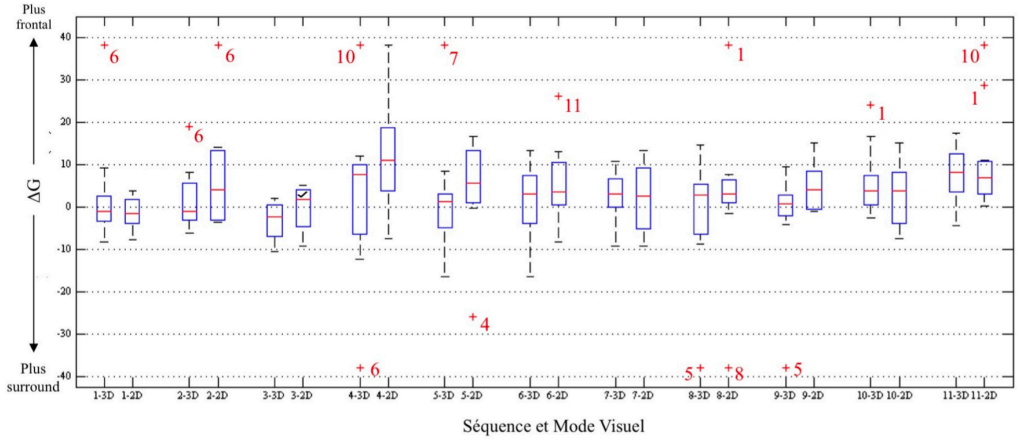


FIGURE 4.9 – Différence de gain avant/arrière ΔG médiane (dans son diagramme en boîte) en fonction de la séquence et du mode visuel associé (3D/2D) pour la session 2 [37].

Ainsi, il semble que le fait de proposer des images en 3D incite les mixeurs à plus profiter des enceintes "surround" même si une période d'apprentissage semble nécessaire.

Une deuxième expérience a ensuite été mise en place afin de déterminer si les attentes des spectateurs en termes d'immersion sont différentes selon qu'ils sont confrontés à des images 2D ou 3D. Ainsi, 8 courtes (environ 30 s) séquences audiovisuelles (image 2D/3D et son multicanal) ont été présentées à des sujets placés en situation de spectateur dans un cinéma. Cette expérience s'est déroulée dans le cinéma "Le Bretagne" à Saint-Renan Figure 4.10(a)). Trois séquences parmi celles proposées lors de l'expérience précédente (sujet en situation de mixeur) ont été réutilisées ici :

— la séquence 4 (Figure 4.7(a)) ayant donné lieu à une balance avant/arrière différente selon

le mode visuel lors de la session 2 (Figure 4.9) ;

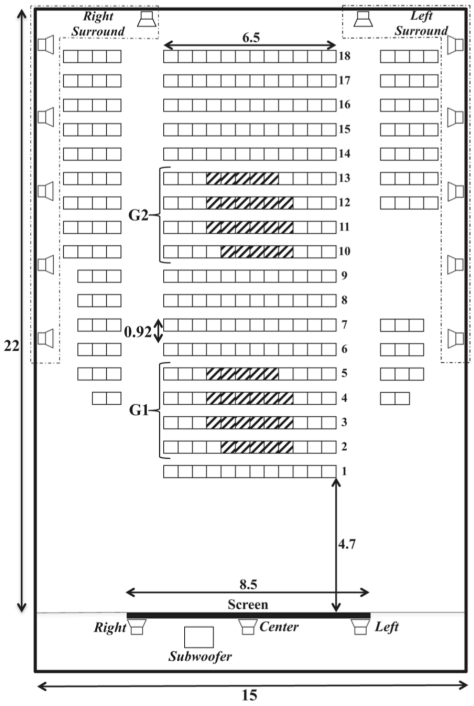
— la séquence 10 (Figure 4.7(b)) ayant donné lieu à une balance avant/arrière différente selon le mode visuel lors de la session 1 (Figure 4.8) ;

— la séquence 11 n'ayant pas donné lieu à une balance avant/arrière différente selon le mode visuel lors de ces deux sessions.

Cinq nouvelles séquences issues de tournages professionnels effectués dans le cadre du projet CCFL2 y ont été ajoutées (exemple en Figure 4.10(b)). Ces 8 séquences ont été projetées dans leurs versions 2D (image classique) et 3D (image stéréoscopique) et avec deux options de mixages sonores réalisés par les expérimentateurs (l'un plutôt frontal, l'autre plutôt "surround"), pour un total de 32 stimuli.

(a)

(b)



(c)

FIGURE 4.10 – Vue intérieure du cinéma "Le Bretagne" (a), exemple de séquence projetée à l'écran (b) et représentation schématique de la salle (c).

Pour chaque stimulus, les sujets devaient indiquer s'ils considéraient que le mixage sonore était adapté aux images projetées. Les jugements étaient à reporter sur une échelle continue allant de "beaucoup trop surround" à "beaucoup trop frontal". La perception de ces sons d'ambiance pouvant dépendre de la position du sujet dans la salle, les sujets, 44 étudiants en Master Image & Son, ont été répartis en 2 groupes de 22 :

— un premier groupe en proximité de l'écran (rangées n° 2, 3, 4 et 5) ;

— un second groupe en position d'écoute de référence (rangées n° 10, 11, 12 et 13) communément appelée "sweet spot" [44] ;

comme indiqué sur la Figure 4.10(c).

Les résultats indiquent [AI11] un effet significatif du mode de présentation mais pour deux séquences seulement (séquences 1 et 2, voir Figure 4.11). Les différents mixages proposés pour

ces deux séquences ont obtenu des degrés d'adaptation aux images 2D et 3D significativement différents, dans le sens ou ceux-ci ont été jugés relativement trop "surround" en 2D. À noter que les séquences 1 et 2 de cette expérience étaient respectivement les séquences 4 et 10 de l'expérience précédente, ayant déjà généré des différences significatives entre les mixages réalisés en présence d'images 2D ou 3D. Ces deux extraits apparaissent comme étant ceux possédant les tailles de "boîte scénique" [45] les plus importantes dans leurs versions 3D. Ces tailles sont déterminées en mesurant (en pixels) la différence entre la disparité stéréoscopique de l'objet le plus loin et de l'objet le plus proche et permettent de décrire la "quantité" de stéréoscopie dans une séquence.

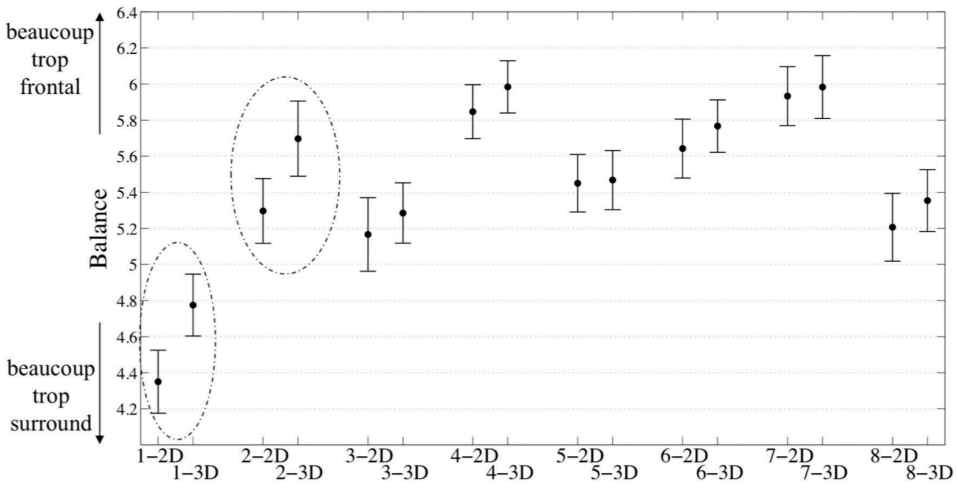


FIGURE 4.11 – Balance moyenne (dans son intervalle de confiance à 95 %) en fonction de la séquence et du mode visuel associé (2D/3D) [37].

Enfin, l'effet du mode de présentation n'apparaît significatif que pour le second groupe de sujets (Figure 4.12), placés au "sweet spot".

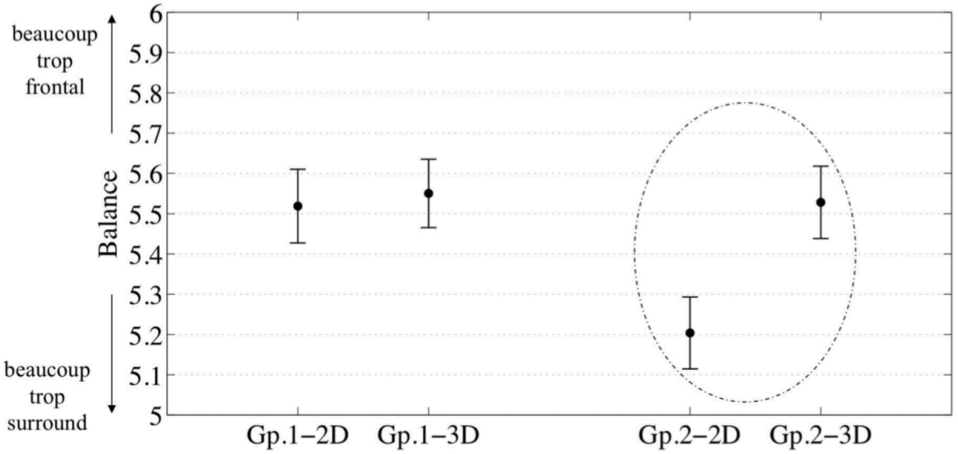


FIGURE 4.12 – Balance moyenne (dans son intervalle de confiance à 95 %) en fonction du groupe et du mode visuel associé (2D/3D) [37].

Ainsi, les attentes en terme d'équilibre avant/arrière des sons d'ambiance sont bien différentes selon que les séquences sont projetées dans leurs versions 2D ou 3D, mais seulement lorsque les séquences exploitent suffisamment les possibilités de la stéréoscopie et lorsque les spectateurs sont dans la position d'écoute de référence.

4.3.2 Influence de l'image stéréoscopique sur la perception des objets sonores

La diffusion sonore traditionnelle en multicanal ne cherche pas la cohérence audiovisuelle pour les objets sonores qui profitent généralement de l'effet de capture visuelle [41]. Ce phénomène est aussi appelé effet ventriloque [46] car il évoque l'illusion d'un ventriloque dont la voix semble provenir de la bouche de sa marionnette plutôt que de la sienne. Cependant de nouvelles techniques de restitution permettent une plus grande cohérence spatiale audiovisuelle en élévation (Dolby Atmos par exemple) et en profondeur (WFS par exemple) [43]. Les possibilités offertes par ces dispositifs sont susceptibles de modifier les attentes des spectateurs en termes de cohérence entre les objets sonores et leurs images respectives, notamment lorsque ces dernières sont projetées en 3D.

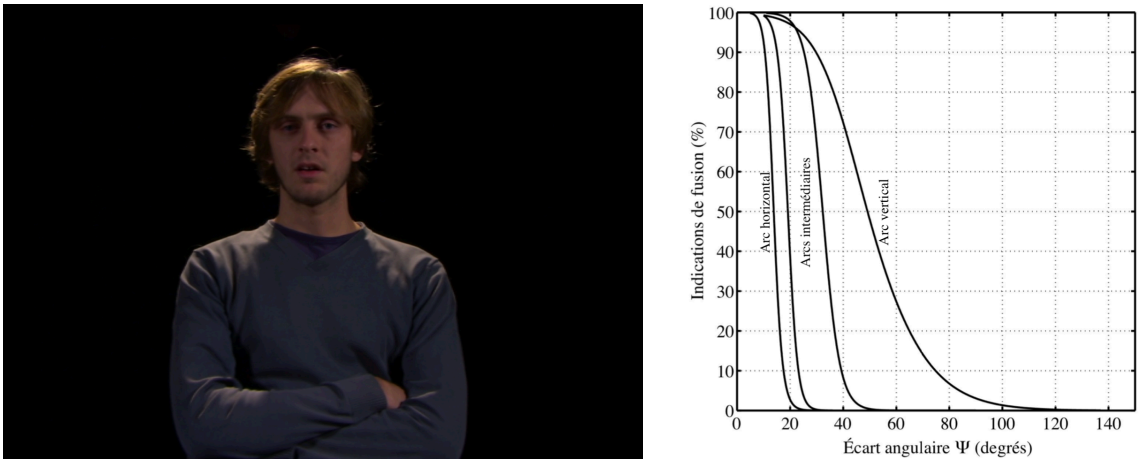
Une expérience préliminaire a été consacrée à l'étude de l'effet ventriloque pour des sources à différents azimuts et élévations. L'image d'un locuteur (Figure 4.13(a)) prononçant diverses phrases était projetée en 3D stéréoscopique sur un écran situé à 2,40 m du sujet. La bouche du locuteur à l'écran se situait exactement dans l'axe avant/arrière (0° d'azimut et 0° d'élévation) du sujet faisant face à l'écran. Le son pouvait être restitué par une des 28 enceintes (Amadeus PMX 4) réparties derrière (ou au-dessus de) l'écran :

— 7 enceintes étaient placées sur un arc horizontal (dans le plan transversal), permettant ainsi de faire varier uniquement l'azimut de la source sonore ;

— 7 enceintes étaient placées sur un arc vertical (dans le plan sagittal), permettant ainsi de faire varier uniquement l'élévation de la source sonore ;

— 14 enceintes étaient placées sur deux arcs intermédiaires (dans des plans inclinés de 45° et 67.5° par rapport au plan transversal), permettant ainsi de faire varier l'azimut et l'élévation de la source sonore ;

afin de faire varier l'écart angulaire Ψ entre la source visuelle (la bouche à l'écran) et la source sonore (l'enceinte).



(a) (b)

FIGURE 4.13 – Image du locuteur projetée à l'écran (a) et pourcentage d'indications de fusion en fonction de l'écart angulaire Ψ (b) pour un sujet typique, de gauche à droite : arc horizontal, arcs intermédiaires et arc vertical [37].

Le sujet devait indiquer si la voix et la bouche semblaient provenir de la même direction. Le test était divisé en 4 sessions d'environ 1 h chacune et 8 sujets naïfs ont participé à cette

expérience. Une indication de fusion signifie que le sujet a reporté que la bouche et la voix provenaient de la même direction. La valeur de Ψ pour laquelle le pourcentage d'indications de fusion est égal à 50% (l'écart angulaire Ψ pour lequel la voix et la bouche semblent provenir de la même direction une fois sur deux) est désigné comme le seuil de fusion. Les résultats indiquent [AI8] que le seuil de fusion est nettement plus élevé pour un écart angulaire en élévation qu'en azimut. La Figure 4.13(b) représente les fonctions psychométriques (pourcentage d'indications de fusion en fonction de l'écart angulaire) d'un sujet illustrant typiquement ces résultats : le seuil de fusion est d'environ 16° sur l'arc horizontal contre environ 55° sur l'arc vertical.

L'effet ventriloque étant très robuste en élévation, l'expérience suivante a eu pour but d'explorer l'effet de la cohérence audiovisuelle dans le plan horizontal (azimut et profondeur) et en étudiant notamment les attentes en terme de cohérence spatiale lorsque le sujet est en présence d'images 2D (classiques) ou 3D (stéréoscopiques). Pour les besoins de cette expérience, 8 courtes (environ 20 s) séquences audiovisuelles (image 2D/3D et son multicanal) ont été réalisées dans le cadre du projet CCFL2 (exemples en Figure 4.14).

(a) (b)

FIGURE 4.14 – Exemples de séquences audiovisuelles utilisées pour la perception des objets sonores [37].

Différentes options de mixage sonore ont été prises afin d'inclure ou non la cohérence audiovisuelle spatiale en azimut :

- mixage "classique" où les objets sonores sont restitués par l'enceinte centrale ;
- mixage "cohérent" où l'azimut des objets sonores correspond à leur position à l'écran, les sources étaient spatialisées par des différences d'amplitude entre enceintes [47] (avec éventuellement du suivi pour les sources en mouvement) ;

et en profondeur :

- mixage "proximité" sans simulation de la profondeur ;
- mixage "distance simulée" en agissant sur le niveau sonore, le timbre et le rapport champ direct à champ réverbéré [36].

Ces mixages ont été réalisés dans une salle d'expérimentation traitée acoustiquement (TR = 0.3 s) et équipée d'enceintes Amadeus PMX 4 constituant un système sonore multicanal [3] auquel 2 enceintes frontales intermédiaires L_C et R_C ($Left_{Center}$ et $Right_{Center}$) ont été ajoutées respectivement entre L et C et entre R et C. Cette disposition à 5 enceintes frontales est notamment recommandée dans les salles de cinéma compatibles Dolby Atmos [43].

Les 8 séquences audiovisuelles ont été proposées à 16 sujets naïfs dans cette même salle d'expérimentation. Les images étaient projetées (vidéo-projecteur 3D Epson EH-TW6000) à l'écran dans leurs versions 2D et 3D stéréoscopique respectivement accompagnées des 4 mixages sonores, pour un total de 64 stimuli. Le sujet (équipé de lunettes 3D actives) était situé à 3.60 m d'un écran acoustiquement transparent derrière lequel étaient disposées les 5 enceintes frontales (Figure 4.15). À chaque stimulus, il devait indiquer si le son était adapté à l'image sur une échelle continue allant de "pas adapté du tout" à "très adapté".

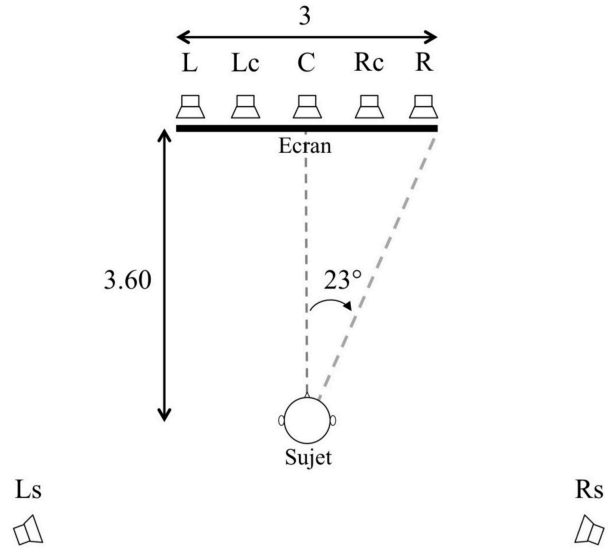


FIGURE 4.15 – Représentation schématique du dispositif expérimental destiné au mixage sonore et à l'évaluation de l'adéquation entre ce mixage et les images projetées [37].

Les résultats montrent [AI7] que le mixage "cohérent" en azimut était considéré comme significativement plus adapté pour 5 séquences sur 8 : les séquences 1 (Figure 4.14(a)), 2, 3, 6 (Figure 4.14(b)) et 7 comme l'indique la Figure 4.16.

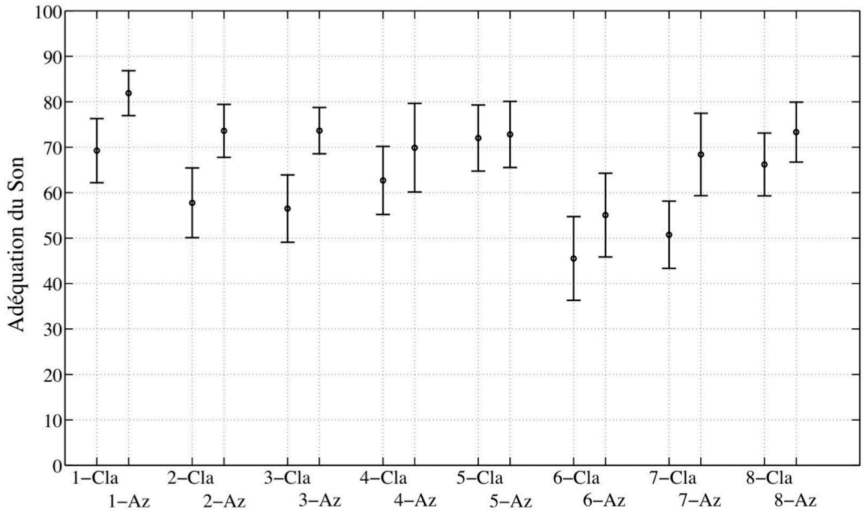


FIGURE 4.16 – Adéquation moyenne (dans son intervalle de confiance à 95 %) du son à l'image en fonction de la séquence et du mixage sonore en azimut : "Cla" pour mixage "classique" avec objets diffusés sur l'enceinte centrale et "Az" pour mixage "cohérent" en azimut [37].

Concernant la profondeur, seule une séquence parmi 8 (séquence 6, voir Figure 4.14(b)) a permis de mettre en évidence le fait qu'un mixage avec "distance simulée" était significativement plus adapté qu'un mixage en "proximité" sans simulation de la profondeur (Figure 4.17).

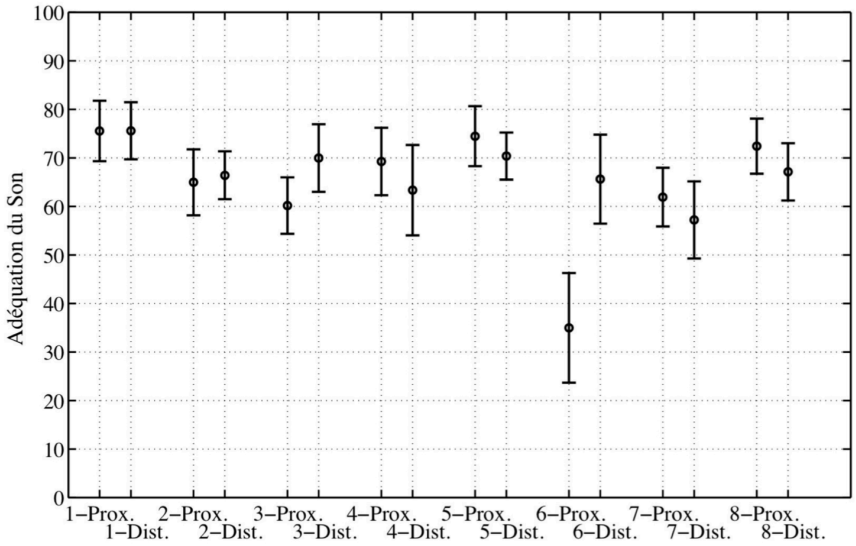


FIGURE 4.17 – Adéquation moyenne (dans son intervalle de confiance à 95 %) du son à l'image en fonction de la séquence et du mixage sonore en profondeur : "Prox" pour mixage "proximité" sans simulation de la profondeur et "Dist" pour mixage "distance simulée" [37].

En revanche, le fait que les images aient été projetées en 2D ou en 3D n'a pas eu d'effet significatif sur l'adéquation du mixage sonore. Ainsi, les auditeurs estiment les mixages sonores proposant de la cohérence audiovisuelle plus adaptés aux images projetées (pour 5 séquences sur 8 en azimut et 1 séquence sur 8 en profondeur), mais indépendemment du mode de projection visuelle. Le fait que les images aient été projetées en 3D n'a donc pas nécessité une cohérence audiovisuelle plus importante qu'en 2D.

4.4 Cohérence audiovisuelle spatiale dans un contexte de concert

De la même manière qu'au cinéma, les habitudes de mixage musical en concert obéissent à des conventions liées au système de diffusion et souvent héritées des techniques stéréophoniques basées sur les différences d'intensité et de temps [4]. En conséquence, la position d'un source sonore dans la rampe stéréophonique coïncide rarement avec sa position réelle sur scène lors d'un concert ou d'un enregistrement. Par exemple, il est d'usage de répartir les éléments d'une batterie sur les extrémités de la rampe stéréophonique bien que le batteur se trouve au centre de la scène. Cependant, dans des contextes de concert également, les systèmes de diffusion stéréophonique traditionnelle tendent à être remplacés par des technologies telles que la WFS qui remettent en cause ces usages en termes de spatialisation sonore. La spatialisation plus robuste et la plus grande cohérence audiovisuelle spatiale permise par la WFS [48] ont ainsi pu être éprouvées lors de concerts sonorisés selon ce principe à la maison de la Radio [49].

Le but de cette étude était d'évaluer l'apport de la cohérence audiovisuelle permise par la WFS à la perception du mixage sonore d'un concert, comparativement à un mixage classique dans lequel la cohérence audiovisuelle n'est pas recherchée. Pour les besoins de cette expérience, des captations (audio et vidéo) ont été réalisées lors de 3 concerts :

— un trio baroque dans une chapelle (Figure 4.18(a)) ;
— un ensemble de jazz dans une salle de concert (Figure 4.18(b)) ;
— un groupe de rock en plein air (Figure 4.18(c)).

Les concerts captés étaient destinés à être restitués lors de projections 3D sonorisées permettant de recréer une situation de concert du point de vue d'un sujet.

(a) (b) (c)

FIGURE 4.18 – Concerts captés : baroque (a), jazz (b) et rock (c).

Pour chacun de ces concerts, 3 ingénieurs du son ont produit chacun 2 mixages sonores différents :

— un mixage "non cohérent" réalisé sans images du concert correspondant : les ingénieurs du son avaient pour consigne de produire le meilleur mixage possible sans contrainte sur le positionnement des sources ;
— un mixage "cohérent" réalisé avec une projection en 3D stéréoscopique des images du concert : les ingénieurs du son avaient pour consigne de produire le meilleur mixage possible tout en veillant à la cohérence audiovisuelle des sources, avec éventuellement du suivi de déplacement pour les sources en mouvement (par exemple le chanteur du groupe de rock visible en Figure 4.18(c)).

Le mixage a été effectué dans une salle d'expérimentation traitée acoustiquement (TR = 0.3 s) et équipée d'un dispositif de restitution sonore composé de 30 enceintes (Amadeus PMX 4), d'un caisson de basse (Genelec 7070A) et d'un écran acoustiquement transparent pour la projection des images (Figure 4.19). Le rendu en WFS était assuré par un processeur (Sonic Emotion Wave 1) disposant d'outils dédiés au mixage "objet". Le mixage "non cohérent" a systématiquement été réalisé en premier pour que les mixeurs n'aient pas d'a priori sur le positionnement des sources.

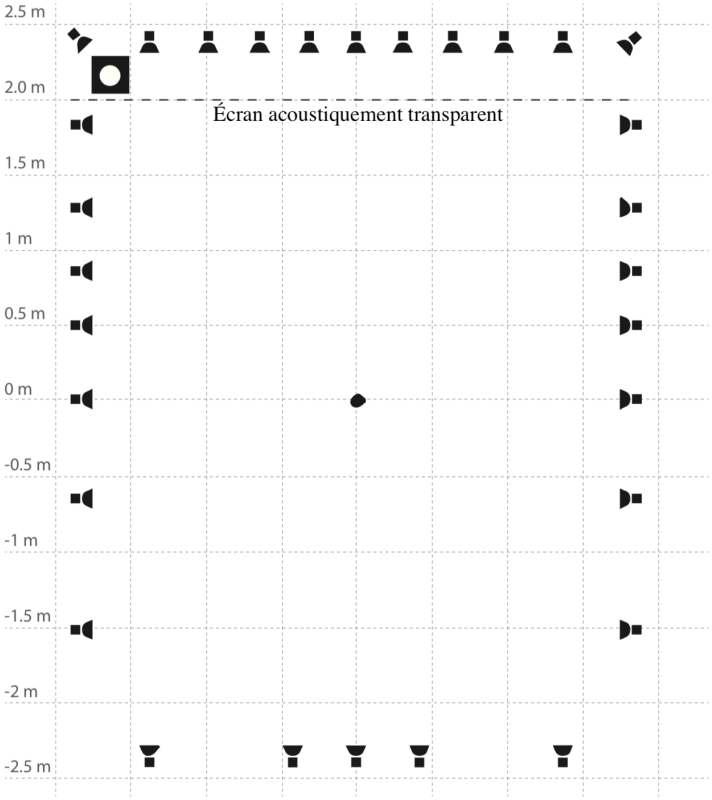


FIGURE 4.19 – Représentation schématique de la salle expérimentale destinée au mixage sonore et à l'évaluation perceptive en WFS.

Ces deux types de mixages ont ensuite été comparés par des sujets placés en situation de concert dans la même salle expérimentale (Figure 4.19) permettant la diffusion audio en WFS et vidéo en 3D stéréoscopique. Une expérience préliminaire a été effectuée afin de vérifier que les mixages "non cohérents" réalisés sans image de concert et "cohérents" donnaient bien lieu à des différences perceptibles. Les résultats ont montré [CI6] que les sujets (13 naïfs et 13 experts) ont systématiquement pu différencier ces deux types de mixages, et qu'ils étaient également bien différenciables de mixages effectués selon des techniques conventionnelles basées sur des différences d'amplitude [50].

Ces deux types de mixages ont ensuite été comparés en terme de préférence par 11 sujets experts. Ces comparaisons ont été effectuées selon deux méthodes de présentation :

— AV (audiovisuel) : diffusion audio en WFS et vidéo en 3D stéréoscopique pour placer les sujets en situation de concert ;

— AS (audio seul) : diffusion audio en WFS sans projection vidéo afin de comparer uniquement les qualités sonores intrinsèques des différents mixages.

Les résultats ont montré [CI7, CI4] que le mixage “cohérent” était significativement préféré au mixage “non cohérent” en présentation audiovisuelle (AV), bien que ce dernier ait été légèrement préféré lorsqu'il était présenté avec en audio seul (AS), comme l'indique la Figure 4.20(a). En d'autres termes, la cohérence audiovisuelle a amené les sujets à préférer des mixages audio dont la qualité sonore avait été évaluée comme intrinsèquement moindre lorsque les comparaisons s'effectuaient sans visuel associé.

Ces observations sont valables quel que soit le concert comme l'indique la Figure 4.20(b) : les mixages cohérents ont été significativement préférés en AV alors qu'une faible préférence pour les mixages non cohérents (voire pas de préférence) a été indiquée en AS.

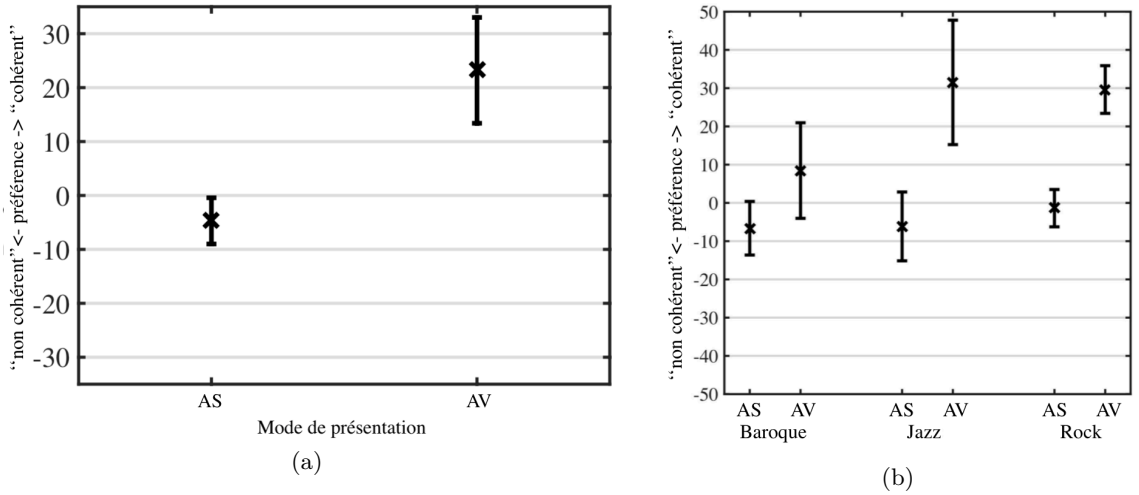


FIGURE 4.20 – Préférence moyenne (dans son intervalle de confiance à 95 %) entre les deux mixages en fonction du mode de présentation associé (a) et selon les différents concerts (b).

Ces constatations ont été appuyées par des comparaisons complémentaires effectuées selon différents attributs (intelligibilité, timbre, immersion, précision spatiale et réalisme) représentatifs de la qualité audio spatiale [51]. Les résultats ont notamment montré [CI4] que les sujets avaient perçu les mixages cohérents comme significativement plus réalistes.

Chapitre 5

Sonie en fonction de la localisation sonore

5.1 Contexte

Ce dernier axe de recherche est né des questionnements sur les variations de niveau sonore perçu (la sonie) en fonction de la localisation d'une source. Lors de tests de qualité sur les systèmes de restitution comme lors de tests de localisation, il s'est avéré que le niveau perçu pouvait varier selon la position de la source sonore. Des études préliminaires ont alors montré que cette dernière pouvait être perçue significativement plus forte lorsque localisée sur le côté de l'auditeur plutôt que face à lui [CJ4, CN7, CN5, AS1, AI9, AI3].

Ces résultats ont été confortés dans le cadre de la thèse de Gauthier Berthomieu [52] qui a confirmé ces effets [CJ3, CJ2, CI5] et montré que leur ampleur pouvait être variable selon la méthode de présentation des stimuli [CJ1, CN1, AI2] (sources réelles ou virtuelles, visibles ou non) et la compréhension de la tâche par le sujet [AI1].

L'effet de la distance perçue sur la sonie a ensuite été étudié en mobilisant les connaissances acquises dans les précédents axes de recherche, notamment sur les interactions audiovisuelles, sur la spatialisation des sources sonores et sur l'influence de la nature de la source (la voix humaine par exemple). Ce travail a permis de mettre à profit l'expérience acquise dans les environnements virtuels pour développer un paradigme expérimental original permettant de contrôler totalement l'environnement audiovisuel dans lequel était immergé le sujet. Ce dernier devait évaluer la distance et la sonie relatives à des sources positionnées dans cet environnement virtuel. Un casque de réalité virtuelle était alors utilisé pour le rendu visuel et sonore (spatialisé et dynamique) de ces sources dans différentes salles. Les résultats ont montré des effets de la distance différents selon que les sujets devaient évaluer la sonie pour du bruit [CI3] ou de la voix parlée [CI1].

5.2 Sonie en fonction de l'azimut

La sonie évoquée par un stimulus sonore peut varier selon l'azimut de sa source. Ce phénomène, appelé sonie directionnelle, peut facilement s'expliquer dans les hautes fréquences où l'ombre acoustique de la tête modifie les pressions atteignant chacune des deux oreilles différemment selon que la source se trouve en face ou sur le côté de l'auditeur [53], affectant ainsi le processus de sommation binaurale [54]. Cependant, ce phénomène a aussi été observé en basse fréquence (400 Hz), où l'influence de la tête est généralement considérée comme négligeable, en diffusant des bruits de largeur de bande égale au tiers d'octave [55].

L'hypothèse de départ des études décrites ci-dessous a donc postulé que les modifications de pression aux oreilles induites par la position d'une source n'étaient pas les seules responsables de ces variations de sonie et que le simple fait de localiser une source à un azimut donné pouvait la modifier. Ainsi les sons proposés ont été spatialisés uniquement d'après l'indice de localisation en basse fréquence : la différence interaurale de temps, communément désignée par son acronyme anglais ITD (Interaural Time Difference). Pour un son pur, la détermination de l'azimut à partir de l'ITD est possible sans ambiguïté tant que la période est supérieure au double de l'ITD maximale possible, lorsque la source possède un azimut de $\pm 90°$. Cette condition peut aussi s'exprimer avec la différence interaurale de phase correspondante (Interaural Phase Difference) :

$$|\mathrm{IPD}| < \pi \tag{5.1}$$

ce qui correspond à une fréquence d'environ 750 Hz, et même jusqu'au double de cette fréquence, soit environ 1500 Hz, si des mouvements de la tête ou de la source sont possibles [53].

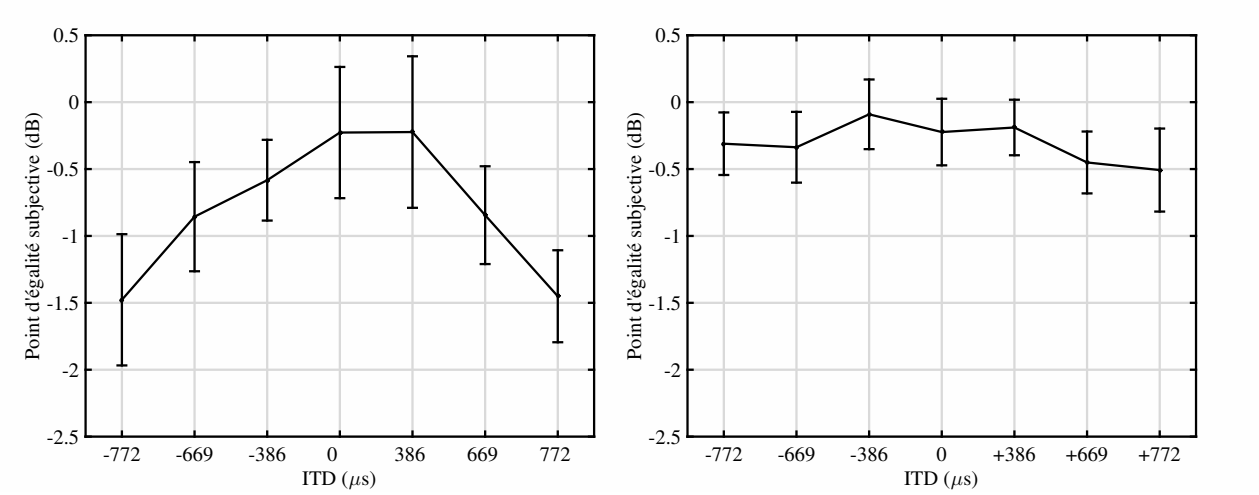
Des sons purs de fréquence 200 et 400 Hz ont été spatialisés par cet indice de localisation en basse fréquence uniquement. Les différences interaurales de temps ont été calculées pour différentes incidences virtuelles θ_{inc} ($0°, \pm 30°, \pm 60°$ et $\pm 90°$) d'après le modèle de Kuhn [56] valable en basses fréquences :

$$\mathrm{ITD} = \frac{3a}{c_0} \sin \theta_{inc} \tag{5.2}$$

où $a = 8,75$ cm désigne le rayon standard de la tête et $c_0 = 340 \text{ m} \cdot \text{s}^{-1}$ la vitesse du son dans l'air. Les différences interaurales obtenues ainsi ont été introduites entre les canaux droit et gauche de sons purs initialement diotiques présentés au casque (modèle Sennheiser HD 650, circum-aural ouvert). Une ITD négative signifie que le stimulus est latéralisé vers la gauche. Les stimuli latéralisés ont été présentés aux auditeurs (11 sujets) dont la tâche consistait en une égalisation de sonie avec une référence diotique.

Les égalisations de sonie ont été réalisées selon une procédure adaptative à 2 intervalles à choix forcé 2I2AFC (2-Interval 2-Alternative Forced Choice), suivant une règle "1-up 1-down" convergeant vers le point d'égalité subjective indiquant une sonie égale. Celui-ci sera désigné ci-après PSE (Point of Subjective Equality).

Un effet significatif de l'ITD sur le PSE a été reporté [AI9] à 40 phones (Figure 5.1(a)) mais pas à 70 phones (Figure 5.1(b)). À 40 phones, le PSE diminue significativement avec l'ITD. Un PSE négatif indique que le stimulus à égaliser a requis un niveau physique moindre que sa référence pour produire la même sonie, on peut donc en déduire que ce stimulus aurait été perçu plus fort que sa référence si présenté au même niveau.



(a) (b)

FIGURE 5.1 – Point d'égalité subjective moyen (dans son intervalle de confiance à 95 %) en fonction de l'ITD pour des sons purs (200 et 400 Hz) à 40 phones (a) et 70 phones (b).

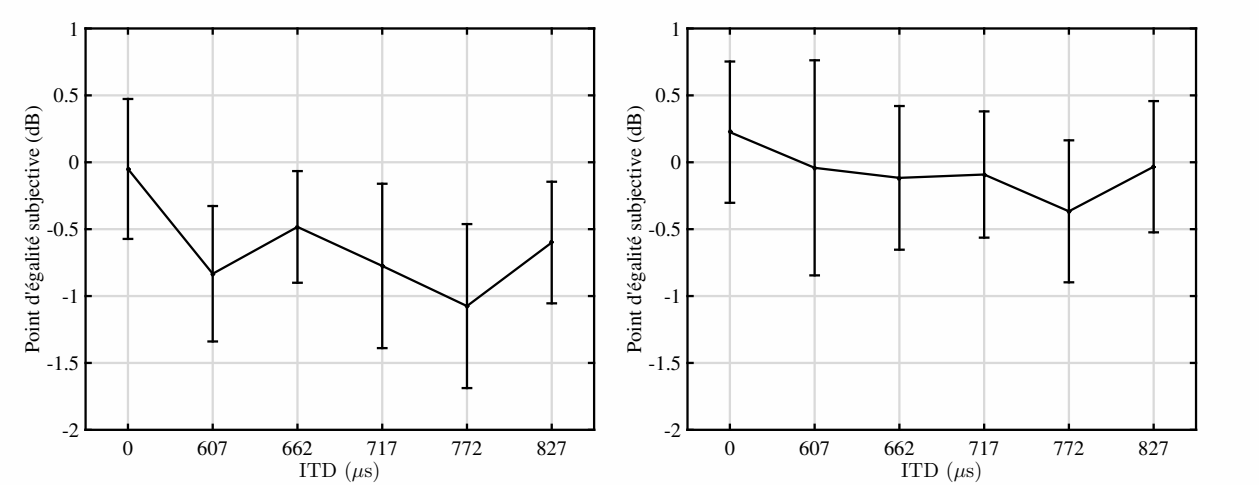
Cet effet s'est également révélé significatif [AI3] à 500 Hz (Figure 5.2(a)) mais pas à 707, 1000, 1404 et 2000 Hz (voir Figure 5.2(b) par exemple) avec 20 sujets pour des valeurs d'ITD relatives à un angle d'incidence θ_{inc} de 90° :

— 772 μs, précédemment obtenue d'après le modèle de Kuhn [56] ;

— 662 μs, obtenue d'après le modèle de Woodworth [57] :

$$\mathrm{ITD} = \frac{a}{c_0}(\theta_{inc} + \sin\theta_{inc}) \tag{5.3}$$

et pour quelques valeurs réparties autour de ces ITD modélisées.

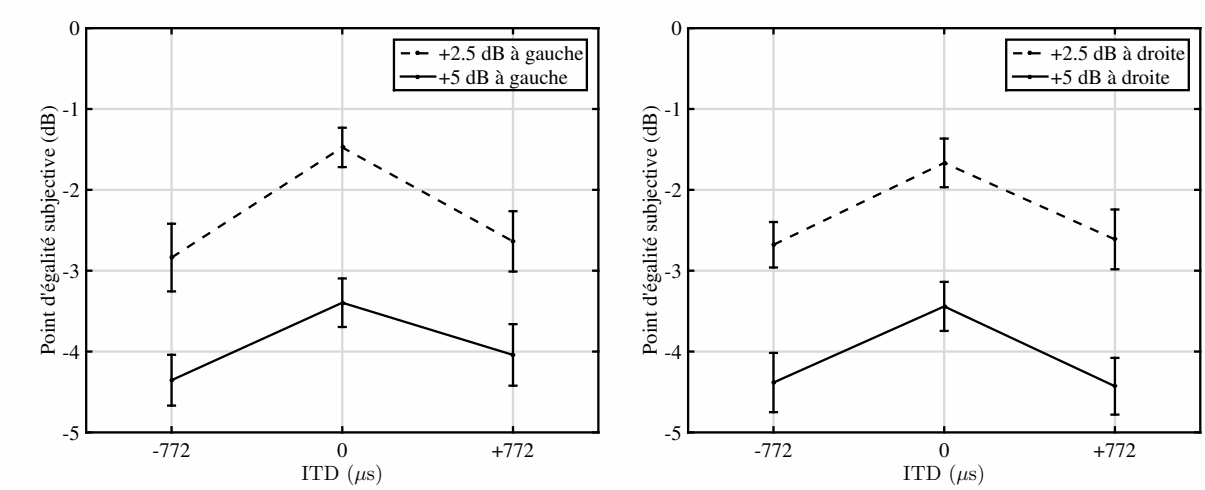


(a) (b)

FIGURE 5.2 – Point d'égalité subjective moyen (dans son intervalle de confiance à 95 %) en fonction de l'ITD obtenu pour des sons purs de fréquence 500 Hz (a) et 2000 Hz (b).

Le fait que l'ITD ait un effet significatif sur la sonie à des fréquences où elle constitue un indice de localisation sans ambiguïté [53] laisse supposer que la localisation de la source sonore et la sonie sont liées. Afin de vérifier cette hypothèse, l'effet de l'ITD a été étudié en

introduisant également une différence interaurale de niveau, communément désignée par son acronyme anglais ILD (Interaural Level Difference). Il s'agit là de l'indice de localisation en azimut complémentaire de l'ITD d'après la théorie "duplex" de Rayleigh [58]. Celui-ci est plutôt utilisé à partir des fréquences où l'ITD devient inexploitable (à partir de 1500 Hz environ), mais la localisation induite par l'ITD peut être compensée par une ILD contradictoire [59]. Les PSE ont alors été mesurés pour 20 sujets et pour différentes valeurs d'ITD en introduisant de surcroît une différence interaurale de niveau. L'ILD a été obtenue en ajoutant 2.5 ou 5 dB sur l'oreille gauche (Figure 5.3(a)) ou droite (Figure 5.3(b)), en présence d'ITD pouvant attirer une localisation à gauche (-772 μs) ou à droite ($+772$ μs). L'ILD et ITD pouvaient être congruentes : $+2.5$ dB à droite et $+772$ μs par exemple. Mais ces différences interaurales pouvaient également être non congruentes et ainsi compenser leurs latéralisations respectives [59] : $+5$ dB à droite et -772 μs par exemple. Les résultats indiquent [AS1] que les effets de l'ITD et de l'ILD s'additionnent, que ces différences soient congruentes ou non (Figure 5.3).



(a) (b)

FIGURE 5.3 – Point d'égalité subjective moyen (dans son intervalle de confiance à 95 %) en fonction de l'ITD obtenu pour des sons purs (200 et 400 Hz) en présence d'ILD menant à gauche (a) et à droite (b).

Ainsi l'effet sur la sonie ne semble pas être dû à la localisation de la source d'après l'ITD mais à l'ITD elle-même. Le fait que ce phénomène se produise à des niveaux faibles (40 phones) mais pas moyens (70 phones) permet d'avancer une autre hypothèse pour expliquer cet effet : des sons présentant une différence interaurale de temps seraient plus facilement séparés du bruit de fond et donc perçus comme plus forts. Un effet similaire sur la sonie a été reporté pour des sons comportant une corrélation interaurale non nulle [60].

Toutefois, l'ITD ne semble pas avoir d'effet au niveau du seuil d'audition puisque la mesure (sur 15 sujets) du seuil de sons purs dont les fréquences étaient comprises entre 125 et 500 Hz n'a pas été significativement affectée par l'ajout d'ITD [AS1]. Le seuil de référence indiqué sur la Figure 5.4 correspond à la moyenne des seuils mesurés sur des sons diotiques (ITD nulle) et a ainsi été défini à 0 dB HL (Hearing Level). Il faut cependant remarquer que la mesure d'un seuil d'audition n'est pas à proprement parler une mesure de sonie et que l'éventuel effet de l'ITD sur la sonie pourrait être inférieur à la précision de la procédure standard utilisée pour ces mesures audiométriques [23].

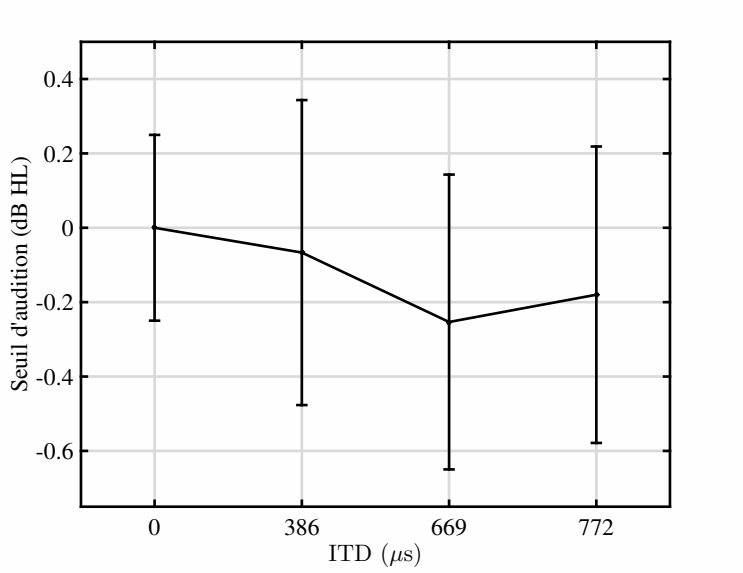


FIGURE 5.4 – Seuil d'audition moyen (dans son intervalle de confiance à 95 %) en fonction de l'ITD obtenu pour des sons purs (125, 200, 250, 400 et 500 Hz).

L'hypothèse d'une meilleure séparation du signal dans le bruit est néanmoins appuyée par les premiers résultats de la thèse de Gauthier Berthomieu [52]. Ceux-ci ont indiqué (pour 22 sujets) que l'effet de l'ITD, étudié sur une plage de niveaux plus étendue (de 30 à 90 phones), diminue significativement [CI5] lorsque le niveau du signal (un son pur de fréquence 200 Hz) augmente (Figure 5.5(a)). Afin de vérifier cette hypothèse, un bruit centré sur 200 Hz et de largeur de bande égale à celle d'un filtre rectangulaire équivalent ERB (Equivalent Rectangular Bandwidth [61]) a été diffusé en plus du son pur. Le niveau du bruit additionnel était de 10 ou 20 phones pour fournir différentes valeurs de rapport signal sur bruit, le niveau du signal étant compris entre 40 et 80 phones. Les résultats indiquent [CI5] que les PSE mesurés (pour 22 sujets) en présence de bruit additionnel (Figure 5.5(b)) ne sont pas significativement différents de ceux mesurés dans le silence (Figure 5.5(a)).

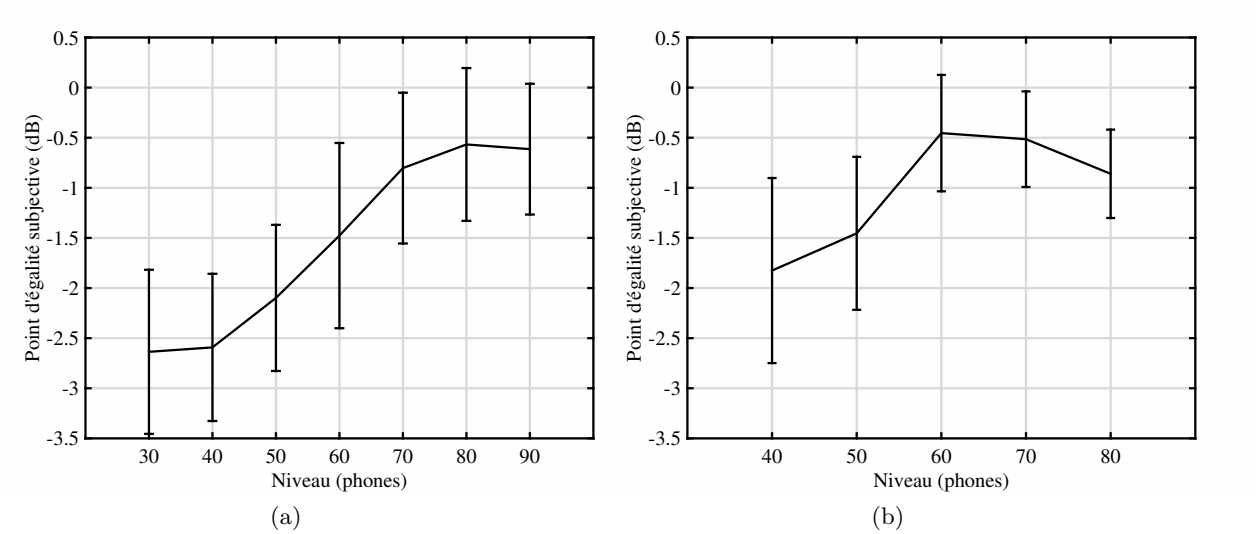


FIGURE 5.5 – Point d'égalité subjective moyen (dans son intervalle de confiance à 95 %) en fonction du niveau du signal obtenu pour des sons purs (200 Hz) présentant une ITD de 772 μs dans le silence (a) et en présence de bruit additionnel (b).

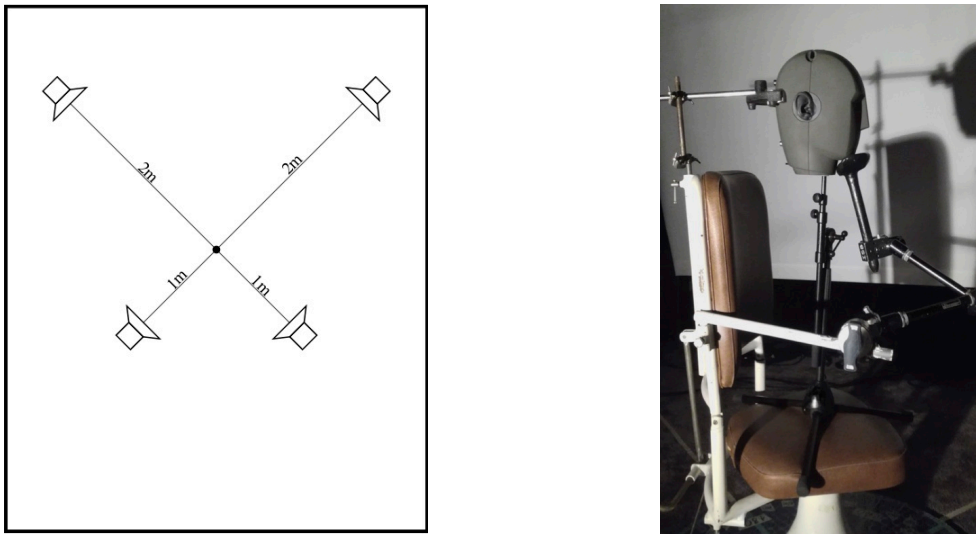
Cette absence d'effet significatif du bruit additionnel ne soutient donc pas l'hypothèse selon laquelle l'ITD permettrait une meilleure séparation du signal dans le bruit. Cet effet pourrait plutôt être expliqué par le fait que l'ITD modifie le processus de sommation binaurale [AI3]. Il a par exemple été montré que l'ITD modifie le taux de potentiels d'action [62] et pourrait ainsi agir sur le phénomène d'inhibition contralatérale, qui décrit le fait qu'un son présenté sur une oreille peut réduire la sonie d'un son présenté sur l'autre oreille [63].

Cet effet a également pu être surévalué par le fait que les stimuli ont été présentés dans des conditions peu écologiquement valides, comme cela a déjà été observé dans des études s'intéressant à la sommation binaurale [64, 65]. Ces études indiquent qu'un effet relevé lors d'une présentation au casque peut se révéler moindre lorsque la présentation est effectuée sur enceinte et en présence d'indices visuels. Ainsi, des stimuli présentés au casque et latéralisés par l'ITD uniquement pourraient avoir produit des PSE exagérément faibles en comparaison de signaux diffusés par des sources réelles et visibles.

Afin de vérifier cette éventualité, ces expériences d'égalisation de sonie ont été reproduites sur enceintes (Amadeus PMX 4), dans une salle d'écoute acoustiquement traitée dont la mesure du temps de réverbération [66] était de 0.3 s à 250 Hz, conformément aux préconisations IEC 6026813 [10] concernant les test d'écoute sur haut-parleurs. Le signal utilisé était un bruit de largeur de bande égale à celle d'un filtre rectangulaire équivalent ERB (Equivalent Rectangular Bandwidth [61]) moins susceptible d'exciter un mode de salle qu'un son pur et de niveau 50 phones. Sa fréquence centrale était de 265 Hz, comprise entre 200 et 400 Hz où l'effet de l'ITD était le plus prononcé [AI9]. Différentes dispositions des enceintes dans la pièce ont été testées de manière à en exciter différemment la réponse. Les sources frontales (de référence) comme latérales (à égaliser) pouvaient ainsi être positionnées à 1 m ou 2 m de la tête de l'auditeur. Ce dernier était assis sur une chaise orientée vers l'une des enceintes. Sa tête était précisément positionnée à l'aide d'un appui-tête et d'une mentonnière, et maintenue dans une position fixe tout au long du test. Le centre de la tête d'un auditeur, entouré par les enceintes, est matérialisé par le point noir sur la Figure 5.6(a). Les points d'égalité subjective ont été déterminés par 20 auditeurs faisant respectivement face à chacune des quatre enceintes et égalisant en sonie les deux enceintes latérales ($\pm 90°$), en présentation réelle.

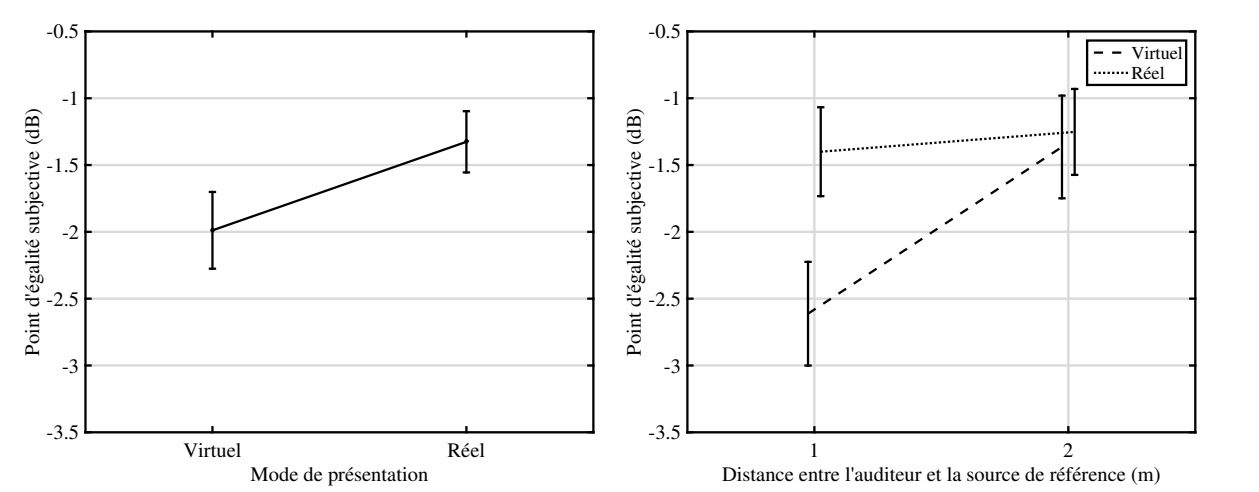
Cette expérience a été reproduite au casque (Sennheiser HD 650), en présentation virtuelle. Les stimuli ont été enregistrés par une tête artificielle (Neumann KU 100) placée et maintenue exactement de la même manière que la tête d'un auditeur, sur la chaise faisant face à chacune des enceintes frontales (Figure 5.6(b)), de manière à reproduire au casque les mêmes stimuli qu'en présentation réelle.

Le bruit a également été simplement latéralisé au casque (772 μs) afin de vérifier que l'effet observé avec un bruit à bande étroite était similaire à celui observé avec des son purs. Le PSE moyen entre ce bruit présenté de manière dichotique et sa référence diotique est d'environ -1.5 dB, du même ordre [AI9] que ce qui avait été mesuré pour des sons purs (Figure 5.1(a)). Les PSE moyens mesurés dans les conditions de présentation réelle et virtuelle ne sont pas significativement différents de cette moyenne, mais sont significativement différents l'un de l'autre (Figure 5.7(a)). Le fait que le PSE soit significativement plus faible dans les conditions de présentation virtuelle confirme que cet effet a pu être surestimé dans des conditions peu écologiquement valides. Cependant, les résultats montrent que cette différence n'est significative que dans les configurations dans lesquelles la source frontale (de référence) est située à 1 m de l'auditeur (Figure 5.7(b)).



(a) (b)

FIGURE 5.6 – Représentation schématique de salle d'écoute utilisée pour les tests sur enceintes (b), tête artificielle placée au point d'écoute dans cette salle (b).



(a) (b)

FIGURE 5.7 – Point d'égalité subjective moyen (dans son intervalle de confiance à 95 %) obtenu pour des bruits à bande étroite (f_c = 265 Hz) diffusés par des sources latérales ($\pm 90°$), en fonction du mode de présentation (a) et en fonction de la distance de la source de référence pour ces deux modes de présentation (b) : virtuel (trait tireté) et réel (trait pointillé).

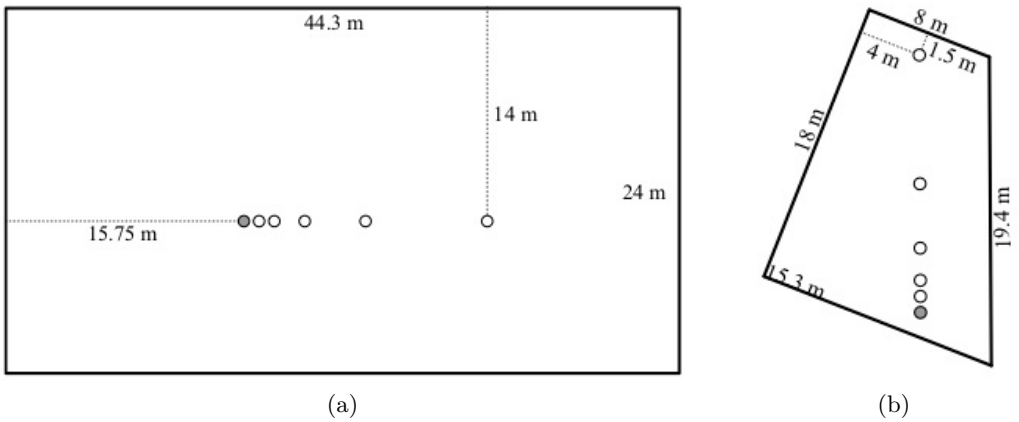
Que les sources aient été situées à 1 m ou 2 m, le niveau sonore restitué au point d'écoute était le même (50 phones). Cependant, les auditeurs ont effectué leurs estimations de sonie tout en étant également capables d'estimer la distance de la source lorsque celle-ci était visible. La distance pouvant être prise en compte dans les estimations de sonie [67], les auditeurs auraient ainsi pu être amenés à estimer différemment la sonie selon les informations disponibles sur la localisation de la source [AI1]. En d'autres termes, le fait que les sources soient réelles et visibles plutôt que virtuelles et invisibles est susceptible de modifier la sonie bien que le niveau sonore soit physiquement le même, comme s'attachera à le montrer la suite de ces études.

5.3 Sonie en fonction de la distance

Cet axe de recherche s'est donc naturellement poursuivi en étudiant la sonie en fonction de la distance, pour laquelle un phénomène de "constance de sonie" a été mis en évidence [67]. Ainsi lorsque la distance entre l'auditeur et la source augmente, la sonie estimée reste constante bien que la pression aux oreilles diminue avec la distance. Ce phénomène n'a néanmoins été observé qu'en présence d'indices permettant à l'auditeur d'estimer la distance le séparant de la source et en orientant la question de telle manière à ce que celle-ci soit prise en compte dans l'estimation. Les expérimentateurs demandaient alors explicitement aux sujets de baser leurs jugements sur la puissance sonore de source. Ce type d'estimation peut également être décrit comme de la sonie *à la source* ("loudness at the source" [54]) par opposition à une estimation de sonie ne se basant que sur les signaux *aux oreilles* ("loudness at the ear" [68]).

La perception de la distance dans un environnement virtuel s'étant révélée conforme à la réalité [AI4], ce paradigme a été utilisé pour étudier les interactions entre distance perçue et sonie (*à la source* et *aux oreilles*) pour des bruits et des signaux de parole. Un environnement virtuel a donc été créé spécifiquement pour permettre aux expérimentateurs de positionner des sources (visibles ou non) à différentes distances de l'auditeur dans plusieurs lieux possibles.

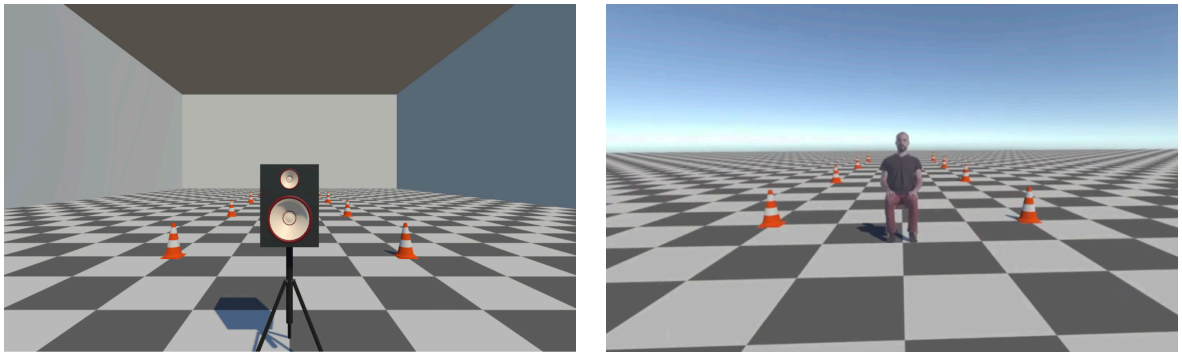
L'expérience s'est déroulée dans une cabine audiométrique où les sujets étaient équipés d'un casque audio (modèle Sennheiser HD 650) pour le rendu sonore binaural et d'un visiocasque (ou HMD pour Head-Mounted Display, modèle HTC Vive) pour le rendu visuel stéréoscopique de l'environnement virtuel. La source pouvait ainsi être virtuellement positionnée à 1, 2, 4, 8 et 16 m de l'auditeur dans 3 "salles" distinctes : une grande salle de sport (TR = 2 s), une petite salle de concert (TR = 0.5 s) et un champ libre (anéchoïque). Deux sources différentes ont été étudiées : un haut-parleur diffusant du bruit blanc et un locuteur prononçant différents mots. Les HRTF (génériques) de la tête artificielle Neumann KU 100 ont été utilisées pour la synthèse binaurale de ces signaux de bruit et de parole, ceux-ci ayant été préalablement convolués par des réponses impulsionnelles spatiales [69] mesurées dans les deux salles (Figure 5.8).



(a) (b)

FIGURE 5.8 – Représentations schématiques de la grande salle de sport (a) et de la petite salle de concert (b) dans lesquelles les réponses impulsionnelles spatiales ont été enregistrées. Dans chaque salle, le cercle plein indique la position du microphone (et donc de l'auditeur), les cercles vides indiquent les différentes positions de la source.

Des modèles 3D ont été créés pour le rendu visuel des salles et des sources sur HMD (voir exemples en Figure 5.9). La restitution sonore comme visuelle se faisait avec suivi des mouvements de la tête. Un panneau occultant pouvait venir se placer entre le sujet et la source pour la masquer visuellement.



(a) (b)

FIGURE 5.9 – Point de vue du sujet sur un haut-parleur situé à 1 m dans la salle de sport (a) et sur un locuteur situé à 4 m en champ libre (b).

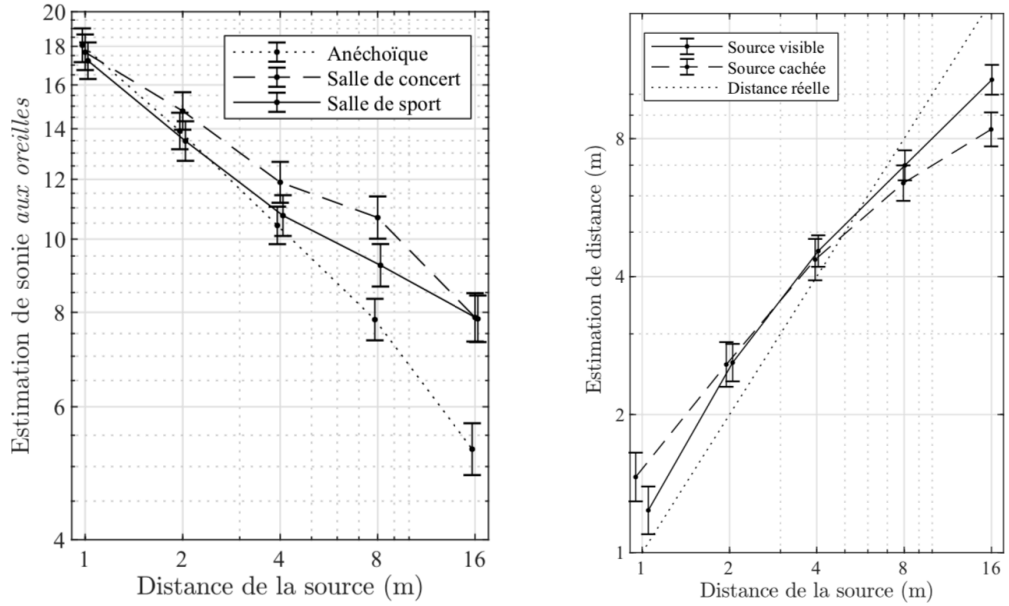
Dans un premier temps, 20 sujets ont eu à produire des estimations de sonie *aux oreilles* (par un nombre reflétant leur sensation sur une échelle libre) et de distance (en m) pour du bruit blanc diffusé par un haut-parleur (Figure 5.9(a)). Les résultats ont montré [CI3] que la sonie *aux oreilles* décroît avec la distance (Figure 5.10(a)) et n'est guidée que par le niveau sonore aux oreilles de l'auditeur (qui peut néanmoins dépendre du champ réverbéré). Les estimations de distance se révèlent pour leur part conformes à la littérature [40] et aux observations déjà effectuées [AI4] (sur-estimation des faibles distance, sous-estimation pour les distances importantes), même lorsque la source n'est pas visible (Figure 5.10(b)). Ces observations suggèrent que l'estimation de la sonie *aux oreilles* ne donne pas lieu à un phénomène de constance selon la distance, bien que cette information ait été rendue disponible aux sujets.

Cette expérience a été répétée avec 17 sujets devant désormais estimer la sonie *à la source* : le sujet devait reporter (sur une échelle libre) un nombre reflétant le fait que le son ait été émis plus ou moins fort. Deux des environnements précédemment décrits ont été conservés ici : la salle de sport (le plus réverbérant des deux environnements non-anéchoïques) et l'environnement anéchoïque, dans lesquels la source pouvait être visible ou non. Seules les estimations de sonie *à la source* effectuées dans l'environnement anéchoïque en présence d'une source non visible (Figure 5.11(a)) présentent une décroissance en fonction de la distance similaire à celle obtenue pour la sonie *aux oreilles* (Figure 5.10(a)).

Cette décroissance de l'estimation de sonie *à la source* en fonction de la distance peut être quantifiée par une régression sur une fonction puissance de forme :

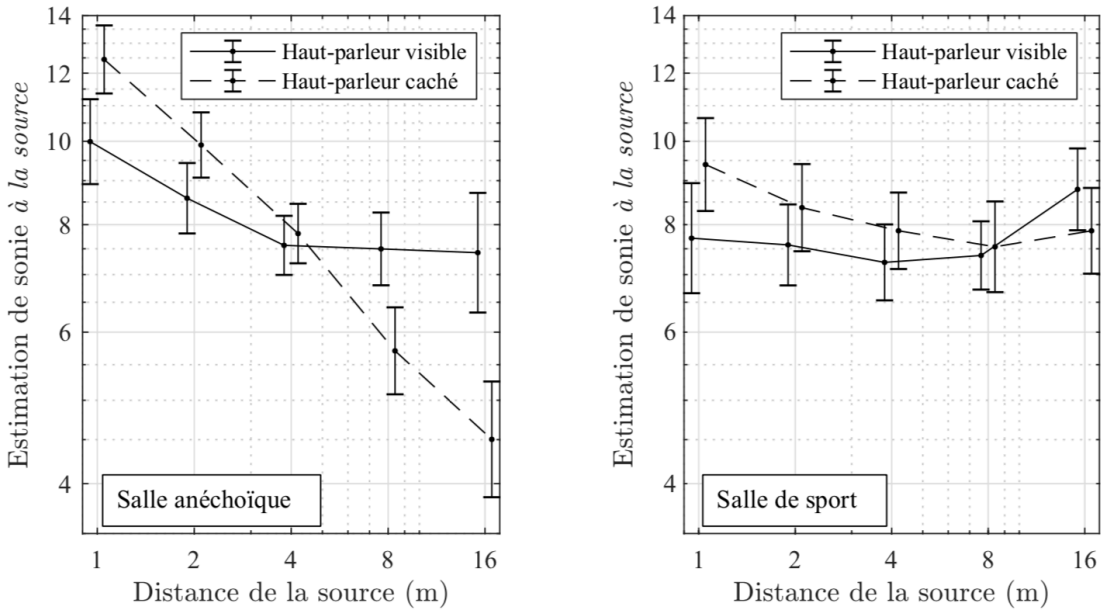
$$L = k \cdot r^b \tag{5.4}$$

où L désigne la quantité de sonie perçue, k une constante, r la distance et b un exposant qui permet d'évaluer la pente de la fonction de sonie [67]. Un exposant $b = 0$ traduirait ainsi une parfaite constance de sonie. La fonction de sonie correspondant aux estimations relatives à la source non visible dans l'environnement anéchoïque (Figure 5.11(a)) présente un exposant $b = -0.37$, indiquant une décroissance importante.



(a) (b)

FIGURE 5.10 – Estimations de sonie *aux oreilles* (a) et de distance (b) moyennes (dans leurs intervalles de confiance à 95 %) en fonction de la distance pour du bruit blanc [52].



(a) (b)

FIGURE 5.11 – Estimation de sonie *à la source* moyenne (dans son intervalle de confiance à 95 %) en fonction de la distance dans l'environnement anéchoïque (a) et dans la salle de sport (b) pour du bruit blanc[52].

En revanche, lorsque la source est visible dans l'environnement anéchoïque (Figure 5.11(a)) ou lorsque celle-ci se trouve dans la salle de sport (visible ou non, Figure 5.11(b)), les exposants b sont nettement plus faibles ($|b| \leq 0.11$) et témoignent de ce qui peut être considéré comme de la constance de sonie [67]. Aussi, dès lors que des informations auditives et/ou visuelles sur la distance de la source sont disponibles, le phénomène de constance de sonie peut être observé sur des estimations de sonie *à la source*.

Finalement, des estimations de sonie *aux oreilles* et de sonie *à la source* ont également été collectées pour des signaux de parole. Ce type de stimulus particulier est susceptible de fournir des informations sur le niveau d'émission de source par le biais de l'effort vocal perçu. Différents mots prononcés par un locuteur ont été proposés à 17 sujets qui ont reporté leurs estimations de sonie sur une échelle libre lors de 2 sessions distinctes (pour estimer respectivement la sonie *aux oreilles* et *à la source*). Le locuteur, visible ou caché, pouvait se trouver dans la salle de sport ou dans l'environnement anéchoïque. La Figure 5.9(b) représente par exemple le locuteur visible dans l'environnement anéchoïque.

Les résultats ont montré [CI1] une nette décroissance ($b = -0.28$) des estimations de sonie *aux oreilles* avec la distance (Figure 5.12(a)), quelle que soit la salle et les conditions de visibilité de la source. En revanche, les estimations de sonie *à la source* indiquent une constance ($b = -0.06$) selon la distance (Figure 5.12(b)), quelle que soit la salle et les conditions de visibilité de la source.

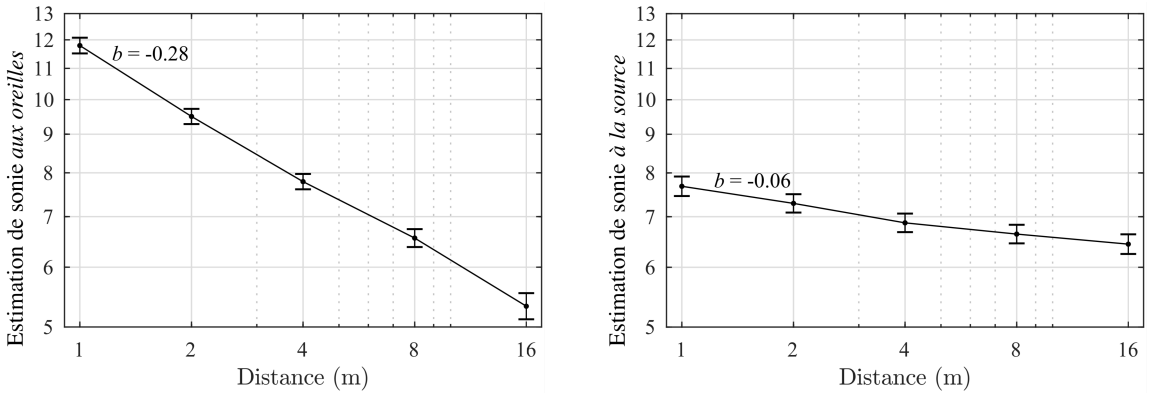


FIGURE 5.12 – Estimations de sonie *aux oreilles* et de sonie *à la source* (b) moyennes (dans leurs intervalles de confiance à 95 %) en fonction de la distance pour des signaux de parole.

Ainsi, la constance de sonie est possible dès lors que que des informations sur la puissance de la source sont disponibles et que le sujet est explicitement amené à estimer de la sonie *à la source*. En revanche, lorsque la question n'est pas spécifiquement posée en ces termes, par exemple lorsqu'il est simplement demandé au sujet de procéder à des égalisations de sonie, celui-ci peut être libre de l'interpréter différemment selon les conditions expérimentales. Les estimations de sonie peuvent alors varier selon que la source est visible ou non [AI2] ou être sujettes à des différences inter-individuelles selon l'azimut de la source [70].

Chapitre 6

Recherches en cours et à venir

Mes projets de recherche actuels et futurs s'inscrivent d'abord dans la continuité des travaux fondamentaux et appliqués précédemment décrits. Du point de vue fondamental, il est ici essentiellement question de localisation sonore, tandis que les applications visent la captation et la restitution sonores spatialisées.

Ensuite, il est prévu que le champ de mes recherches s'élargisse à de nouvelles thématiques. Des études en cours sur les relations entre propriétés physiques et perceptives des instruments de musique vont permettre de développer un axe de recherche déjà abordé par notre équipe de recherche. Enfin, des questionnements sur les niveaux sonores à ne pas dépasser en concert sonorisé ont fait émerger une nouvelle problématique portant sur la protection des dangers liés aux surexpositions sonores.

6.1 Aspects fondamentaux de la localisation sonore

Dans la continuité des travaux sur la sonie directionnelle ou sur la localisation en distance, l'étude des principes fondamentaux de la localisation sonore permettra de poursuivre l'exploration des mécanismes de la perception sonore spatiale. Trois exemples sont donnés ci-dessous.

6.1.1 Diplacousie binaurale dysharmonique

Des travaux antérieurs ont révélé un effet significatif de la position d'une source sur la sonie (sensation de force sonore), mais il apparaît que la hauteur tonale perçue peut elle aussi être affectée. Cette sensation peut varier d'une oreille à l'autre selon les individus. Ce phénomène, appelé diplacousie binaurale dysharmonique, touche principalement des sujets malentendants, présentant des asymétries de sensibilité mais concerne également un auditeur normo-entendant sur cinq [71]. Dans la plupart des cas, ces légères différences de hauteur ne sont pas perçues et les deux sensations sont alors fusionnées en une hauteur tonale unique. Cependant ces différences peuvent être révélées par l'ajout de différences interaurales (de temps et/ou de niveau). Ainsi des sons présentant naturellement de telles différences interaurales, du fait de la position de leur source dans l'espace, pourraient révéler ou exagérer une diplacousie. Des résultats préliminaires ont par ailleurs montré qu'une ITD de 772 μs pouvait révéler une différence de hauteur tonale perçue entre les oreilles pour des sons purs de fréquence 200 Hz présentés à des sujets normo-entendants.

6.1.2 Perception de la distance auditive

La distance auditive a jusqu'ici été étudiée dans des environnements virtuels [AI4, CI3], en présence ou non d'indices visuel, et dans des conditions où la source était à distance constante de l'auditeur. Les résultats ont montré une sous-estimation de la distance similaire à ce qui avait été observé dans des conditions réelles [40, 41]. Cette tendance à la sous-estimation de la distance a également été reportée pour des sources en mouvement, procurant des indices dynamiques de localisation. De plus, les sons s'approchant étaient perçus comme partant et s'arrêtant plus proches que des sons équidistants s'éloignant [72, 73], révélant un mécanisme d'alerte pour une source se rapprochant de l'auditeur. Ces résultats ont toutefois été obtenus pour des mouvements relativement lents de sources réelles [72] ou pour des mouvements relativement rapides mais simulés [73].

Une première étude a été menée avec des enregistrements binauraux d'une source réelle se déplaçant rapidement dans un champ libre ou dans un environnement réverbérant. Cette source – diffusant du bruit blanc, un signal carré ou un bruit de klaxon – pouvait s'approcher ou s'éloigner d'un auditeur qui devait estimer (en m) les distances respectives des points de départ et d'arrivée. De manière surprenante, aucune sous-estimation de la distance n'a été relevée [CN2], que la source s'approche ou s'éloigne de l'auditeur. Les différences entre ces résultats et ceux de la littérature pouvant être dus aux conditions expérimentales (vitesse de déplacement, déplacement réel plutôt que simulé), des études complémentaires seront réalisées avec des déplacements simulés identiques.

6.1.3 Largeur de source apparente

Dans un environnement anéchoïque, seul le son direct issu d'une source parvient à l'auditeur et celle-ci est alors perçue comme ponctuelle. Dans une salle, le rayonnement de la source est réverbéré par les parois et l'auditeur perçoit de multiples réflexions en plus du trajet direct. La localisation de la source, habituellement basée sur ce premier front d'onde, peut être perturbée par des réflexions arrivant très rapidement après le son direct (moins de 80 ms plus tard). La source sera alors perçue comme "large" du fait de ce flou de localisation [74]. Cette "largeur de source apparente" (ou ASW pour Apparent Source Width) dépend essentiellement du rayonnement de la source, des propriétés (géométrie et absorption) de la salle dans laquelle elle se situe et de leur interaction selon la position de la source dans la salle. Des tests préliminaires ont ainsi montré que la largeur perçue dépendait de l'écart angulaire entre le son direct et ses réflexions ainsi que de leurs amplitudes, confirmant que celle-ci était liée à la fraction d'énergie latérale précoce LF_E [74]. Le contenu fréquentiel et le niveau sonore sont également susceptibles de modifier la largeur apparente [75]. Enfin, le processus de localisation résultant souvent d'interactions multisensorielles, la largeur perçue pourrait également être affectée par une représentation visuelle ou une simple connaissance préalable de la source.

Ce dernier aspect fondamental de la localisation auditive trouve également une application directe dans l'étude des systèmes de restitution sonore multi-enceintes (stéréophonique, multicanal...) composés de sources réelles restituant une scène sonore virtuelle. Selon leurs positions dans une pièce donnée, ces sources réelles peuvent être perçues de manière plus ou moins large, et la qualité de l'image sonore restituée peut alors s'en ressentir (sources virtuelles plus floues et difficilement séparable spatialement). De plus, la largeur de la source virtuelle peut être due au principe de spatialisation lui-même. Par exemple, une source virtuelle spatialisée exactement entre deux enceintes sera généralement perçue plus large qu'une source proche d'une des deux

enceintes [76]. Des résultats préliminaires montrent par ailleurs que la largeur perçue est plus importante pour des sources spatialisées par différence de temps que par différence d'intensité. La poursuite de ces travaux permettra d'étudier les interactions entre les dispositifs de spatialisation et les réflexions dues à la salle d'écoute dans la perception de la largeur de source apparente.

6.2 Captation et restitution sonores spatialisées

Les travaux sur la captation et la restitution sonores spatialisées se poursuivront dans notre équipe qui accueille cette année deux nouveaux doctorants (Clément Rappin et Tom Colas) respectivement dans le cadre d'une convention industrielle de formation par la recherche et d'un contrat doctoral d'établissement. Les connaissances acquises lors des recherches résumées dans les chapitres 1 et 2 me permettront de contribuer à ces travaux.

La thèse de Clément Rappin (CIFRE), réalisée en partenariat avec Feichter Electronics (Lannion), permettra de mettre à profit les travaux antérieurs portant sur la conception et l'évaluation des réseaux microphoniques [CI11, CI14, CI15, CI20] pour développer un système de captation ambisonique. Ce travail sera dirigé par Mathieu Paquier et encadré par Julian Palacino (Feichter Electronics). Les signaux ambisoniques captés présentent l'avantage de pouvoir être décodés et restitués selon le système de diffusion sonore à disposition mais seront ici plus particulièrement adaptés à une restitution binaurale sur casque, qui fera l'objet d'une évaluation perceptive. La prise de son ambisonique permet notamment une restitution binaurale dynamique de la scène sonore, en reproduisant les mouvements relatifs de la tête de l'auditeur et des sources. Ce rendu dynamique avec suivi de mouvement favorise l'"externalisation" des sources [77] qui seraient, en restitution binaurale statique, perçues à l'intérieur de la tête pour des signaux synthétisés avec des HRTF génériques [78].

La thèse de Tom Colas (CDE) sera intégralement dédiée à la compréhension du phénomène d'externalisation en écoute binaurale non-individualisée. Ce travail sera dirigé par Mathieu Paquier et co-encadré par Etienne Hendrickx et Nicolas Faruggia (Lab-STICC – IMT Atlantique). Il cherchera à faire le lien entre des mesures subjectives de l'externalisation, classiquement reportée par l'auditeur lui-même [79], et des mesures objectives de l'activité cérébrale par électroencéphalographie (EEG) [80]. Ces nouvelles données, complémentaires des mesures purement comportementales, permettront de mieux comprendre le phénomène d'externalisation et d'établir des modèles permettant sa prévision et son amélioration en agissant sur les stimuli ou les mouvements de tête de l'auditeur. En outre, ce travail donnera lieu à une collaboration inter-équipes dans notre laboratoire, Nicolas Faruggia étant membre de l'équipe 2AI (Algorithm Architecture Interactions) et spécialisé dans les neurosciences et l'intelligence artificielle.

6.3 Lien entre physique et perception des instruments de musique

Le lien entre les matériaux utilisés en facture instrumentale et la perception du son émis a déjà fait l'objet de différentes études dans l'équipe Perception Sonore. Ces expériences avaient pour objectif de déterminer l'influence perceptive du matériau employé pour les anches de cornemuse écossaise [AI6] ou les hautbois de cornemuse du Centre France [81].

Cette thématique de recherche a été récemment relancée par l'accueil en délégation CNRS de Bruno Gazengel, professeur des universités au Laboratoire d'Acoustique de l'Université du Mans (LAUM UMR CNRS 6613), durant l'année universitaire 2020-2021. Le projet de recherche développé avait pour but de relier les caractéristiques physiques d'anches de saxophone à des critères perceptifs du point de vue du musicien en situation de jeu. Les caractéristiques physiques (raideur, débit d'air...) sont déterminées à l'aide d'un banc de mesure simulant le comportement de l'anche en situation de jeu [82]. D'un point de vue perceptif, l'originalité de la démarche repose dans le fait que les attributs verbaux utilisés pour décrire les anches ne sont pas imposés aux musiciens comme cela a pu être proposé dans de précédentes études [83]. À ce stade, un protocole expérimental de caractérisation physique (mesures objectives) et perceptive (mesures subjectives) a pu être déterminé à partir d'essais préliminaires sur un musicien-testeur. La poursuite de ces travaux est envisagée sous la forme d'une thèse en co-tutelle entre le LAUM et le Lab-STICC, co-dirigée par Bruno Gazengel et moi-même.

6.4 Prévention des risques auditifs liés à la musique amplifiée

Cette thématique de recherche a été suscitée par le décret du 7 août 2017 [84] fixant les niveaux maximums de diffusion autorisés "dans les lieux accueillant des activités impliquant la diffusion de sons amplifiés à des niveaux sonores élevés". Les nouvelles limites fixées par ce décret sont désormais de 102 dBA/118 dBC (niveaux intégrés sur 15 minutes) mais soulèvent encore de nombreuses questions dans la communauté scientifique ainsi que parmi les ingénieurs du son. D'une part, les effets à long terme de telles expositions ne sont pas connus, et d'autre part, il est difficile de connaître l'exposition réelle des sujets [CN3].

Le but de cette recherche est donc double : mesurer l'exposition elle-même ainsi que ses effets sur les auditeurs. Ces études seront basées sur un important panel de sujets spectateurs de concerts, auxquels il sera proposé une mesure de leurs capacités auditives juste avant et juste après le concert, mais aussi quelques heures, jours voire semaines suivant l'exposition. Les mesures audiométriques seront réalisées sur la base du volontariat et favorisées par les partenariats avec des salles de spectacle déjà existants dans le cadre du Master "Ingénierie de l'image, Ingénierie du son" :

— Le Quartz, scène nationale de Brest ;

— La Carène, salle des musiques actuelles de Brest.

Concernant la mesure de l'exposition, chaque sujet volontaire sera équipé d'un dosimètre permettant la mesure du niveau sonore tout au long de l'exposition. Une collaboration est déjà en cours avec Feichter Electronics qui a développé un système capable de recueillir simultanément les données issues d'un important panel de dosimètres connectés en Bluetooth et d'ainsi connaître les expositions individuelles. La mise en relation des résultats audiométriques et des données individuelles d'exposition à ces forts niveaux permettra de mettre en évidence leurs conséquences à plus ou moins long terme et leurs potentiels dangers. Les résultats de ces travaux permettront de sensibiliser les ingénieurs du son (étudiants comme professionnels) aux dangers de telles expositions et de réaliser des actions de prévention en partenariat avec la société Audiolite (principal prestataire de sonorisation dans le grand ouest).

Conclusion

Ce mémoire présente une synthèse des travaux de recherche que j'ai réalisés ou auxquels j'ai significativement contribué depuis mon arrivée à l'Université de Bretagne Occidentale en 2006. Ceux-ci portent sur la perception du son restitué et ont été regroupés ici selon 5 axes représentatifs qui abordent des problématiques liées à la psychoacoustique fondamentale comme appliquée :

— l'évaluation perceptive des systèmes de captation sonore spatialisée ;

— l'évaluation perceptive des systèmes de restitution sonore ;

— la modélisation de la qualité vocale en téléphonie mobile ;

— les interactions audiovisuelles ;

— la sonie en fonction de la localisation sonore.

Ces travaux m'ont essentiellement amené à développer ou à contribuer au développement de protocoles expérimentaux pour l'évaluation perceptive. Ils m'ont ainsi permis de consolider une expertise allant du design expérimental jusqu'à l'analyse des résultats.

Les différentes études menées ont par exemple eu pour but d'évaluer la qualité perçue des systèmes de captation/restitution sonore, trouvant leurs applications dans la diffusion musicale, la téléphonie, la réalité virtuelle et le cinéma. D'un point de vue plus fondamental, elles ont permis l'exploration et la compréhension des mécanismes de la localisation sonore et de la perception de la sonie. De manière générale, les études entreprises trouvent souvent leur raison d'être dans les évolutions technologiques de la restitution sonore, accompagnée ou non d'image.

Les connaissances acquises et les résultats obtenus grâce à ces nombreuses expériences me permettent de porter un projet de recherche s'inscrivant à la fois dans la continuité des travaux effectués et dans l'ouverture vers de nouveaux axes de recherche. Ainsi, mes travaux futurs poursuivront l'investigation des aspects fondamentaux de la localisation sonore et appliqués du son spatialisé. Ils se tourneront également vers le lien entre la physique et la perception des instruments de musique et la prévention des risques auditifs liés à la musique amplifiée.

Bibliographie

[1] Simeon Delikaris-Manias. « Parametric spatial audio processing utilising compact microphone arrays ». Thèse de doctorat. Espoo : Aalto University, nov. 2017.

[2] Simeon Delikaris-Manias. *Simulations of second order microphones in audio coding*. Rapport de mobilité doctorale. Université Européenne de Bretagne. Brest : Université de Bretagne Occidentale, juill. 2011.

[3] ITU-R BS.775-3. *Multichannel stereophonic sound system with and without accompanying picture*. International Telecommunication Union. Geneva, Switzerland, août 2012.

[4] David M. Leakey. « Some measurements on the effects of interchannel intensity and time differences in two channel sound systems ». *The Journal of the Acoustical Society of America* 31:7 (1959), p. 977-986. doi : 10.1121/1.1907824.

[5] Michael A. Gerzon. « Periphony: With-height sound reproduction ». *Journal of the audio engineering society* 21:1 (1973), p. 2-10.

[6] Mikkel Nymand. *Microphone techniques for surround sound*. International Tonmeister Symposium. Schloss Hohenkammer, Germany, oct. 2005.

[7] Günther Theile. « Natural 5.1 music recording based on psychoacoustic principals ». *Proceedings of the 19th Audio Engineering Society Conference: Surround Sound – Techniques, Technology, and Perception*. Paper 1904. Schloss Elmau, Germany, juin 2001.

[8] Peter G. Craven et Michael A. Gerzon. *Coincident microphone simulation covering three dimensional space and yielding various directional outputs*. U.S. Patent no. 4,042,779. Août 1977.

[9] Stéphanie Bertet, Jérôme Daniel et Sébastien Moreau. « 3D sound field recording with higher order ambisonics – Objective measurements and validation of spherical microphone ». *Proceedings of the 120th Audio Engineering Society Convention*. Paper 6857. Paris, France, mai 2006.

[10] IEC 60268–13. *Sound system equipment – Part 13: Listening tests on loudspeakers*. International Electrotechnical Commission. Geneva, Switzerland, mars 1998.

[11] ISO 8586. *Sensory analysis – General guidelines for the selection, training and monitoring of selected assessors and expert sensory assessors*. International Organization for Standardization. Geneva, Switzerland, jan. 2014.

[12] Sylvain Choisel et Florian Wickelmaier. « Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference ». *The Journal of the Acoustical Society of America* 121:1 (2007), p. 388-400. doi : 10.1121/1.2385043.

[13] Floyd E. Toole. « Subjective measurements of loudspeakers: A comparison of stereo and mono listening ». *Proceedings of the 74th Audio Engineering Society Convention*. Paper 2023. New York City, NY, USA, oct. 1983.

[14] Allan Devantier, Sean Hess et Sean Olive. « Comparison of loudspeaker-room equalization preferences for multichannel, stereo, and mono reproductions: Are listeners more discriminating in mono? » *Proceedings of the 124th Audio Engineering Society Convention*. Paper 7492. Amsterdam, The Netherlands, mai 2008.

[15] Ken I. McAnally et Russell L. Martin. « Variability in the headphone-to-ear-canal transfer function ». *Journal of the Audio Engineering Society* 50:4 (2002), p. 263-266.

[16] Pierre-Yves Diquelou, David Kerneis et Hmaied Shaiek. *Procédé d'élaboration des filtres de compensation des modes acoustiques d'un local*. Brevet FR2965685. Avr. 2012.

[17] ITU–R BS.1284-2. *General methods for the subjective assessment of sound quality*. International Telecommunication Union. Geneva, Switzerland, jan. 2019.

[18] AES20–1996 (s2008). *AES recommended practice for professional audio – Subjective evaluation of loudspeakers*. Audio Engineering Society. New York City, NY, USA, fév. 2008.

[19] Henrik Møller. « Fundamentals of binaural technology ». *Applied acoustics* 36:3–4 (1992), p. 171-218. doi : 10.1016/0003-682X(92)90046-U.

[20] Abhijit Kulkarni et H. Steven Colburn. « Variability in the characterization of the headphone transfer-function ». *The Journal of the Acoustical Society of America* 107:2 (2000), p. 1071-1074. doi : 10.1121/1.428571.

[21] Klaus A. J. Riederer. « Repeatability analysis of head-related transfer function measurements ». *Proceedings of the 105th Audio Engineering Society Convention*. Paper 4846. San Francisco, CA, USA, sept. 1998.

[22] Henrik Møller, Dorte Hammershøi, Clemen Boje Jensen et Michael Friis Sørensen. « Transfer characteristics of headphones measured on human ears ». *Journal of the Audio Engineering Society* 43:4 (1995), p. 203-217.

[23] ANSI/ASA S3.21. *Methods for manual pure-tone threshold audiometry*. American National Standards Institute. New York City, NY, USA, avr. 2019.

[24] Nicolas Côté. « Integral and diagnostic intrusive prediction of speech quality ». Thèse de doctorat. Technische Universität Berlin, juin 2010.

[25] ITU–T P.863. *Methods for objective and subjective assessment of speech quality: Perceptual Objective Listening Quality Assessment (POLQA)*. International Telecommunication Union. Geneva, Switzerland, mars 2018.

[26] ITU–T P.862. *Perceptual Evaluation of Speech Quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*. International Telecommunication Union. Geneva, Switzerland, fév. 2001.

[27] Marcel Wältermann, Alexander Raake et Sebastian Möller. « Quality dimensions of narrowband and wideband speech transmission ». *Acta Acustica united with Acustica* 96:6 (2010), p. 1090-1103. doi : 10.3813/AAA.918370.

[28] Alexander Raake. *Speech quality of VoIP : assessment and prediction*. Chichester, United Kingdom : John Wiley & Sons, 2006.

[29] Lu Huo, Marcel Wältermann, Ulrich Heute et Sebastian Möller. « Estimation of the speech quality dimension "Discontinuity" ». *Proceedings of the 8th ITG (InformationsTechnische Gesellschaft) Conference on Voice Communication*. Aachen, Germany, oct. 2008.

[30] Barbara J. McDermott. « Multidimensional analyses of circuit quality judgments ». *The Journal of the Acoustical Society of America* 45:3 (1969), p. 774-781. doi : 10.1121/1.1911465.

[31] Brian R. Glasberg et Brian C. J. Moore. « A model of loudness applicable to time-varying sounds ». *Journal of the Audio Engineering Society* 50:5 (2002), p. 331-342.

[32] ITU–T P.800. *Methods for objective and subjective assessment of quality*. International Telecommunication Union. Geneva, Switzerland, août 1996.

[33] ITU–T COM 12-34. *TOSQA – Telecommunication Objective Speech Quality Assessment*. International Telecommunication Union. Geneva, Switzerland, déc. 1997.

[34] Nitay Shiran et Ilan D. Shallom. « Enhanced PESQ algorithm for objective assessment of speech quality at a continuous varying delay ». *Proceedings of the 1st International Workshop on Quality of Multimedia Experience*. San Diego, CA, USA, 2009, p. 157-162. doi : 10.1109/QOMEX.2009.5246960.

[35] Marcel Wältermann, Alexander Raake et Sebastian Möller. « Direct quantification of latent speech quality dimensions ». *Journal of the Audio Engineering Society* 60:4 (2012), p. 246-254.

[36] Nicolas Côté. *Perception multimodale de la distance dans un environnement virtuel*. Rapport post-doctoral. Conseil Général du Finistère. Brest : Université de Bretagne Occidentale, juill. 2011.

[37] Etienne Hendrickx. « Influence de la stéréoscopie sur la perception du son : cas de mixages sonores pour le cinéma en relief ». Thèse de doctorat. Brest : Université de Bretagne occidentale, déc. 2015.

[38] Douglas R. Campbell, Kalle J. Palomaki et Guy J. Brown. « A Matlab simulation of "shoebox" room acoustics for use in research and teaching ». *Computing and Information Systems* 9:3 (2005), p. 48-51.

[39] Marco Jeub, Magnus Schafer et Peter Vary. « A binaural room impulse response database for the evaluation of dereverberation algorithms ». *16th International Conference on Digital Signal Processing*. 2009. doi : 10.1109/ICDSP.2009.5201259.

[40] Pavel Zahorik, Douglas S. Brungart et Adelbert W. Bronkhorst. « Auditory distance perception in humans: A summary of past and present research ». *Acta Acustica united with Acustica* 91:3 (2005), p. 409-420.

[41] Jörg Lewald, Walter H. Ehrenstein et Rainer Guski. « Spatio-temporal constraints for auditory–visual integration ». *Behavioural brain research* 121:1-2 (2001), p. 69-79. doi : 10.1016/S0166-4328(00)00386-7.

[42] Michel Chion. *L'Audio-vision : son et image au cinéma*. 5e éd. Paris, France : Armand Colin, 2021.

[43] Francis Rumsey. « Immersive audio: Objects, mixing, and rendering ». *Journal of the Audio Engineering Society* 64:7/8 (2016), p. 584-588.

[44] EBU R90. *The subjective evaluation of the quality of sound programme material*. European Beoadcasting Union. Geneva, Switzerland, déc. 2000.

[45] Samuel Moulin. « Quel son spatialisé pour la vidéo 3D ? Influence d'un rendu Wave Field Synthesis sur l'expérience audio-visuelle 3D ». Thèse de doctorat. Paris : Sorbonne Université, avr. 2015.

[46] Charles E. Jack et Willard R. Thurlow. « Effects of degree of visual association and angle of displacement on the "ventriloquism" effect ». *Perceptual and motor skills* 37:3 (1973), p. 967-979. doi : 10.1177/003151257303700360.

[47] Ville Pulkki et Matti Karjalainen. « Localization of amplitude-panned virtual sources I: stereophonic panning ». *Journal of the Audio Engineering Society* 49:9 (2001), p. 739-752.

[48] Augustinus J. Berkhout, Diemer de Vries et Peter Vogel. « Acoustic control by wave field synthesis ». *The Journal of the Acoustical Society of America* 93:5 (1993), p. 2764-2778. doi : 10.1121/1.405852.

[49] Etienne Corteel, Raphaël Foulon et Frédéric Changenet. « A hybrid approach to live spatial sound mixing ». *Proceedings of the 140th Audio Engineering Society Convention*. Paper 9527. Paris, France, mai 2016.

[50] Ville Pulkki. « Virtual sound source positioning using vector base amplitude panning ». *Journal of the Audio Engineering Society* 45:6 (1997), p. 456-466.

[51] Alexander Lindau, Vera Erbes, Steffen Lepa, Hans-Joachim Maempel, Fabian Brinkman et Stefan Weinzierl. « A spatial audio quality inventory (SAQI) ». *Acta Acustica united with Acustica* 100:5 (2014), p. 984-994.

[52] Gauthier Berthomieu. « Influence de la position d'une source sur le niveau sonore perçu ». Thèse de doctorat. Brest : Université de Bretagne occidentale, déc. 2019.

[53] Brian C. J. Moore. « Space perception ». *An introduction to the psychology of hearing*. 6^{e} éd. Leiden, The Netherlands : Brill, 2013. Chap. 7, p. 245-281.

[54] Ville Pekka Sivonen et Wolfgang Ellermeier. « Binaural loudness ». *Loudness*. Sous la dir. de Mary Florentine, Arthur N. Popper et Richard R. Fay. New York City, NY, USA : Springer, 2011. Chap. 7, p. 169-197.

[55] Ville Pekka Sivonen et Wolfgang Ellermeier. « Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation ». *The Journal of the Acoustical Society of America* 119:5 (2006), p. 2965-2980. doi : 10.1121/1.2184268.

[56] George F. Kuhn. « Model for the interaural time differences in the azimuthal plane ». *the Journal of the Acoustical Society of America* 62:1 (1977), p. 157-167. doi : 10.1121/1.381498.

[57] Robert S. Woodworth. « Hearing ». *Experimental psychology*. Sous la dir. d'Harold Schlosberg. New York City, NY, USA : Holt, 1938. Chap. 20, p. 501-539.

[58] John W. Strutt (3rd Baron Rayleigh). « On our perception of sound direction ». *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 13:74 (1907), p. 214-232. doi : 10.1080/14786440709463595.

[59] Gerard G. Harris. « Binaural interactions of impulsive stimuli and pure tones ». *The Journal of the Acoustical Society of America* 32:6 (1960), p. 685-692. doi : 10.1121/1.1908181.

[60] Barrie A. Edmonds et John F. Culling. « Interaural correlation and the binaural summation of loudness ». *The Journal of the Acoustical Society of America* 125:6 (2009), p. 3865-3870. doi : 10.1121/1.3120412.

[61] Brian R. Glasberg et Brian C. J. Moore. « Derivation of auditory filter shapes from notched-noise data ». *Hearing research* 47:1-2 (1990), p. 103-138. doi : 10.1016/0378-5955(90)90170-T.

[62] Victor Benichoux, Marc Rébillat et Romain Brette. « On the variation of interaural time differences with frequency ». *The Journal of the Acoustical Society of America* 139:4 (2016), p. 1810-1821. doi : 10.1121/1.4944638.

[63] Brian C. J. Moore et Brian R. Glasberg. « Modeling binaural loudness ». *The Journal of the Acoustical Society of America* 121:3 (2007), p. 1604-1612. doi : 10.1121/1.2431331.

[64] Michael Epstein et Mary Florentine. « Binaural loudness summation for speech and tones presented via earphones and loudspeakers ». *Ear and hearing* 30:2 (2009), p. 234-237. doi : 10.1097/AUD.0b013e3181976993.

[65] Michael Epstein et Mary Florentine. « Binaural loudness summation for speech presented via earphones and loudspeaker with and without visual cues ». *The Journal of the Acoustical Society of America* 131:5 (2012), p. 3981-3988. doi : 110.1121/1.3701984.

[66] ISO 3382–1. *Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces*. International Organization for Standardization. Geneva, Switzerland, juin 2009.

[67] Pavel Zahorik et Frederic L. Wightman. « Loudness constancy with varying sound source distance ». *Nature neuroscience* 4:1 (2001), p. 78-83. doi : 10.1038/82931.

[68] Donald H. Mershon, Douglas H. Desaulniers, Stephan A. Kiefer, Thomas L. Amerson Jr et Jeanne T. Mills. « Perceived loudness and visually-determined auditory distance ». *Perception* 10:5 (1981), p. 531-543. doi : 10.1068/p100531.

[69] François Salmon. « Contrôle des impressions spatiales dans un environnement acoustique virtuel ». Thèse de doctorat. Rennes : IRT b<>com / Université de Bretagne occidentale, mars 2021.

[70] Sabine Meunier, Sophie Savel, Jacques Chatron et Guy Rabau. « Interindividual differences in directional loudness ». *The Journal of the Acoustical Society of America* 140:4 (2016), p. 3268. doi : 10.1121/1.4970368.

[71] Edward M. Burns. « Pure-tone pitch anomalies. I. Pitch-intensity effects and diplacusis in normal ears ». *The Journal of the Acoustical Society of America* 72:5 (1982), p. 1394-1402. doi : 10.1121/1.388445.

[72] John G. Neuhoff. « An adaptive bias in the perception of looming auditory motion ». *Ecological Psychology* 13:2 (2001), p. 87-110. doi : 10.1207/S15326969EC01302_2.

[73] John G. Neuhoff, Rianna Planisek et Erich Seifritz. « Adaptive sex differences in auditory motion perception: looming sounds are special ». *Journal of Experimental Psychology: Human Perception and Performance* 35:1 (2009), p. 225-234. doi : 10.1037/a0013159.

[74] Toshiyuki Okano, Leo L. Beranek et Takayuki Hidaka. « Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E), and apparent source width (ASW) in concert halls ». *The Journal of the Acoustical Society of America* 104:1 (1998), p. 255-265. doi : 10.1121/1.423955.

[75] Ingo B. Witew et Johannes A. Buechler. « The perception of apparent source width and its dependence on frequency and loudness ». *The Journal of the Acoustical Society of America* 120:5 (2006), p. 3224. doi : 10.1121/1.4788196.

[76] Ville Pulkki, Matti Karjalainen et Jyri Huopaniemi. « Analyzing virtual sound source attributes using a binaural auditory model ». *Journal of the Audio Engineering Society* 47:4 (1999), p. 203-217.

[77] Etienne Hendrickx, Peter Stitt, Jean-Christophe Messonnier, Jean-Marc Lyzwa, Brian F. G. Katz et Catherine De Boishéraud. « Improvement of externalization by listener and source movement using a binauralized microphone array ». *Journal of the Audio Engineering Society* 65:7/8 (2017), p. 589-599. doi : 10.17743/jaes.2017.0018.

[78] Durand R. Begault et Elizabeth M. Wenzel. « Headphone localization of speech ». *Human Factors* 35:2 (1993), p. 361-376. doi : 10.1177/001872089303500210.

[79] Etienne Hendrickx, Peter Stitt, Jean-Christophe Messonnier, Jean-Marc Lyzwa, Brian F. G. Katz et Catherine De Boishéraud. « Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis ». *The Journal of the Acoustical Society of America* 141:3 (2017), p. 2011-2023. doi : 10.1121/1.4978612.

[80] Nicolas Farrugia, Alix Lamouroux, Christophe Rocher, Jules Bouvet et Giulia Lioi. « Beta and theta oscillations correlate with subjective time during musical improvisation in ecological and controlled settings: a single subject study ». *Frontiers in Neuroscience* 15 (2021). doi : 10.3389/fnins.2021.626723.

[81] Mathieu Paquier, Etienne Hendrickx et Raphaël Jeannin. « Effect of wood on the sound of oboe as simulated by the chanter of a 16-inch French bagpipe ». *Applied Acoustics* 103 (2016), p. 47-53. doi : 10.1016/j.apacoust.2015.10.008.

[82] Alberto Muñoz Arancón, Bruno Gazengel et Jean-Pierre Dalmont. « Comparison of human and artificial playing of a single reed instrument ». *Acta Acustica united with Acustica* 104:6 (2018), p. 1104-1117. doi : 10.3813/AAA.919275.

[83] Jean-François Petiot, Pierric Kersaudy, Gary Scavone, Stephen McAdams et Bruno Gazengel. « Investigation of the relationships between perceived qualities and sound parameters of saxophone reeds ». *Acta Acustica united with Acustica* 103:5 (2017), p. 812-829. doi : 10.3813/AAA.919110.

[84] « Décret n° 2017-1244 du 7 août 2017 relatif à la prévention des risques liés aux bruits et aux sons amplifiés ». *Journal Officiel de la République Française* 0185 (août 2017).

DOCTORAT BRETAGNE LOIRE / MATHSTIC

UBO
Université de Bretagne Occidentale

Titre : Aspects perceptifs de la restitution sonore

Mots-clés : Perception sonore, Psychoacoustique, Localisation sonore, Son spatialisé, Qualité sonore, Sonie

Résumé : Ce mémoire résume les activités de recherche que j'ai menées depuis 2006 à l'Université de Bretagne Occidentale (Brest). Celles-ci se sont d'abord déroulées dans le cadre du Laboratoire d'Informatique des Systèmes Complexes (LISyC EA 3883, de 2006 à 2012) et actuellement dans le cadre du Laboratoire des Sciences et Techniques de l'Information, de la Communication et de la Connaissance (Lab-STICC UMR CNRS 6285, de 2012 jusqu'à présent). Ces recherches relèvent de la psychoacoustique et portent principalement sur la perception du son dans des contextes où celui-ci est restitué : la réalité virtuelle, le cinéma, la diffusion musicale (avec ou sans image associée) et les télécommunications.

Les travaux présentés dans ce mémoire portent sur des aspects fondamentaux comme appliqués de la perception sonore, avec pour exemples respectifs la compréhension des mécanismes de localisation auditive et l'évaluation de la qualité sonore. Ils sont ordonnés selon cinq axes présentant parfois des problématiques et méthodologies communes. Les principales expériences relatives à chacun de ces axes sont décrites ici de manière synthétiques. Enfin, les perspectives ouvertes par leurs principaux résultats permettent de définir un projet de recherche s'inscrivant à la fois dans la poursuite des études passées et dans le développement de nouveaux axes de recherche.

Title: Perceptual aspects of sound reproduction

Keywords: Sound perception, Psychoacoustics, Sound localization, Spatial sound, Sound Quality, Loudness

Abstract: This dissertation summarizes the research activities that I have carried out since 2006 at the University of Brest (Université de Bretagne Occidentale). They first took place within the Laboratory for Computer Science of Complex Systems (LISyC EA 3883, from 2006 to 2012) and then within the Laboratory for Sciences and Techniques of Information, Communication and Knowledge (Lab-STICC UMR CNRS 6285, from 2012 to present). This research is related to psychoacoustics and is mainly about the perception of sound in contexts where it is reproduced: virtual reality, cinema, music (with or without accompanying picture) and telecommunications.

The research studies presented in this dissertation address both fundamental and applied aspects of sound perception, taking auditory localization mechanisms and sound quality assessment as respective examples. The presentation of this research work is organized along five thematic axes that may share common scientific issues and experimental methodologies. The principal experiments related to each of these axes are here briefly described. Finally, the prospects opened by their main findings enable to define a research project that encompasses both the continuation of past studies and the development of new research axes.