HAL
open science

Influence de la position d'une source sur le niveau sonore perçu

Gauthier Berthomieu

To cite this version:

Gauthier Berthomieu. Influence de la position d'une source sur le niveau sonore perçu. Acoustique [physics.class-ph]. Université de Bretagne occidentale - Brest, 2019. Français. NNT : 2019BRES0115 . tel-02469434v2

HAL Id: tel-02469434

https://theses.hal.science/tel-02469434v2

Submitted on 4 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE

L'UNIVERSITE DE BRETAGNE OCCIDENTALE
COMUE UNIVERSITE BRETAGNE LOIRE

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

Par
Gauthier BERTHOMIEU

Influence de la position d'une source sur le niveau sonore perçu.

Thèse présentée et soutenue à Brest, le 12 décembre 2019
Unités de recherche : Lab-STICC, CNRS, UMR 6285

Rapporteurs avant soutenance :

Wolfgang ELLERMEIER	Professeur des Universités, TU, Darmstadt
Sabine MEUNIER	Chargée de Recherche, CNRS, Marseille

Composition du Jury :

Président du Jury :	Etienne PARIZET	Professeur des Universités, INSA, Lyon
Examinateurs :	Wolfgang ELLERMEIER	Professeur des Universités, TU, Darmstadt
	Sabine MEUNIER	Chargée de Recherche, CNRS, Marseille
	Patrick SUSINI	Directeur de Recherche, IRCAM, Paris
Dir. de thèse :	Mathieu PAQUIER	Professeur des Universités, Université de Bretagne Occidentale, Brest
Co-encadrant :	Vincent KOEHL	Maître de Conférences, Université de Bretagne Occidentale, Brest

Invité :

Etienne HENDRICKX	Maître de Conférences, Université de Bretagne Occidentale, Brest

Remerciements

Je souhaite, dans un premier temps, remercier chaleureusement Mathieu Paquier et Vincent Koehl. Je les remercie évidemment pour m'avoir respectivement dirigé et encadré au cours de cette thèse, mais également pour les trois années durant lesquelles j'étais leur étudiant en Licence et Master Image & Son. Ces six années passées à leurs côtés représentent pour moi un sacré bout de chemin, au cours duquel j'ai appris énormément. J'en garderai sans nul doute un précieux souvenir, alimenté par les innombrables jeux de mots vaseux et moqueries à base d'adaptateurs Apple qui ponctuaient nos journées.

Je remercie bien sûr Sabine Meunier et Wolfgang Ellermeier d'avoir accepté avec enthousiasme d'être les rapporteurs de ce manuscrit. Je tiens également à remercier Étienne Parizet et Patrick Susini, chacun à deux reprises, pour leur participation au jury ainsi que pour avoir été les membres de mon comité de suivi individuel durant ces trois années.

J'adresse à Étienne Hendrickx et Julian Palacino un certain « Merci à vous, merci beaucoup ! » qui devrait leur faire fredonner un extrait de funk à haut-de-forme pour le restant de la journée. Ce fut un réel plaisir de partager ce bureau avec eux deux. Comme promis, nous célébrerons la fin de cet épisode par une fondue savoyarde tant attendue.

Je remercie sincèrement toute l'équipe de l'ISB, Luc Pennamen et François Barvec, Vincent Mazo, Karine Peron, pour leur soutien permanent et nécessaire. Je tiens à remercier tous les membres de l'Université de Bretagne Occidentale avec lesquels j'ai eu la chance de travailler, Emanuel Radoi, David Spenato, Anne Le Roux, Michèle Kerleroux ainsi que Gaël Le Roux pour son aide précieuse lors de la confection de la chaise de torture utilisée lors de la troisième étude. Merci également à François Salmon, qui m'a épaulé lors des enregistrements de réponses impulsionnelles.

Je tiens à remercier ma famille, qui a toujours su encourager et supporter mes choix d'orientation bien que ceux-ci aient imposé une migration à quelques 750 km de mon berceau. Merci également à tous mes amis, ceux des Landes, ceux de Bordeaux, ceux de Brest et ceux d'ailleurs. Ils sont heureusement trop nombreux pour être cités, mais tous peuvent se reconnaître dans ces lignes et, je l'espère, prendre conscience de l'importance du rôle qu'ils jouent ou ont joué dans cette aventure. Petite mention spéciale à Olivier, qui a dû répéter inlassablement trois mots en boucle deux après-midi durant pour les bienfaits de la science. Et merci à Sylvain.

Enfin, je remercie Justine du fond du cœur. Si les trois années que je viens de vivre étaient si belles, c'est aussi parce que nous les avons partagées.

Table des matières

Introduction **1**

1 Etat de l'art **3**

1.1 Bases physiologiques de l'audition 3

1.2 Localisation auditive 8

1.2.1 Localisation en azimut 8

Différences interaurales de temps 8

Différences interaurales de niveau 11

Utilisation dynamique des différences interaurales 12

1.2.2 Localisation en élévation 13

1.2.3 Perception de distance 15

Niveau sonore 15

Rapport champ direct à champ réverbéré 17

Indices spectraux 18

Indices binauraux 18

1.2.4 Performances de localisation auditive 19

En azimut 20

En élévation 21

En distance 22

1.2.5 Interactions audiovisuelles et localisation auditive 23

Localisation visuelle 24

Performances de localisation visuelle 27

Localisation en présentation audiovisuelle 28

1.3 Perception d'intensité sonore 32

1.3.1 Sonie et isosonie : le sone et le phone 32

1.3.2 Principaux facteurs physiques influant sur la sonie monaurale 33

Niveau sonore 33

Fréquence 35

Durée 36

Largeur de bande 38

1.3.3 Sonie de sons non-stationnaires 40

1.3.4 Sonie binaurale 42

Gain binaural et rapport de sonie binaural-à-monaural 42

1.3.5 Modéliser la sonie 44

1.3.6 Sonie dans la vie de tous les jours . . . 49
Nature du stimulus . . . 49
Interactions multisensorielles . . . 50
Contexte d'écoute . . . 50
1.4 Interactions entre position et sonie . . . 52
1.4.1 En azimut et en élévation . . . 52
Sonie directionnelle . . . 52
Démasquage binaural . . . 55
1.4.2 En distance . . . 58

I Sonie et azimut 67

2 Étude de l'effet de l'ITD sur la sonie de sons purs BF à différents niveaux 69
2.1 Introduction . . . 69
2.2 Matériel et méthode . . . 69
2.2.1 Stimuli . . . 69
2.2.2 Procédure . . . 70
2.3 Résultats & Discussion . . . 72

3 Influence d'un bruit externe sur l'effet de l'ITD sur la sonie 77
3.1 Introduction . . . 77
3.2 Matériel et méthode . . . 78
3.2.1 Stimuli . . . 78
3.2.2 Procédure . . . 79
3.3 Résultats & Discussion . . . 80

4 Sonie directionnelle de bruits à bande étroite en conditions d'écoute naturelle 87
4.1 Introduction . . . 87
4.2 Matériel et méthode . . . 88
4.2.1 Stimulus . . . 89
4.2.2 Procédure . . . 89
4.2.3 Présentation réelle . . . 90
4.2.4 Présentation virtuelle . . . 91
4.3 Résultats & Discussion . . . 94
4.3.1 Résultats . . . 94
4.3.2 Discussion . . . 99
Effet des Conditions de présentation . . . 99
Indices auditifs . . . 99
Indices visuels . . . 100
Autres facteurs . . . 101
4.3.3 Conclusion . . . 102

Conclusion de la première partie **103**

II Sonie et distance **105**

5 Sonie *aux oreilles* de bruit en fonction de la distance **107**
5.1 Introduction . . . 107
5.2 Matériel et méthode . . . 108
5.2.1 Stimuli auditifs . . . 110
5.2.2 Stimuli visuels . . . 111
5.2.3 Procédure . . . 112
5.3 Résultats & Discussion . . . 116
5.3.1 Estimations de sonie *aux oreilles* . . . 116
5.3.2 Estimations de distance . . . 120
5.3.3 Conclusion . . . 122

6 Sonie *à la source* de bruit en fonction de la distance **123**
6.1 Introduction . . . 123
6.2 Matériel et méthode . . . 124
6.2.1 Stimuli auditifs . . . 124
6.2.2 Stimuli visuels . . . 124
6.2.3 Procédure . . . 125
6.3 Résultats & Discussion . . . 126

7 Sonie *aux oreilles* et sonie *à la source* de signaux de parole en fonction de la distance **133**
7.1 Introduction . . . 133
7.2 Matériel et méthode . . . 134
7.2.1 Stimuli auditifs . . . 134
7.2.2 Stimuli visuels . . . 137
7.2.3 Procédure . . . 138
7.3 Résultats & Discussion . . . 140
7.3.1 Estimations de sonie *aux oreilles* . . . 140
7.3.2 Estimations de sonie *à la source* . . . 145
7.3.3 Conclusion . . . 148

Conclusion de la deuxième partie **151**

Conclusion générale **153**

A Création des stimuli binauraux utilisés pour les expériences sur la distance **155**

B Mesures effectuées dans les salles **159**

Introduction

La sonie décrit la sensation de force sonore engendrée par un son. Ainsi, plus un auditeur perçoit un son comme étant fort, plus la sonie correspondant à ce son est élevée.

La sonie est une grandeur subjective qui dépend de nombreux paramètres et qui ne peut être mesurée avec des outils de mesure physique. La caractéristique physique du son ayant le rapport le plus évident avec la sonie est son niveau de pression sonore. Plus le niveau d'un son est élevé, plus la sensation de force sonore qu'il engendre est susceptible d'être importante. Cependant, il n'existe pas de relation directe entre une caractéristique donnée d'un son (pas même le niveau sonore) et sa sonie. Ainsi, deux sons de même niveau mais différents par ailleurs (fréquence, largeur de bande, répartition spectrale de l'énergie, évolutions temporelles de l'amplitude ou du spectre, etc.) peuvent évoquer des sonies radicalement différentes.

Durant leurs trajets entre la source et chaque oreille d'un auditeur, les sons subissent des modifications caractéristiques, dépendantes de la position relative de la source par rapport à l'auditeur. Certaines de ces modifications sont susceptibles d'affecter la sonie des sons émis, donnant lieu à des relations entre la position (voire à la localisation qui s'ensuit) d'une source sonore et la sonie des sons qu'elle émet.

La sonie directionnelle décrit les relations entre la direction d'incidence d'un son (du point de vue de l'auditeur) et la sonie. Ces relations sont largement expliquées par les niveaux de pression aux oreilles de l'auditeur, qui dépendent de cette direction surtout en hautes fréquences. La sonie est également affectée par la différence interaurale de temps (ITD), qui est un indice de localisation en basses fréquences. La première partie de cette thèse s'est intéressée aux relations entre sonie de sons basses fréquences et azimut dans le plan horizontal et se compose de trois études :

- La première étude s'est intéressée à l'effet de l'ITD sur la sonie de sons purs basses fréquences. Cette expérience avait pour objectif de comparer la sonie de sons purs comportant ou non une ITD lorsque ceux-ci étaient restitués à différents niveaux sonores.
- La seconde étude s'est intéressée à l'évolution de l'effet de l'ITD sur la sonie de sons purs basses fréquences lorsqu'ils étaient présentés dans du bruit diffusé à différents niveaux.
- La dernière étude de cette partie avait pour objectif d'étendre les effets observés dans des conditions d'écoute peu écologiques lors des deux premières expériences à des conditions d'écoute plus proches de conditions réelles. Les stimuli étaient alors diffusés par des sources réelles (haut-parleurs visibles positionnés autour des auditeurs) ou virtuelles (restituées au casque en binaural).

Les relations entre sonie et localisation auditive ne se limitent pas à la sonie directionnelle. Le niveau de pression sonore aux oreilles est directement dépendant de la distance de la source sonore. Plus une source est lointaine, moins le niveau aux oreilles est important. De ce fait, plus une source est lointaine, plus la sonie des sons qu'elle émet est susceptible d'être faible.

Au sens strict, la sonie désigne le niveau sonore perçu. Dans certaines études, les expérimentateurs prennent le parti d'orienter l'attention des auditeurs vers le niveau sonore reçu *aux oreilles* ou vers le niveau sonore émis *à la source* dans leurs évaluations de sonie. Ainsi, les évaluations de sonie *aux oreilles* correspondent à des évaluations de niveau sonore « entendu » tandis que les évaluations de sonie *à la source* sont équivalentes à des évaluations de puissance sonore d'une source. La sonie *à la source* peut ainsi se révéler constante avec la distance de la source, malgré les variations de niveau sonore aux oreilles des auditeurs que la distance implique. La deuxième partie de cette thèse s'est intéressée aux relations entre ces différentes formes de sonie et la distance de sources sonores et se compose également de trois études (réalisées en réalité virtuelle) :

- La première étude a observé la sonie *aux oreilles* et la distance perçue de bruits large bande diffusés depuis différentes distances avec différents niveaux d'information visuelle et auditive.
- La seconde étude a utilisé le même dispositif expérimental que la première mais avait pour objectif d'évaluer les jugements de sonie *à la source* de ces sons.
- La dernière étude réalisée lors de cette thèse s'est intéressée à la sonie *à la source* et à la sonie *aux oreilles* de signaux de parole, toujours avec différents niveaux d'information visuelle et auditive.

Chapitre 1

Etat de l'art

1.1 Bases physiologiques de l'audition

La première partie de ce chapitre présentera brièvement les différents mécanismes physiologiques permettant à l'être humain de percevoir les sons qui lui parviennent. Ces sons subissent une série de modifications tout au long de leur trajet depuis l'extérieur de la tête jusqu'à leur codage en signaux électriques à l'intérieur de l'oreille. Ces modifications seront utilisées par le cerveau pour localiser les sources sonores ainsi que pour en coder l'intensité subjective, qui seront les points abordés par les trois autres sections de cet état de l'art.

L'oreille humaine peut être séparée en trois parties distinctes qui sont l'oreille externe, l'oreille moyenne et l'oreille interne (voir figure 1.1). Cette section décrit les principaux mécanismes d'acquisition d'un son par l'oreille et a été largement alimentée par le site *Voyage au centre de l'audition*, proposé par Pujol (2018).

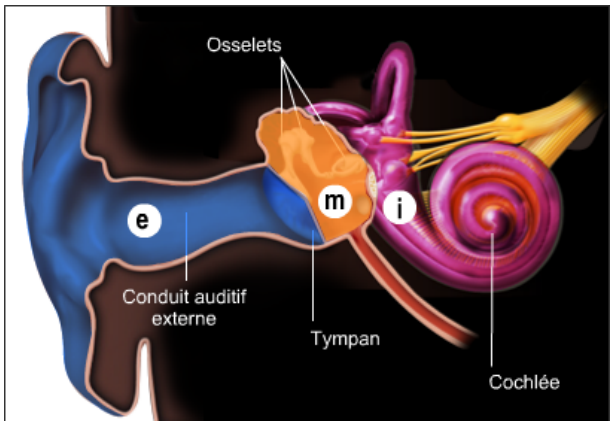


FIGURE 1.1 – Parties externe (e, en bleu), moyenne (m, en orange) et interne (i, en rose) de l'oreille humaine. Image extraite de *Voyage au Centre de l'Audition*, www.cochlea.org, par R. Pujol *et al.*, NeurOreille, Montpellier.

Oreille externe

En premier lieu, le pavillon, partie visible de l'oreille externe, guide les ondes sonores environnantes vers l'entrée du conduit auditif externe (CAE) tout en les modifiant. Ces ondes sonores se

dirigent ensuite dans le CAE qui, couplé à la conque, produit une résonance acoustique. Le tympan est situé en fin de CAE et joue le rôle d'interface entre oreille externe et moyenne. Il transmet l'onde sonore à cette dernière sous forme de vibration mécanique. Au cours de son voyage entre la surface du pavillon et le tympan, l'onde sonore subit une série de modifications caractéristiques. Dans un premier temps, la diffraction du pavillon entraîne un filtrage particulièrement prononcé en hautes-fréquences pour lesquelles il dépend largement de l'angle d'incidence de l'onde sonore. Le filtrage du pavillon est compris dans les fonctions de transfert de la tête (HRTF, pour *Head Related Transfer Functions*), permettant notamment la localisation des sources sonores. La forme de corne acoustique du CAE, dont le diamètre rétrécit entre l'entrée et la sortie (Stinson et Lawton, 1989), permet une amplification de l'onde sonore jusqu'à son arrivée sur le tympan. Cette amplification, basée notamment sur des phénomènes de résonance acoustique, est dépendante de la fréquence mais indépendante de l'angle d'incidence du son sur le pavillon. La figure 1.2 représentant le gain acoustique appliqué à l'onde sonore par l'oreille externe en fonction de la fréquence.

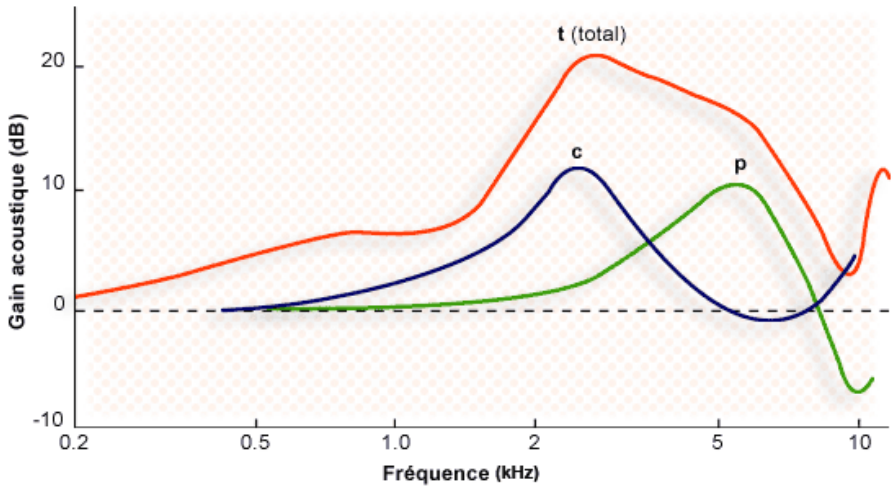


FIGURE 1.2 – Gain acoustique apporté par le pavillon (p, en vert) et le CAE (c, en bleu) pour un son dont la source est située dans le plan horizontal, à un azimut $\theta = 45°$. La somme des deux courbes (t, en rouge) représente l'effet global de l'oreille externe sur l'amplitude du signal perçu. Image extraite de *Voyage au Centre de l'Audition*, www.cochlea.org, par R. Pujol *et al.*, NeurOreille, Montpellier.

Oreille moyenne

Le tympan agit comme la membrane d'un microphone et vibre au contact de l'onde acoustique. Cette vibration mécanique est transmise à l'oreille interne par le biais de la chaîne ossiculaire composée de trois ossicules qui sont le marteau, l'enclume et l'étrier. Le marteau, fixé au centre du tympan, vibre avec celui-ci et transmet sa vibration à l'enclume qui en fait de même avec l'étrier, dont la platine est fixée à la fenêtre ovale, interface entre oreille moyenne et interne (voir figure 1.3). Le rapport de surface entre le tympan et la fenêtre ovale, de l'ordre de 20 pour 1 (la surface du tympan est d'environ $0.6\,\mathrm{cm}^2$ et celle de la fenêtre ovale d'environ $0.03\,\mathrm{cm}^2$),

entraîne une amplification de la pression, renforcée par le système de levier que forment les ossicules, et permet sa transduction depuis le tympan vers le milieu liquide de l'oreille interne. Si l'onde acoustique entrait directement en contact avec les liquides de l'oreille interne, 99.9% de l'énergie serait absorbée lors de la réflexion entre les deux milieux, entraînant une chute du niveau sonore de 30 dB. L'amplification mécanique de l'oreille moyenne est dépendante de la fréquence des ondes à transmettre et est de l'ordre d'une vingtaine de décibels aux fréquences les mieux amplifiées, comme en témoigne la fonction de transfert de l'oreille moyenne représentée figure 1.4.

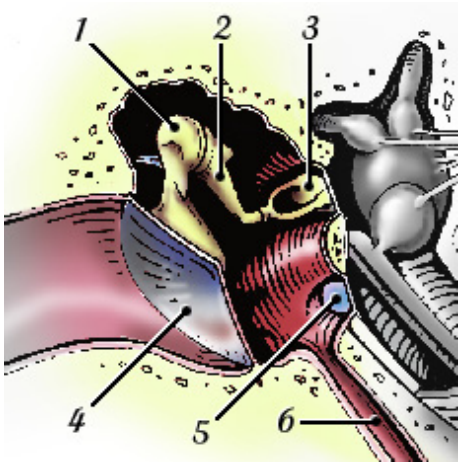


FIGURE 1.3 – Schéma de l'oreille moyenne représentant le tympan (4), le marteau (1), l'enclume (2) et l'étrier (3), dont la platine est fixée à la fenêtre ovale. La fenêtre ronde (5) communique elle aussi avec l'oreille interne et sert pour celle-ci d'échappement de pression. La trompe d'Eustache (6) communique avec la cavité buccale et permet de réguler la pression dans l'oreille moyenne. Image par S. Blatrix, extraite de *Voyage au Centre de l'Audition*, www.cochlea.org, par R. Pujol *et al.*, NeurOreille, Montpellier.

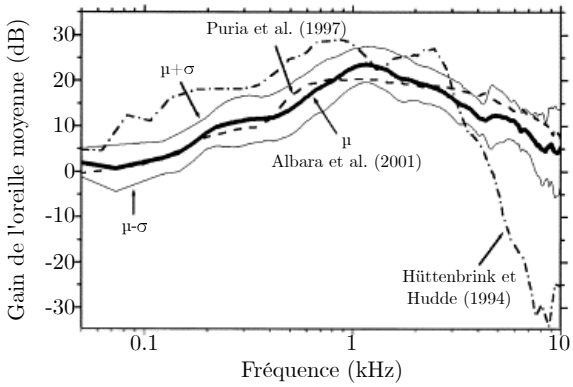


FIGURE 1.4 – Gain de l'oreille moyenne humaine. μ et σ désignent respectivement la moyenne et l'écart-type des mesures effectuées sur onze oreilles par Aibara *et al.* (2001). D'après Aibara *et al.* (2001), Puria *et al.* (1997) et Hüttenbrink et Hudde (1994).

Le marteau et l'étrier sont liés à des muscles qui ont un rôle protecteur dans la chaîne auditive. Si un son de fort niveau ($> 80\,\mathrm{dB}$) est détecté par l'oreille interne, le cerveau établit une boucle réflexe qui entraîne la contraction de ces muscles, rigidifiant la chaîne ossiculaire et atténuant son amplification pour le contenu basses fréquences. Ce réflexe, appelé réflexe stapédien du nom du muscle de l'étrier, a cependant le désavantage d'être mis en oeuvre après un certain délai (d'environ 150 ms pour des niveaux situés au seuil de déclenchement du réflexe et de 25 ms à 35 ms à plus forts niveaux d'après Dancer (1991)) nécessaire au voyage de l'information entre oreille interne, cerveau et oreille moyenne. L'appareil auditif est donc peu protégé contre les sons impulsifs de fort niveau, tels qu'un coup de feu, dont le temps d'attaque est très court. De plus, ce réflexe ne réduit la transmission des sons qu'en basses fréquences (Moore, 2012b), la rigidification de la chaîne ossiculaire ne permettant pas d'atténuer les fréquences supérieures à 4 kHz (Loth, 1997). Le réflexe stapédien est fatigable lors d'utilisations intensives. Enfin, le réflexe stapédien intervient également lorsque l'auditeur parle, dans le rôle d'atténuer le son qu'il perçoit de sa propre voix.

Oreille interne

Cette dernière partie de l'oreille est constituée de deux organes, à savoir le vestibule (organe de l'équilibration) et la cochlée (organe de l'audition). La cochlée est un organe creux, enroulé sur lui-même sur deux spires et demi allant de la base (où se situent les fenêtres ovale et ronde) à l'apex. Trois canalicules parcourent la cochlée en longueur, à savoir le canal cochléaire, rempli d'un liquide appelé endolymphe, la rampe vestibulaire et la rampe tympanique, toutes deux remplie d'un liquide appelé périlymphe.

Lorsque la fenêtre ovale entre en mouvement, entraînée par l'étrier, elle transmet sa vibration à la périlymphe contenue dans la rampe vestibulaire au niveau de la base de la cochlée. Cette vibration se propage le long de la cochlée et fait vibrer une membrane, la membrane basilaire, avec une amplitude progressive qui atteint un maximum à l'endroit où la vibration passe de la rampe vestibulaire à la rampe tympanique avant de se diriger vers la fenêtre ronde. La zone dans laquelle ce phénomène se produit dépend de la fréquence de l'onde, les basses fréquences agissant à l'apex et les hautes fréquences à la base de la cochlée comme schématisé figure 1.5. Cette répartition par fréquence le long de la membrane basilaire est appelée tonotopie cochléaire.

La transduction entre la vibration de la périlymphe et le signal électrique envoyé au cerveau est effectuée par des cellules, les cellules ciliées internes (CCI) se trouvant sur la membrane basilaire. Lorsque la membrane basilaire se déforme, les cils des CCI entrent en contact avec une autre membrane, la membrane tectoriale. Les CCI délivrent alors un neurotransmetteur, signalant au cerveau qu'elles ont capté un son. La cochlée possède également un mécanisme actif d'amplification efficace pour les sons de faible niveau. Pour représenter schématiquement ce phénomène, d'autres cellules, les cellules ciliées externes (CCE), tirent sur la membrane tectoriale pour favoriser son entrée en contact avec les cils des CCI. Ce phénomène amplifie (jusqu'à $+50\,\mathrm{dB}$) la vibration de la membrane basilaire sur une portion très étroite autour de la fréquence du son qui la stimule. Puisque le codage de la fréquence du son se fait en fonction de la zone excitée de la cochlée, ce mécanisme a également pour effet d'affiner le codage en fréquence sur la fréquence

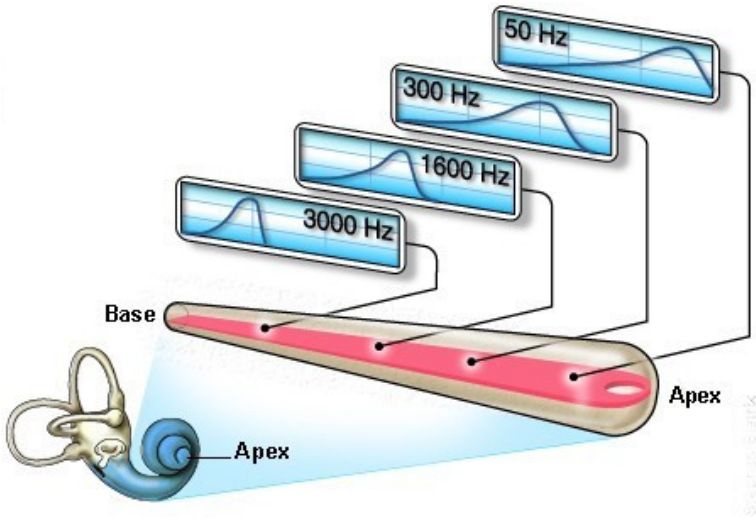


FIGURE 1.5 – Vibration de la membrane basilaire en fonction de la fréquence de l'onde sonore perçue. La zone où se trouve le maximum d'amplitude de vibration de la membrane basilaire dépend de la fréquence de l'onde. Image extraite de *Voyage au Centre de l'Audition*, www.cochlea.org, par R. Pujol *et al.*, NeurOreille, Montpellier.

de stimulation. La sélectivité fréquentielle, décrivant la capacité du système auditif à discriminer deux sons de fréquences différentes, sera ainsi améliorée par ce mécanisme actif d'amplification des sons.

1.2 Localisation auditive

La seconde partie de ce chapitre présente les différents mécanismes permettant à l'être humain de localiser dans l'espace les sources des sons qu'il entend. Lors de son trajet entre sa source et les oreilles d'un auditeur, un son voit ses paramètres physiques modifiés par l'environnement dans lequel il évolue ainsi que par le corps de l'auditeur. Ces modifications caractéristiques peuvent être utilisées par l'auditeur pour déterminer en partie la position de la source sonore. De plus, les chemins reliant une source sonore à chacune des deux oreilles de l'auditeur sont différents puisque ses deux oreilles sont situées à des positions différentes de l'espace. De ce fait, un même son émis par une source unique sera capté par ses deux oreilles sous forme de deux signaux différents. Ces différences interaurales, qui résultent des différences de parcours entre la source et chacune des deux oreilles, sont largement utilisées dans la détermination de la direction de provenance des sons. Enfin, la localisation auditive peut être appuyée voire dirigée par d'autres sens, en particulier la vue, lorsque le stimulus est présenté de façon multimodale.

1.2.1 Localisation en azimut

La localisation d'un son en azimut est rendue possible par les différences physiques distinguant le son atteignant notre tympan gauche et le son atteignant notre tympan droit. Lorsqu'une onde sonore provient de la gauche d'un auditeur, elle atteint d'abord son oreille gauche puis après un certain laps de temps son oreille droite, créant ainsi une différence interaurale de temps (ITD, pour *Interaural Time Difference*) entre les signaux perçus par les deux oreilles. De plus le corps de l'auditeur, en particulier sa tête, se trouve sur le chemin de l'onde atteignant l'oreille droite puisqu'il se situe entre celle-ci et la source sonore, située à gauche. Le corps va alors atténuer l'onde sonore lors de son trajet vers l'oreille droite, créant une différence interaurale de niveau (ILD, pour *Interaural Level Difference*) entre les signaux perçus par les deux oreilles. Le rôle de ces différences interaurales dans la localisation auditive a été proposé au début du XXe siècle par Lord Rayleigh (1907) dans sa théorie dite « duplex ».

Différences interaurales de temps

Lorsqu'un son provient du côté d'un auditeur, il atteint d'abord l'oreille située du même côté – l'oreille ipsilatérale – puis l'autre oreille – l'oreille controlatérale – après un retard temporel dépendant de la différence de marche et de la célérité du son. La différence de marche, correspondant à la différence de longueur entre les chemins reliant la source sonore aux oreilles gauche et droite, dépend directement de la direction à laquelle est située la source. En modélisant la tête sous forme de sphère, Woodworth a modélisé la relation entre différence de marche et azimut d'une source lorsque celle-ci est suffisamment lointaine pour que les fronts d'onde qu'elle émet soient plans lorsqu'ils atteignent la tête de l'auditeur (Woodworth, 1954). La Figure 1.6 représente ce cas schématique lorsque la source est positionnée à un angle $\theta \neq 0°$ d'azimut. Le chemin à parcourir entre la source et l'oreille ipsilatérale (l'oreille gauche) est de même longueur que le chemin séparant la source et le point A. La différence de marche est donc le trajet que doit suivre le son entre le point A et l'oreille controlatérale, soit une distance $r \sin\theta + r \cdot \theta$.

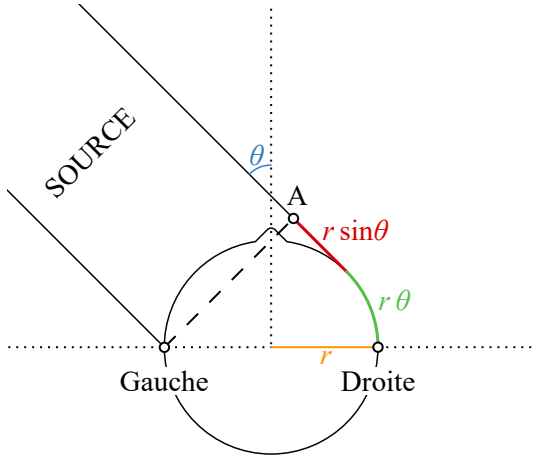


FIGURE 1.6 – Différence de marche entre les signaux atteignant l'oreille gauche et droite d'une tête de rayon r pour une source située à un azimut $\theta \neq 0°$.

De cette différence de marche naît une ITD, qui correspond au temps nécessaire à l'onde sonore pour parcourir ce trajet supplémentaire. Pour un signal sinusoïdal, cette ITD est équivalente à une différence interaurale de phase (IPD, pour *Interaural Phase Difference*). Par exemple, pour un son pur de fréquence 400 Hz dont la période dure 2500 µs, une ITD de 250 µs correspond à un retard d'un dixième de la période, soit une IPD de $\frac{2\pi}{10}$. L'IPD constitue ainsi un indice de localisation efficace en basses fréquences car l'ITD causée par la latéralisation d'une source reste petite devant la période de ces fréquences. Le système auditif peut dans ce cas sans ambiguïté identifier une même période du signal détectée par chacune des deux oreilles à des temps différents. Pour des fréquences plus élevées, l'ITD provoquée par la latéralisation d'une source devient du même ordre de grandeur, voire supérieure à la période temporelle du signal. Le système auditif ne peut alors plus faire correspondre une IPD donnée à une ITD donnée. Pour un son pur de fréquence 2000 Hz dont la période dure 500 µs, la même ITD de 250 µs induit désormais une IPD de π. Les sons captés par l'oreille gauche et droite sont en opposition de phase (séparés dans le temps d'une demi-période), de même que pour une ITD induisant une IPD de $-\pi$ (soit une ITD de -250 µs, en avance sur l'autre oreille). Une ambiguïté apparaît alors, puisque deux ITD distinctes engendrent deux IPD identiques à 2π près que le système auditif ne peut pas distinguer l'une de l'autre. Cette ambiguïté apparaît dès les fréquences auxquelles une IPD de π peut être engendrée par la position de la source sonore, soit les fréquences dont la période vaut approximativement le double de l'ITD maximale (atteinte pour une source localisée à $\pm 90°$ d'azimut). La valeur de cette ITD maximale peut être estimée pour une tête moyenne de rayon $r = 8.75$ cm en divisant la différence de marche correspondant à un azimut de $\theta = 90°$ par la célérité du son dans l'air ($340\,\mathrm{m\,s^{-1}}$) et donne une valeur de 660 µs. Ainsi, l'IPD constitue un indice de localisation sans ambiguïté approximativement jusqu'à la fréquence correspondant à une période 1.32 ms, soit 755 Hz. En pratique, cette fréquence limite peut être étendue à son double par l'intermédiaire de petits mouvements de tête ou lorsque la source est en mouvement (Moore, 2012c). L'interprétation de l'évolution dynamique de l'IPD couplée aux mouvements de tête effectués ou au mouvement perçu de la source sonore permet de lever l'ambiguïté dans des

cas où une IPD fixe peut correspondre à deux positions données d'une source (voir **utilisation dynamique des différences interaurales** ci-après).

Puisque la tête de l'auditeur se situe sur le chemin de l'onde atteignant l'oreille controlatérale, son parcours est perturbé par des phénomènes de diffraction à la surface de la tête en basses fréquences. Kuhn (1977) a modélisé l'ITD en prenant en compte ces phénomènes en utilisant les approches de Schwarz (1943) et de Rzhevkin (1963) sur l'évaluation de la pression totale à la surface d'une sphère rigide. En basses fréquences (lorsque $k \cdot r << 1$, où $k = \frac{\omega}{c_0}$ est le nombre d'onde, c_0 la célérité du son dans l'air et ω la pulsation du signal), l'ITD vaut ainsi :

$$\mathrm{ITD_{BF}} = \frac{3 \cdot r}{c_0} \sin(\theta) \tag{1.1}$$

En régime sinusoïdal, cette expression est équivalente à une différence de phase d'expression :

$$\mathrm{IPD_{BF}} = \mathrm{ITD_{BF}} \cdot \omega = 3 \cdot k \cdot r \sin(\theta) \tag{1.2}$$

En effectuant des mesures en chambre anéchoïque en délivrant des sons purs à un mannequin composé d'un torse et d'une tête artificiels, Kuhn (1977) a relevé une ITD indépendante de la fréquence en basses fréquences jusqu'à approximativement 500 Hz et en hautes fréquences au-delà d'approximativement 3 kHz. Plus récemment, Benichoux *et al.* (2016) ont observé des résultats similaires en analysant une large base de données d'HRTF mesurées sur des têtes humaines. Entre ces deux fréquences limites, l'ITD générée par une source positionnée à une direction unique varie en fonction de la fréquence. Il est alors impossible, à ces fréquences, d'établir une relation directe entre direction d'incidence et ITD.

En hautes fréquences, l'utilisation de l'ITD n'est pas possible puisque les différences interaurales de temps sont grandes devant la période des sons. En revanche, pour des sons non-stationnaires, le retard induit par la position d'une source est perceptible au niveau de l'enveloppe du signal. La sensibilité du système auditif humain aux délais interauraux d'enveloppe a été relevée pour de nombreux signaux non-stationnaires (Middlebrooks et Green, 1991) mais deviendrait moins efficace lorsque la fréquence de la porteuse se situe au-delà de 4 kHz ou lorsque la fréquence de modulation se situe au-delà de 500 Hz. La valeur de l'ITD est, à ces hautes fréquences, convenablement estimée par le modèle de Woodworth (1954), qui ne prend en compte que la différence de marche entre les signaux captés par les deux oreilles :

$$\mathrm{ITD_{HF}} = \frac{a}{c_0}(\theta + \sin(\theta)) \tag{1.3}$$

Bien que le système auditif soit sensible à ces différences au niveau de l'enveloppe, leur utilité en tant qu'indice de localisation n'est pas pleinement définie. Middlebrooks et Green (1990) ont effectué des mesures de délais interauraux d'enveloppe au niveau du CAE pour différentes positions d'une source sonore mobile. Ces mesures montrent une augmentation monotone de ces délais interauraux au niveau de l'enveloppe des sons lorsque l'azimut de la source augmente. D'après cette observation, ces délais pourraient apporter au système auditif une information utile

à la localisation de la source sonore. Pourtant, ils engendrent une faible sensation de latéralisation (Blauert, 1982; Bernstein et Trahiotis, 1985; Trahiotis et Bernstein, 1986) nécessitant un temps d'attaque court au niveau des variations d'enveloppe (Dietz *et al.*, 2015).

Blauert (1997b) a émis l'hypothèse selon laquelle le système auditif n'évalue pas chaque détail des différences interaurales complexes entre les signaux, mais identifie des attributs plus simples desquels il peut aisément extraire l'information de localisation. Ainsi, lorsqu'un son complexe est perçu par le système auditif, ce dernier n'évalue pas chaque fine variation interaurale de phase pour chaque fréquence du spectre. Plutôt, il détermine une ITD globale sur les composantes basses fréquences du spectre sur laquelle il base la sensation de latéralisation du son (Kulkarni *et al.*, 1999).

Différences interaurales de niveau

En plus du retard temporel dû à la différence de marche, le positionnement d'une source sonore à un azimut non nul a pour effet de placer la tête de l'auditeur entre la source et l'oreille controlatérale. L'onde sonore en direction de cette dernière subit de ce fait l'ombre acoustique de la tête et voit son niveau diminuer, donnant naissance à une ILD. En basses fréquences, la tête filtre peu les sons émis par des sources lointaines qui la contournent puisque ses dimensions sont petites devant leur longueur d'onde. L'ILD est de ce fait faible et ne constitue a priori pas un indice de localisation. En hautes fréquences, l'ombre acoustique de la tête n'est plus négligeable. Les ondes incidentes sont réfléchies par la surface de la tête, engendrant des ILD allant jusqu'à environ 35 dB lorsque la source est située à 90° d'azimut d'après les mesures réalisées à 10 kHz par Middlebrooks *et al.* (1989). Feddersen *et al.* (1957) ont relevé que la latéralisation induite par une ILD fixe dépendait de la fréquence des sons purs sur lesquels ils l'appliquaient, révélant que le système auditif prend en compte la fréquence du son qu'il perçoit dans l'utilisation de l'ILD comme indice de localisation. En hautes fréquences, les phénomènes de diffraction au niveau du corps de l'auditeur, en particulier de ses pavillons, sont complexes et mènent à une évolution complexe de l'ILD avec la fréquence. Cependant, comme pour l'ITD et en accord avec la supposition de Blauert (1997b), le système auditif n'utilise pas chaque variation complexe d'ILD correspondant à chaque fréquence du spectre pour en tirer l'information de localisation. Plutôt, il intègre l'ILD observée sur des bandes de fréquences discrètes (Macpherson et Middlebrooks, 2002) pour en tirer une information globale.

Dans la nature, les sons basses fréquences ne sont pas ou très peu filtrés par la tête et pour ces sons l'ILD est donc nulle ou très faible quelle que soit la position de leur source. Cependant, lorsque des sons basses fréquences sont présentés au casque avec une ILD, une sensation de latéralisation des sons émerge. Il est alors possible de comparer les latéralisations induites par l'ITD et l'ILD (Klemm, 1920; Wittmann, 1925; Harris, 1960). Cette relation dépend de nombreux paramètres et ne peut pas être décrite par une fonction linéaire simple (pour une revue complète, voir Blauert (1997b)).

L'efficacité de l'ITD en basses fréquences et celle de l'ILD en hautes fréquences constituent les fondements de la théorie « duplex » de Lord Rayleigh (1907), qui stipule que le système auditif utilise l'un de ces indices selon la gamme fréquentielle du son auquel il est soumis. Pour des sons large bande, il semblerait que la localisation se base en priorité sur l'information délivrée par l'ITD du contenu basses fréquences (Carlile, 1996).

Lorsque la source sonore est très proche de l'auditeur (dans un périmètre inférieur à 1 m), les fronts d'onde qu'elle émet ne peuvent plus être considérés comme plans. l'ILD augmente alors largement (Duda et Martens, 1998), notamment en basses fréquences, où elle peut atteindre jusqu'à 20 dB à 0.12 m (Brungart et Rabinowitz, 1999). L'ILD devient alors un indice potentiel de localisation en distance pour ces sources proches (Brungart *et al.*, 1999a).

Utilisation dynamique des différences interaurales

Les différences interaurales de temps et de niveau renseignent l'auditeur sur l'azimut de la source. Cependant, il existe différentes positions dans l'espace auxquelles une source sonore engendre des différences interaurales identiques. Si l'on garde la représentation schématisant la tête sous forme de sphère rigide et que l'on se place dans le plan horizontal, la différence de marche provoquée par une source localisée à un azimut θ est identique à celle provoquée par une source localisée à un azimut $\pi - \theta$ (voir Figure 1.7). L'ITD et l'ILD qui en découlent seront donc identiques. En conséquence, le système auditif est incapable de déterminer si la source est positionnée à l'azimut θ ou $\pi - \theta$, menant à une confusion avant/arrière. Il est cependant possible de lever cette confusion et de correctement déterminer l'angle d'incidence des sources en effectuant des mouvements de tête (Van Soest, 1929). Si l'on se place dans le cas de la figure 1.7 et que la tête de l'auditeur est fixée, il est impossible de différencier la source A et la source B au moyen de la différence de marche. Si l'auditeur tourne sa tête dans le sens des aiguilles d'une montre, l'oreille gauche se rapproche de la source A et s'éloigne de la source B, tandis que l'oreille droite se rapproche de la source B et s'éloigne de la source A. En conséquence, lors de ce mouvement, les différences interaurales de temps et de niveau augmentent si le son est délivré par la source A et diminuent si le son est émis par la source B et la confusion avant/arrière est levée. Cette interprétation de l'évolution de l'IPD avec les mouvements de tête permet également de doubler la fréquence limite de l'utilisation de l'IPD comme indice de localisation, qui atteint alors approximativement 1.5 kHz.

Si l'on sort du plan horizontal et que l'on considère un espace à trois dimensions, les positions de source menant à des différences interaurales constantes dessinent des cônes appelés cônes de confusion (travaux de von Hornbostel et Wertheimer (1920) sur les points menant à des différences de marche identiques, d'après Blauert (1997b), voir figure 1.8). Une multitude de positions de l'espace peut alors engendrer des ITD et ILD similaires. Les mouvements de tête peuvent ici aussi informer en partie l'auditeur sur la position d'une source. À titre d'exemple, une source localisée directement en face de l'auditeur engendrera schématiquement une ITD et une ILD nulle, de même qu'une source localisée directement au dessus de l'auditeur. Pour une source frontale, des mouvements de tête réalisés par l'auditeur autour de l'axe longitudinal (voir

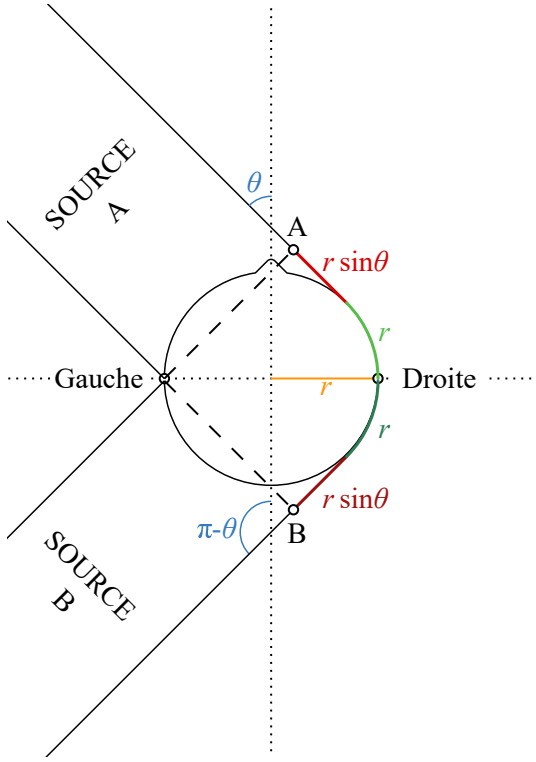


FIGURE 1.7 – Confusion avant/arrière dans le plan horizontal. Les sources A et B, bien que positionnées à des azimuts différents, engendrent la même différence de marche lorsque la tête est modélisée par une sphère rigide.

figure 1.9) engendreront de grandes variations d'ITD et d'ILD, ce qui ne sera pas le cas pour une source située au dessus de l'auditeur, pour laquelle l'ITD et l'ILD resteront nulles. Si la source est située sur le plan sagittal (voir 1.9) à une élévation $\varphi = 45°$, l'ITD et l'ILD sont nulles ici aussi. Les mêmes mouvements de tête engendreront dans ce cas des variations de ces indices dont l'amplitude sera moindre que dans le cas de la source située en face de l'auditeur ($\varphi = 0°$). Cet indice dynamique peut alors être utilisé pour déterminer l'élévation φ de la source sonore (Wallach, 1940). En pratique, la capacité de l'être humain à localiser les sources sonores en élévation excède les capacités que prédirait la seule utilisation dynamique des différences interaurales de temps et de niveau, par exemple pour des sons de durée trop courte pour qu'une interprétation dynamique des indices puisse être faite. Cette différence prouve les limites de la théorie duplex, qui ne peut expliquer à elle seule les performances de localisation auditive de l'être humain.

1.2.2 Localisation en élévation

La morphologie de l'être humain, en particulier de sa tête, de son torse et de ses oreilles, est composée de nombreuses irrégularités, l'éloignant du modèle de sphère rigide utilisé jusqu'alors. Ces irrégularités engendrent une interaction complexe entre la surface du corps et les ondes acoustiques incidentes lors de leur trajet vers les tympans. Cette interaction marque le son d'une empreinte caractéristique, propre à la morphologie de chacun et dépendante de la position relative

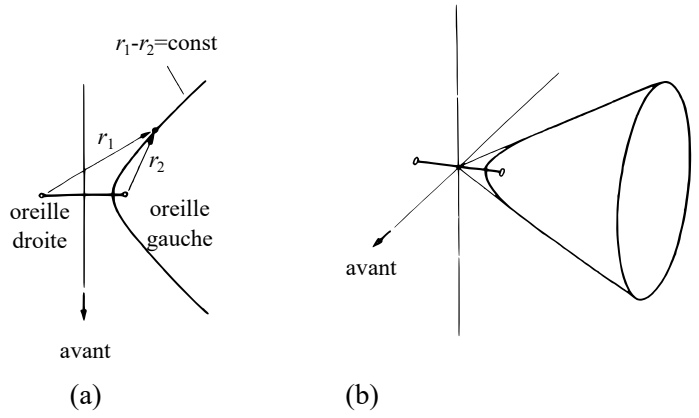


FIGURE 1.8 – Points menant à une différence de marche égale entre les deux oreilles, d'après le modèle de von Hornbostel et Wertheimer (1920). Les points forment une hyperbole dans le plan horizontal (a) et un cône dans un espace à trois dimensions (b). D'après Blauert (1997b).

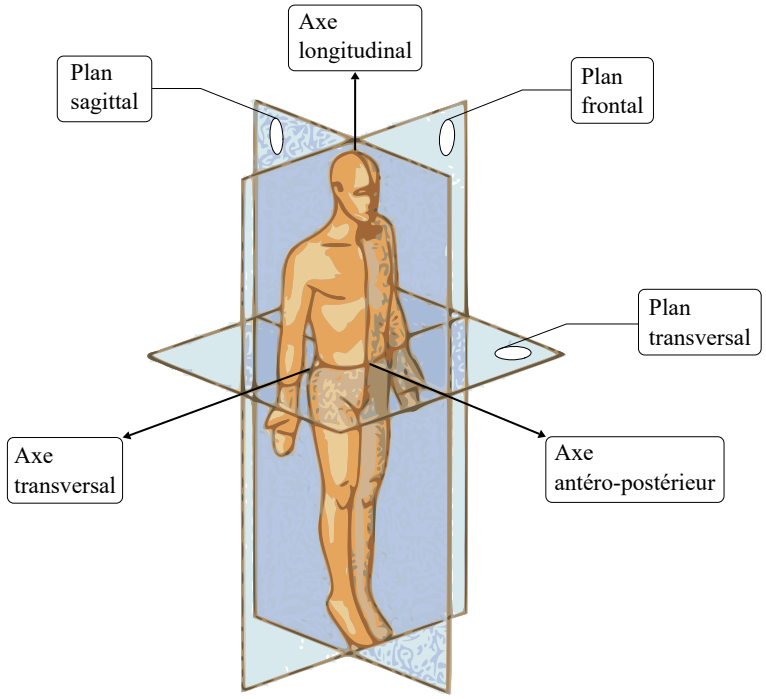


FIGURE 1.9 – Plans et axes du corps humain.

de la source par rapport à l'auditeur. Ainsi, à chacune des positions d'une sphère ayant pour centre le centre de la tête de l'auditeur correspond un couple de fonctions de transfert marquant les sons atteignant les oreilles droite et gauche. Ces fonctions de transfert sont les HRTF de l'auditeur, dont le spectre d'amplitude est directement défini par l'interaction entre le corps de l'auditeur et les ondes acoustiques. Le système auditif peut alors analyser le signal qu'il perçoit pour reconnaître ces indices spectraux. Une fois les empreintes identifiées, il y fait correspondre les HRTF associées et en déduit la position de la source sonore (voir Guillon (2009) pour un exposé détaillé). Ces indices spectraux sont des indices dits monauraux dans le sens où ils ne

reposent pas sur la comparaison des signaux perçus par chacune des deux oreilles. Le pavillon, grâce à sa géométrie complexe, joue un rôle important pour la localisation en élévation (Butler, 1969), pour lever la confusion avant/arrière, ou même en localisation en azimut (Batteau, 1967). Un recueil d'études concernant le rôle de l'oreille externe est disponible dans Moore (2012c).

Les indices spectraux ne sont utilisables que sur des signaux dont le spectre est relativement étendu. Prenons l'exemple d'une source se déplaçant à différentes positions dans le plan médian (engendrant ainsi des différences interaurales nulles). À chacune de ces positions correspond une paire d'HRTF (une HRTF par oreille) par laquelle le son que délivre la source sera modifié avant d'atteindre l'oreille correspondante. Si cette source délivre un son pur, dont le spectre n'est composé que d'une fréquence unique, le résultat des modifications avec chaque paire d'HRTF ne se traduira que par une différence de niveau. Le système auditif ne peut alors pas utiliser les indices spectraux pour identifier les HRTF. Pour les signaux à bande étroite, les modifications du spectre engendrées par les HRTF peuvent également ne pas être assez larges pour permettre au système auditif d'identifier les HRTF à partir d'indices spectraux (voir figure 1.10). Dans ce cas de figure, la localisation en élévation repose uniquement sur la zone du spectre qui contient l'énergie, indépendamment de la direction d'incidence des sons dans le plan médian. En conséquence, le système auditif peut pour ces signaux localiser une source sonore dans une direction totalement différente de sa direction réelle (Blauert, 1969).

1.2.3 Perception de distance

La perception de distance d'une source sonore se base sur différents indices, dont l'utilisation peut être propre à la localisation dans l'espace péripersonnel (à portée de main de l'auditeur, pour des sources approximativement situées à une distance $d < 1\,\mathrm{m}$ de l'auditeur) ou extrapersonnel (loin de l'auditeur). Une revue récente décrivant les différents indices de la perception auditive de la distance, son développement chez l'être humain, ses bases neurologiques ainsi que son utilisation pour des humains atteints de pertes auditives ou visuelles a été réalisée par Kolarik *et al.* (2016).

Niveau sonore

Le premier des indices de distance d'une source sonore est le niveau global du son qui parvient aux oreilles de l'auditeur. De manière générale, plus une source s'éloigne d'un auditeur, plus le son qu'elle émet sera faible lorsqu'il atteindra ce dernier et inversement. Ainsi, si le niveau d'un son aux oreilles d'un auditeur décroît, la distance que celui-ci perçoit de sa source augmente.

Le niveau aux oreilles de l'auditeur dépend à la fois de la distance de la source et de son niveau de diffusion. De ce fait, l'auditeur est susceptible de mal interpréter le niveau à ses oreilles en confondant distance et niveau de diffusion (Stevens et Guirao, 1962), rendant impossible l'estimation absolue de distance de la source basée sur ce seul indice. Cependant, si deux sources délivrant des sons identiques sont placées à différentes distances de l'auditeur, celui-ci peut comparer le niveau sonore qu'il perçoit de chacune d'elle pour déterminer laquelle des deux est

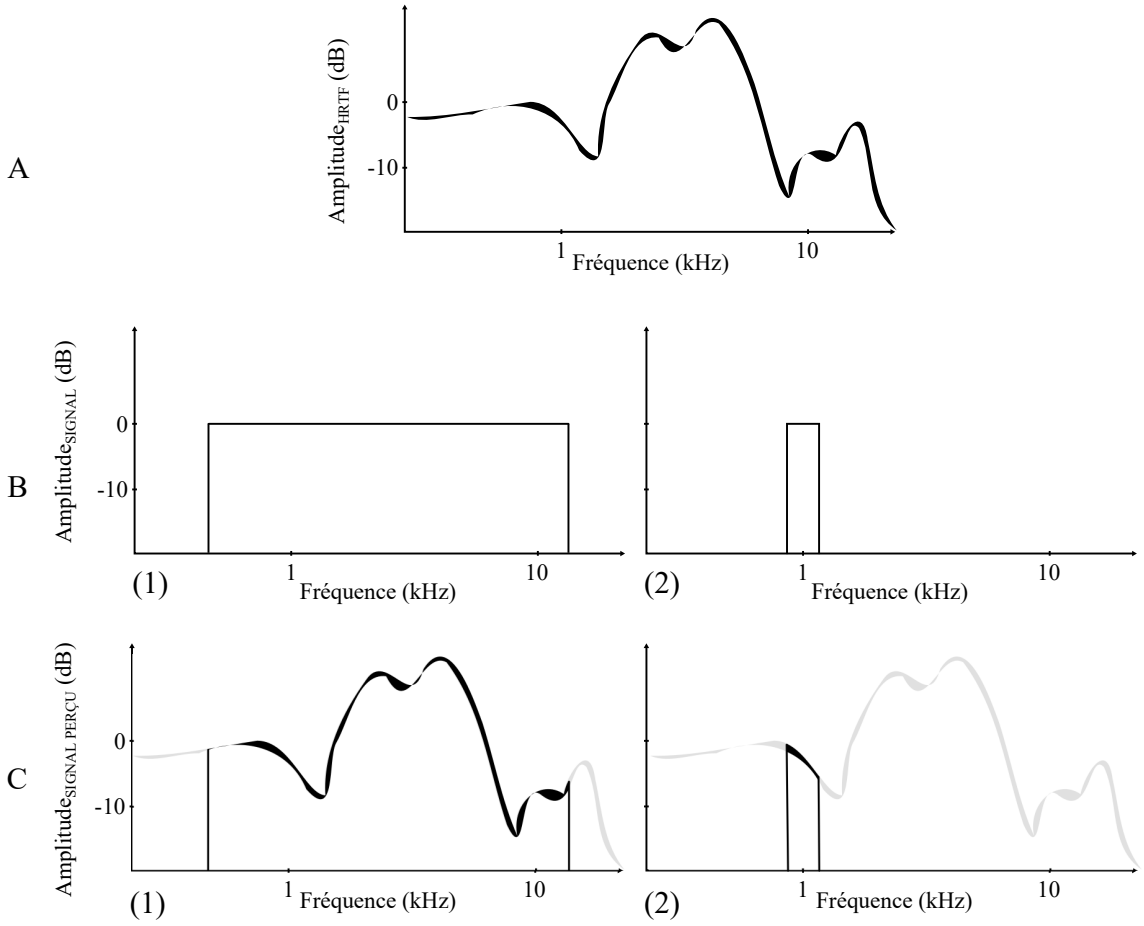


FIGURE 1.10 – Utilisation des indices spectraux dans la localisation auditive. La partie A représente un spectre d'amplitude schématique d'une HRTF correspondant à une oreille et une direction d'incidence Ψ donnée. La partie B représente deux signaux sonores, (1) étant un bruit blanc large bande et (2) un bruit blanc à bande étroite. Si ces sons sont diffusés par une source située dans la direction Ψ, ils seront modifiés par le spectre d'amplitude de la HRTF, représenté en A, et les signaux captés par l'oreille seront respectivement les signaux (1) et (2) représentés en C. Dans le cas du signal large bande (1), le spectre d'amplitude de la HRTF est identifiable et le son peut être localisé comme provenant de la direction Ψ, ce qui n'est pas le cas pour le signal à bande étroite (2).

la plus proche. Le niveau sonore est donc un indice relatif de distance, qui peut être utilisé pour des sources localisées dans l'espace péripersonnel et extrapersonnel et à tout azimut.

Le niveau sonore peut permettre une estimation absolue correcte de distance dans le cas où la puissance de la source sonore est connue. C'est le cas par exemple pour la voix, qui contient des informations intrinsèques sur l'effort vocal de sa source (un chuchotement est aisément identifié comme ayant un niveau à la source plus faible qu'un cri). Il est alors possible, pour l'auditeur, de mettre en commun l'effort vocal ressenti et le niveau perçu pour déterminer la distance de la source sonore (Brungart et Scott, 2001). La perception auditive de distance est de ce fait plus précise pour de la voix que pour des signaux non-familiers tels que du bruit (Zahorik, 2002). Pour de la parole inversée temporellement (qui contient les mêmes propriétés acoustiques que de

la parole normale, mais dont le timbre n'est pas familier à l'être humain), les performances en estimation de distance sont plus faibles que pour de la parole normale (McGregor *et al.*, 1985; Brungart et Scott, 2001). L'étude de Wisniewski *et al.* (2012) indique que ces différences de performance ne sont pas liées à la compréhension du lexique porté par la parole, qui n'est possible que pour la voix normale. Dans cette expérience, des participants anglophones estimaient la distance apparente de voix, inversées temporellement ou non, parlées en anglais ou en bengali. Les performances étaient meilleures pour les paroles normales qu'inversées temporellement, sans distinction entre l'anglais ou le bengali. Les bonnes performances d'estimation de la distance pour la voix semblent donc être expliquées par les indices contenus par le timbre, au sens de la combinaison des évolutions temporelles et spectrales du son, qui permettent à l'auditeur d'interpréter l'effort vocal du locuteur indépendamment de la compréhension du message que la parole véhicule.

Rapport champ direct à champ réverbéré

Lorsqu'un auditeur entend le son émis par une source sonore dans une salle, il perçoit l'onde sonore voyageant directement de la source à ses oreilles (le champ direct) ainsi que toutes les ondes qui atteignent ses oreilles après s'être réfléchies sur les différentes surfaces de l'environnement (murs, objets, sol, etc.). La somme de toutes ces ondes réfléchies constitue le champ réverbéré, dont l'énergie est approximativement constante en tous points de la salle si celle-ci est suffisamment réverbérante. Le son perçu par l'auditeur est alors constitué d'un champ direct, dont l'énergie décroît de 6 dB par doublement de distance et d'un champ réverbéré, dont l'énergie est approximativement constante avec la distance. Il est alors possible de déterminer la distance de la source sonore en évaluant le rapport entre l'énergie du champ direct et celle du champ réverbéré (DRR, pour *Direct-to-Reverberant energy Ratio*). Plus la distance séparant la source et l'auditeur augmente, plus l'énergie du champ direct est faible au niveau du point d'écoute tandis que l'énergie du champ réverbéré reste constante, menant à un faible DRR. Dans le cas où la source et l'auditeur sont proches, l'énergie du champ direct est forte et celle du champ réverbéré toujours constante, menant à un DRR maximal. Le DRR constitue ainsi un indice absolu dans la perception auditive de distance dont l'importance est grande dans les lieux clos (Mershon et King, 1975). Cette indice peut lui aussi être utilisé pour des sources localisées dans l'espace péripersonnel et extrapersonnel, pour des sources frontales et latérales, bien que le seuil de discrimination du DRR soit minimal autour de la distance critique, à laquelle le DRR vaut 0 dB (Larsen *et al.*, 2008).

Lorsque le niveau et le DRR sont manipulés de façon séparée l'un de l'autre, la discrimination entre deux sources de distances différentes est plus précise en utilisant le niveau que le DRR lorsque l'environnement est faiblement réverbérant et devient équivalente pour les deux indices lorsque l'environnement est hautement réverbérant (Kolarik *et al.*, 2013). Le seuil de discrimination du DRR seul est relativement élevé, Larsen *et al.* (2008) ayant mesuré des seuils de 2 à 3 dB à des DRR de 0 et +10 dB et de 6 à 8 dB pour des DRR de −10 et +20 dB. L'utilisation relative du DRR semble ainsi être plus précise autour de la distance critique (Jetzt, 1979) à

laquelle l'énergie du champ direct et celle du champ réverbéré sont égales (DRR de 0 dB) et moins efficace lorsque la source est soit très proche, soit très éloignée de l'auditeur. De plus, une large variabilité intra-sujet a été observée sur les jugements de distance apparente basés sur le DRR recueillis par Zahorik (2002). En définitive, le DRR semble être un indice peu performant en perception de distance relative, et le système auditif doit combiner l'information apportée par le DRR et par le niveau sonore pour réaliser un estimation précise de la distance. Zahorik (2005) suppose que dans cette mise en commun des deux indices, le DRR fournit une information absolue sur la distance de la source et que les petites variations de niveau sonore permettent au système auditif d'affiner l'estimation autour de la valeur absolue déterminée à partir du DRR.

Indices spectraux

Au fur et à mesure qu'une onde sonore se propage dans l'air, son énergie est peu à peu absorbée par le milieu. La quantité d'énergie ainsi absorbée est plus importante pour les hautes fréquences que pour les basses fréquences, mais reste toutefois relativement faible (de l'ordre de 3 à 4 dB tous les 100 m à 4 kHz d'après Ingård (1953)). À de longues distances (supérieures à 15 m) d'une source sonore, l'atténuation créée dans le haut du spectre devient cependant perceptible et constitue un indice de perception en distance (Blauert, 1997b). Un bruit large bande filtré en passe-bas paraîtra ainsi plus lointain qu'un bruit large bande filtré en passe-haut (Butler *et al.*, 1980). Little *et al.* (1992) ont effectué une expérience similaire avec des fréquences de coupure plus représentatives de la réalité. Les résultats de leur expérience indiquent que les participants jugeaient les bruits filtrés en passe-bas comme étant plus lointains seulement après plusieurs présentations des sons. Cette observation a mené les auteurs à estimer que l'utilisation d'indices spectraux dus à l'absorption des hautes fréquences par l'air est un indice relatif de distance de la source sonore.

En champ proche, la diffraction de la tête est dépendante de la fréquence et de la distance entre la source et l'auditeur. Il en résulte une accentuation des basses fréquences pour les sources très proches de l'auditeur (Brungart *et al.*, 1999b) et le contenu basses fréquences devient nécessaire pour localiser précisément les sources sonores très proches (Brungart, 1999).

En résumé, les indices spectraux renseignent l'auditeur sur la distance de la source en champ très proche (< 1 m) grâce à l'augmentation du contenu basses fréquences et pour des distances très grandes (> 15 m) grâce à la diminution du contenu hautes fréquences du signal sonore.

Indices binauraux

En champ proche, les basses fréquences du son augmentent à mesure que la distance entre la source et l'auditeur diminue. Si la source est localisée sur le côté de l'auditeur, les basses fréquences seront plus accentuées pour le son atteignant l'oreille ipsilatérale (proche de la source) que pour le son atteignant l'oreille controlatérale (plus éloignée de la source). Il en résulte une ILD dépendante de la distance, qui constitue un indice important de localisation pour les sources très proches (Brungart *et al.*, 1999b) nécessitant que le son contienne des basses fréquences

(Brungart, 1999). Cette observation semble de prime abord contraire à la théorie duplex, selon laquelle la localisation du contenu basses fréquences se fait exclusivement par le biais de l'ITD. Cette théorie est cependant fondée sur la localisation de la direction d'une source (et non de sa distance) et valable aux distances lointaines, où l'ILD est très faible en basses fréquences quelle que soit la position de la source sonore. Les résultats de Kopčo et Shinn-Cunningham (2011) présentent, dans un environnement réverbéré, les mêmes erreurs de localisation en distance pour des sources proches qu'elles soient frontales (sans ILD) ou latérales (avec ILD). Les auteurs émettent l'hypothèse que le système auditif base ses estimations de distance en champ proche sur le DRR (de façon indépendante de la direction de la source) même si l'ILD est présente et théoriquement exploitable. L'ITD varie très peu avec la distance (Duda et Martens, 1998), de même que l'ILD pour des sources situées à plus de 1 m de l'auditeur.

Lorsque la source sonore est relativement proche de l'auditeur, il existe également un phénomène de parallaxe acoustique entre les sons captés par les deux oreilles. L'angle entre la source et le plan sagittal passant par l'oreille gauche est différent de l'angle entre la source et le plan sagittal passant par l'oreille droite (si la source n'est pas localisée dans le plan médian), et cette différence augmente à mesure que la distance de la source diminue. Si l'on considère que la direction relative à chaque oreille peut être déterminée à partir de la modification des HRTF droite et gauche, il est possible de déterminer un angle de parallaxe d'après la différence entre les directions apparentes de chaque oreille, $\alpha_{\mathrm{L}} - \alpha_{\mathrm{R}}$, dépendant de la distance (voir figure 1.11). Lorsque l'on s'éloigne de l'espace péripersonnel, les HRTF sont quasi-indépendantes de la distance (Otani *et al.*, 2009). Les angles α_{L} et α_{R} ne varient plus significativement avec la distance de la source et l'angle de parallaxe ne permet plus de localiser la source en distance. Cette distance limite se situe approximativement à 1 m de l'auditeur d'après Kim *et al.* (2001). D'après Zahorik (2005), la parallaxe acoustique pourrait également expliquer les performances de localisation en distance de sources proches localisées dans le plan médian en l'absence d'indices liés à l'ILD ou au niveau, telles qu'observées par Ashmead *et al.* (1990).

Le tableau 1.1 résume les conditions dans lesquelles chaque indice auditif de distance vu ci-dessus peut être utilisé par le système auditif.

1.2.4 Performances de localisation auditive

Lorsqu'un auditeur évalue de façon absolue la position de la source d'un son qu'il entend, l'écart entre la position réelle et la position perçue détermine l'acuité du système auditif. L'acuité moyenne du système auditif pour une combinaison de paramètres donnée (position de la source, type de stimulus, etc.) peut alors être estimée en comparant la position moyenne perçue pour un grand nombre d'auditeurs à la position réelle de la source sonore. La dispersion des résultats autour de cette valeur moyenne renseigne sur la précision du système auditif, qui est habituellement représentée par l'écart type des valeurs de positions perçues (Letowski et Letowski, 2011). Enfin, la plus petite différence angulaire perceptible (MAA, pour *Minimum Audible Angle*) représente la capacité du système auditif à différencier des sources situées à deux positions angulaires différentes (relatives à l'auditeur).

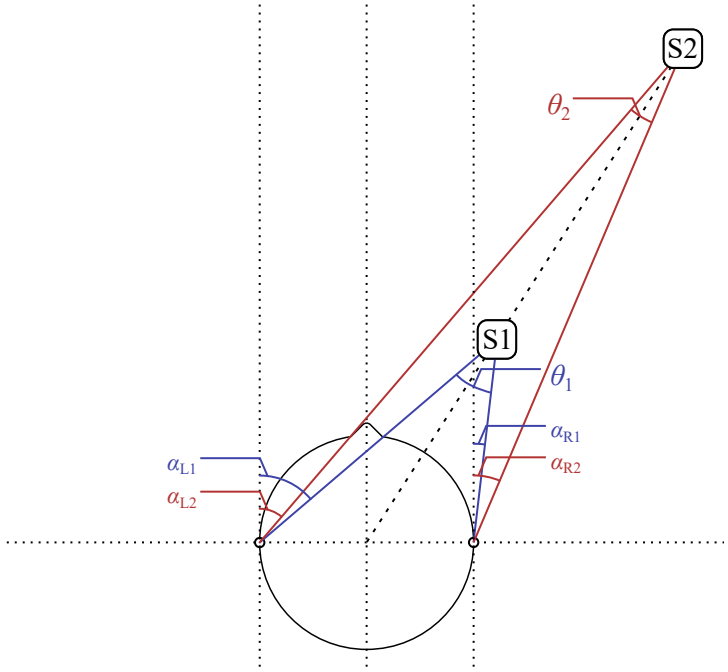


FIGURE 1.11 – Parallaxe acoustique pour deux sources S1 et S2 situées à un même azimut, l'une (S1) étant plus proche que l'autre (S2). La comparaison des azimuts apparents de chaque oreille ($\alpha_{L_{1,2}}$ et $\alpha_{R_{1,2}}$) permet au système auditif de déduire l'angle de parallaxe $\theta_{1,2}$ correspondant à chaque source. Cet angle est plus grand pour les sources proches que pour les sources lointaines ($\theta_1 > \theta_2$) et constitue un indice de localisation en distance.

En azimut

Dans deux expériences à grande échelle (regroupant respectivement 600 et 900 participants), Preibisch-Effenberger (1966) puis Haustein et Schirmer (1970), tels que décrits par Blauert (1997b), ont mesuré l'acuité et la précision du système auditif dans le plan horizontal. Leurs résultats, présentés figure 1.12, montrent une erreur de localisation bien plus faible, de l'ordre de 1°, pour les sources situées à l'avant (0°) ou à l'arrière (180°) des auditeurs que pour les sources situées sur les côtés (90° et 270°), où elle était de l'ordre de 10°. La précision de la localisation est également meilleure pour les sources situées devant et derrière les auditeurs, avec un avantage pour les sources avant (flou de ±3.6° contre ±5.5° à l'arrière), que pour les sources situées sur les côtés (flou d'environ ±10°). Des résultats similaires ont été obtenus par Carlile *et al.* (1997). Ces erreurs sont néanmoins dépendantes du type et de la fréquence des sons joués. Le système auditif semble notamment être moins précis entre 1.5 kHz et 3 kHz, possiblement car cette zone fréquentielle correspond à la zone de transition entre les indices de localisation en azimut (IPD et ILD) dans laquelle aucun de ces deux indices n'est pleinement efficace (Middlebrooks et Green, 1991).

Le MAA est également minimal lorsque la source sonore est située à un azimut de 0°. Depuis cette position, des déplacements représentant des variations angulaires de 1° en azimut peuvent être repérés à des fréquences inférieures à 1 kHz (Mills, 1958, 1972). Le MAA est bien plus élevé

	Indice auditif de distance			
Condition	Niveau	DRR	Indices spectraux	Indices binauraux
Environnement anéchoïque	✓	✗	✓	✓
Environnement réverbérant	✓	✓	✓	?
Jugements absolus de distance	✗	✓	✗	✓
Jugements relatifs de distance	✓	✓	✓	✓
Sources frontales	✓	✓	✓	?
Sources latérales	✓	✓	✓	✓
Espace péripersonnel	✓	✓	✓	✓
Espace extrapersonnel	✓	✓	✓ (>15 m)	✗

Table 1.1 – Résumé des conditions dans lesquelles chacun des indices auditifs de distance peut être utilisé. Chaque indice est exploitable dans les conditions pour lesquelles apparaît une coche (✓) et ne peut pas être exploité dans les conditions pour lesquelles apparaît une croix (✗). Un point d'interrogation (?) indique que l'efficacité de l'indice donné dans les conditions données n'est pas connue ou claire. D'après Kolarik *et al.* (2016).

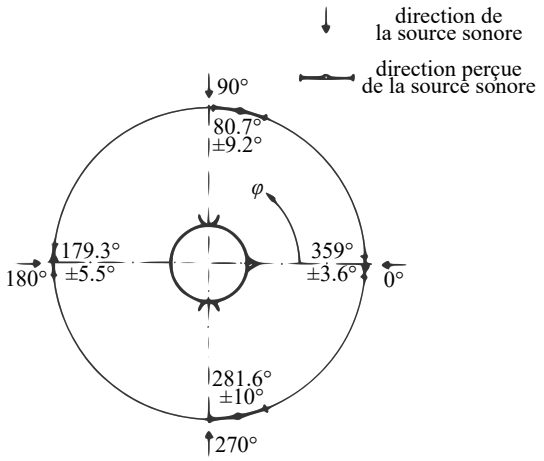


Figure 1.12 – Acuité et flou de localisation dans le plan horizontal pour des bruits blancs de durée 100 ms, mesurés sur 600 (Preibisch-Effenberger, 1966) puis 900 (Haustein et Schirmer, 1970) sujets dont la tête était immobilisée. D'après Blauert (1997b).

entre 1.5 kHz et 3 kHz, conformément aux observations faites dans les tâches de localisation. Lorsque la source est située sur le côté, le MAA augmente. Pour une source située à un azimut de 75°, l'auteur a mesuré un MAA de 7° à 8° aux fréquences inférieures à 1 kHz. Le MAA atteignait des valeurs trop élevées pour être mesuré aux fréquences allant de 1.5 kHz à 2 kHz à ce même azimut.

En élévation

Dans le plan médian, le système auditif ne peut déterminer la position qu'à partir d'indices spectraux, dont la fiabilité est moindre que celle des indices binauraux. Les performances de localisation sont de ce fait moins bonnes en élévation qu'en azimut et dépendent grandement

du type de stimulus utilisé, notamment de sa familiarité avec l'auditeur. Le flou de localisation en élévation à 0° d'azimut et d'élévation a été mesuré à environ 17° pour de la voix parlée par une personne non-familière des auditeurs par Blauert (1970), environ 9° pour de la voix parlée par une personne familière par Damaske et Wagener (1969) et 4° pour du bruit blanc par Wettschurek (1970). Les résultats de Damaske et Wagener (1969) montrent que l'erreur et le flou de localisation pour de la voix familière augmentent avec l'élévation. Ces résultats sont présentés figure 1.13 tels que transcrits par Blauert (1997b). Cette dépendance de la familiarité de la source est compréhensible puisque la localisation en élévation se fait par l'identification, sur le signal perçu, de filtres (les HRTF) ayant modifié le spectre du signal émis par la source. Si le signal original est connu (car sa source est familière), il est plus évident pour le système auditif d'identifier ces filtres. Si l'on reprend l'exemple d'une voix parlée par un locuteur qu'il connaît, le système auditif peut comparer le signal en sortie de filtre (la voix qu'il perçoit) avec le signal original (la voix telle qu'il la connaît) pour déterminer la nature du filtre. Pour un son non-familier, le système auditif doit déduire la nature du filtre en ne disposant que du signal à sa sortie. Le faible flou de localisation observé par Wettschurek (1970) pour du bruit blanc pourrait être expliqué par le fait que ce stimulus couvre l'ensemble du spectre, rendant plus simple l'utilisation des indices spectraux par le système auditif (voir figure 1.10).

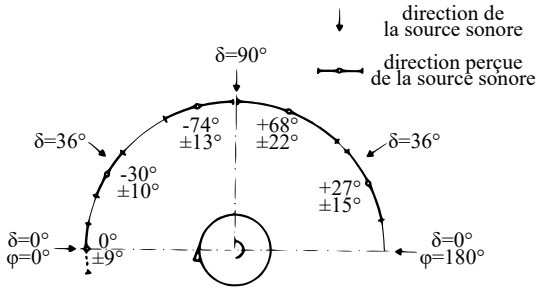


FIGURE 1.13 – Acuité et flou de localisation dans le plan médian pour de la voix parlée par un locuteur familier des auditeurs (Damaske et Wagener, 1969). D'après Blauert (1997b).

Wettschureck (1973) a mesuré le MAA pour des sources se déplaçant en élévation. Ses résultats montrent que le MAA en élévation atteignait 3° pour du bruit blanc dont la source était située devant l'auditeur (à un point de départ de 0° en azimut et en élévation). Lorsque les hautes-fréquences du signal étaient enlevées, le MAA augmentait considérablement. Lorsque le point de départ de la source se situait au-dessus des auditeurs, le MAA en élévation augmentait jusqu'à environ 10°.

En distance

L'étude des performances de localisation a longtemps été étudiée davantage pour la localisation en azimut et en élévation que pour la localisation en distance. L'étude de la localisation en distance comporte certains désavantages d'ordre pratique. À titre d'exemple, le fait de placer un certain nombre de haut-parleurs à différentes distance d'auditeurs sur un même axe aurait pour effet de placer les haut-parleurs les plus proches sur le trajet des ondes acoustiques émises par les

haut-parleurs lointains. Depuis une vingtaine d'années, les avancées technologiques permettent de pallier ce genre de problème en restituant les sons au casque tout en respectant la nature des stimuli sonores. Il est par exemple possible d'effectuer des enregistrements binauraux des sons joués par un haut-parleur positionné à différentes distances (Zahorik, 2002) afin de manipuler indépendamment les différents indices auditifs de distance et d'en évaluer les rôles dans la localisation auditive.

De manière globale, la distance perçue par le système auditif a tendance à être surestimée pour des sources proches et sous-estimée pour des sources lointaines (Kearney *et al.*, 2012; Zahorik et Wightman, 2001; Zahorik, 2002). Cette erreur systématique de la distance perçue par rapport à la distance réelle a été relevée pour de nombreuses combinaisons de stimuli, d'environnements acoustiques et de procédures d'évaluation (Zahorik, 2005). De façon générale, la relation entre distance estimée et distance réelle peut être modélisée par une fonction puissance compressive de forme :

$$r' = k \cdot r^a \tag{1.4}$$

où r' est la distance estimée, r la distance réelle de la source, k et a des constantes ajustables où $a < 1$. De nombreuses méthodes existent pour obtenir une estimation de distance de la part d'un auditeur et ces méthodes peuvent, pour un même stimulus, mener à des estimations très différentes. Les protocoles d'action dirigée par la perception (voir figure 1.14) mènent notamment à des estimations de distance plus précises que les estimations reportées verbalement. Dans ces derniers, l'auditeur donne une estimation, dans une unité de mesure, de la distance à laquelle se trouve la source (Fukusima *et al.*, 1997; Loomis *et al.*, 1998; Russell et Schneider, 2006; Andre et Rogers, 2006). Les actions dirigées donnent également des estimations plus précises que celles obtenues par actions imaginées. Dans ces dernières, les participants imaginent par exemple se déplacer sur la position de la source, et le temps qu'ils passent à imaginer s'y déplacer est utilisé pour estimer la distance perçue de la source. Des biais systématiques sont susceptibles d'expliquer ces différences, apparaissant lorsque les auditeurs utilisent une unité de mesure pour décrire une représentation mentale de distance ou lors de la conversion d'une grandeur physique directement mesurée (par exemple le temps que met un participant à s'imaginer marcher sur la position de la source) en une estimation indirecte de distance (Grechkin *et al.*, 2010). Il est courant d'observer que le protocole utilisé influence les estimations reportées. Proffitt *et al.* (1995) ont par exemple obtenu des estimations correctes de pente apparente de plans inclinés lorsque les participants reportaient leur estimation en faisant varier l'inclinaison d'une base avec leur main. Ces mêmes participants réalisaient cependant des estimations incorrectes lorsqu'ils effectuaient des reports verbaux ou qu'ils ajustaient une représentation visuelle de la pente.

1.2.5 Interactions audiovisuelles et localisation auditive

L'information que l'on reçoit du monde qui nous entoure est transmise sous différentes modalités sensorielles. Les informations venant d'un objet et captées par chacun de nos sens sont

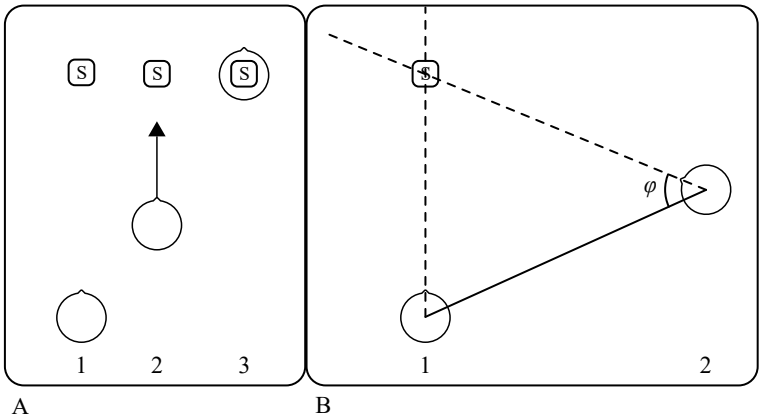


FIGURE 1.14 – Exemples de protocoles d'action dirigée par la perception. La source S joue un son lorsque le sujet, qui ne la voit pas, est dans la position 1. Dans la *marche à l'aveugle* (A), après que le son soit joué, le sujet marche dans la direction de la source (position 2) et s'arrête lorsqu'il a atteint la distance à laquelle il a perçu la source sonore (position 3). Dans la *triangulation* (B), après que le son soit joué, le sujet se déplace en position 2 puis doit viser la position à laquelle il a perçu la source sonore. La distance perçue est déduite de l'angle φ.

généralement liées et le cerveau les met en commun dans un processus de perception multisensorielle. L'audition et la vision sont deux sens particulièrement liés et la perception auditive peut, sous certaines conditions, être largement influencée par la perception visuelle.

Localisation visuelle

Comme pour l'audition, la vision se fait à l'aide de deux organes, nos yeux. Chaque scène visuelle est transcrite en une image différente par nos yeux droit et gauche, du fait de leur position différente dans l'espace. Chaque objet de la scène est alors perçu à une direction différente par chaque oeil, nommées directions oculocentriques. Les images perçues par les yeux droit et gauche sont ensuite fusionnées par le cerveau, et chaque objet est perçu dans une direction unique, la direction égocentrique, résultant des deux directions oculocentriques.

Tout comme la perception de distance auditive, la perception de distance visuelle se fait au moyen de nombreux indices impliquant ou non la comparaison des stimuli perçus par nos deux organes (indices monoculaires et binoculaires, respectivement) dont l'utilité est restreinte à une gamme de distances. Le cerveau combine ensuite les différents indices disponibles selon la distance de l'objet visuel pour affiner la sensation de profondeur qu'ils apportent. Une revue détaillée de ces indices a été réalisée par Jouhaneau (2012).

Les indices monoculaires se séparent en deux catégories. D'une part, les indices physiologiques sont donnés par l'état dans lequel le système visuel réagit au stimulus visuel et interviennent sans que le cerveau n'analyse les images captées :

- **L'accommodation** : lorsque l'oeil humain vise un objet proche ou lointain, son cristallin se déforme pour que l'image se formant sur la fovéa de sa rétine soit nette quelle que soit la distance à laquelle cet objet se trouve. La déformation est dépendante de la distance à

laquelle se trouve l'objet. Ainsi, la déformation du cristallin est maximale lorsque l'objet est au niveau du *punctum proximum* (point le plus proche de l'oeil auquel la vision reste nette). Lorsque l'objet est au *punctum remotum* (à l'infini), son image se forme sur la fovéa de la rétine lorsque l'oeil est au repos. Dans ces conditions, l'oeil n'accommode donc pas. L'accommodation peut donc être interprétée par le cerveau en tant qu'indice de distance.

- **La parallaxe** : si l'oeil d'un observateur est dirigé vers un objet distant et que cet observateur se déplace latéralement d'une position à une autre, un angle de parallaxe est créé entre les chemins joignant l'objet distant et la position de l'oeil dans les deux positions de l'observateur. Cet angle sera grand pour un objet proche et faible pour un objet lointain, constituant de ce fait un indice de distance.

Les indices cognitifs sont extraits des images captées par le système visuel et nécessitent un apprentissage de la part du cerveau pour être exploitables :

- **La taille de l'image** : un objet donne lieu à une image sur la rétine dont la taille diminue lorsque cet objet s'éloigne de l'observateur. Si l'observateur connaît la taille réelle de l'objet, il peut la comparer avec sa taille apparente pour effectuer une estimation absolue de la distance à laquelle il se trouve. Si deux objets de dimensions similaires sont perçus à des tailles apparentes différentes, celui dont la taille apparente est la plus petite sera perçu comme étant plus distant que l'autre (Delorme et Flückiger, 2003). Voir figure 1.15.
- **La hauteur dans le champ visuel** : si deux objets sont placés sur une surface plane se trouvant en-dessous de l'oeil de l'observateur, l'objet dont les points de contact avec la surface sont les plus hauts sur l'image perçue sera considéré comme étant le plus distant et ce jusqu'à une hauteur limite située sur la ligne d'horizon. L'efficacité de cet indice décroît donc avec la distance. Voir figure 1.15.
- **L'occlusion** : lorsqu'un objet est partiellement masqué par un autre objet, le système visuel détermine que l'objet masquant est plus proche que l'objet masqué. Cet indice est exclusivement relatif et ordinal puisqu'il n'informe pas l'observateur sur la distance séparant les objets. Voir figure 1.15.
- **Les ombres** : le fait que la lumière vienne généralement d'en haut peut permettre au cerveau de déterminer des informations de distance à partir de la projection des ombres des objets. À titre d'exemple, un objet peut être perçu proche d'un mur si son ombre y est projetée. Voir figure 1.15.
- **La perspective linéaire** : les lignes parallèles, telles que les lignes d'un chemin de fer, convergent vers un point de fuite avec la distance. Il est possible d'utiliser cette convergence en tant qu'indice relatif de distance. Voir figure 1.15 et 1.16.
- **La perspective de texture** (Tyler, 1975) : les détails complexes de la texture d'un objet apparaissent plus détaillés lorsque cet objet est proche que lorsqu'il est lointain. La partie lointaine d'une forêt apparaît ainsi relativement uniforme comparée à la texture détaillée que forment les feuillages proches de l'observateur, constituant un indice relatif de distance. Voir figure 1.16.

- **La perspective de diffusion** : le filtrage effectué par l'atmosphère, la pollution, le brouillard ou tout autre obstacle se trouvant entre l'objet et l'observateur rend les objets lointains flous et peu contrastés, fournissant un indice absolu de distance pour des distances élevées. Voir figure 1.16.
- **La perspective chromatique** : de même que pour la perspective de diffusion, la teinte des objets est modifiée avec la distance, donnant un aspect bleuté aux objets lointains. Voir figure 1.16.

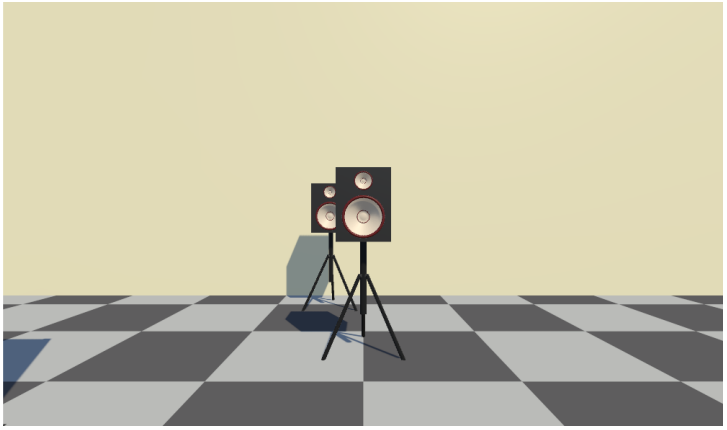


FIGURE 1.15 – Cette figure contient plusieurs indices monoculaires de distance. La taille relative des deux enceintes est différente, donnant pour impression que la plus grande est plus proche. L'indice de hauteur dans le champ visuel est représenté par la différence de hauteur dans l'image des points de contact entre les pieds des deux enceintes et le sol texturé. L'enceinte dont ces points de contact sont les plus haut est considéré comme étant la plus lointaine. De plus, l'enceinte la plus grande fait occlusion à l'autre enceinte, elle est donc située devant cette dernière. La perspective linéaire présente sur les lignes du sol texturé permettent également de déterminer la position relative des deux enceintes. Enfin, la position relative des ombres des deux enceintes renseigne sur leur éloignement relatif et se cumule au fait que l'ombre projetée par l'une des enceintes sur le mur permet de déduire qu'elle se trouve plus près de celui-ci que l'autre enceinte.

Pour certains auteurs (e.g Cutting et Vishton (1995)), certains de ces indices (tels que les perspectives linéaire et de texture) ne sont pas des indices en tant que tels, mais plutôt une combinaison d'autres indices (tels que la taille et la densité relative dans ces cas).

Les indices binoculaires sont basés sur la différence entre les images créées sur la rétine de chaque oeil. Puisque nos deux yeux sont séparés dans l'espace d'un écart inter-pupille d'environ 62 mm pour les femmes et 65 mm pour les hommes (Dodgson, 2004), les images qu'ils captent représentent la scène visuelle sous deux points de vue distincts que le cerveau fusionne pour créer une image unique comportant une sensation de profondeur (Julesz, 1971). Ces indices sont :

- **La convergence** : lorsqu'un observateur fixe un objet, l'axe optique de chacun de ses deux yeux vise vers cet objet. Ainsi, les deux axes optiques convergent vers cet objet. Plus l'objet est proche, plus la convergence est importante. Cet indice fournit une information absolue de distance.

FIGURE 1.16 – Les perspectives linéaire, chromatique, de texture et de diffusion sont clairement visibles dans cette extrait de *Stand by me*, film réalisé par Rob Reiner (1987). Les lignes du chemin de fer convergent en s'éloignant du point d'observation (perspective linéaire). La colline présente à l'arrière plan est peu contrastée (perspective de diffusion) et a une teinte bleutée (perspective chromatique), indiquant son éloignement du point d'observation. On remarque également que la verdure y est moins détaillée que celle apparaissant au premier plan (perspective de texture), de même que celle du gravier sur lequel les personnages marchent, dont le détail diminue avec l'éloignement.

· **La disparité binoculaire** (Tyler, 1975) : les images captées par chacun des deux yeux sont décalées en azimut. Ce décalage est dépendant de la distance de l'objet dans la scène visuelle (Voir figure 1.17). Ainsi, plus le décalage azimutal d'un objet entre les images captées par les deux yeux est grand, plus cet objet sera perçu proche du point d'observation. Cet indice fournit une information relative de distance par rapport au point de fixation.

La figure 1.18 représente les seuils différentiels des indices visuels de profondeur en fonction de la distance séparant les objets et l'observateur.

Performances de localisation visuelle

Les performances de localisation visuelle sont nettement supérieures aux performances de localisation auditive. L'acuité visuelle, qui désigne l'angle minimum nécessaire entre deux points pour qu'il soient distingués l'un de l'autre, peut atteindre une minute d'arc (un soixantième de degré) dans le champ de vision central (Howard, 1982). L'acuité visuelle binoculaire (ou stéréoscopique), qui désigne la disparité binoculaire à partir de laquelle il est possible de distinguer la profondeur séparant deux objets distincts, peut atteindre 2 secondes d'arc (soit une différence de profondeur de 4 mm pour des objets situés à 5 m) pour certains individus, et est inférieure à 30 secondes d'arc pour la majeure partie de la population (Coutant et Westheimer, 1993).

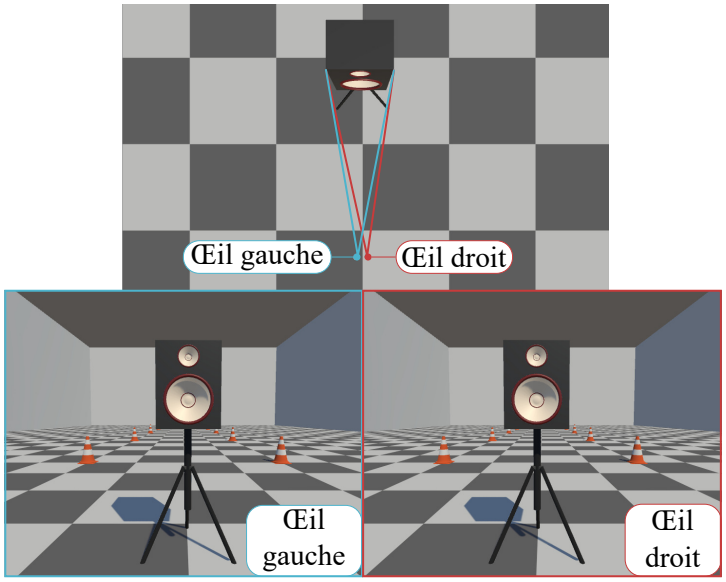


FIGURE 1.17 – Disparité binoculaire. L'image du dessus montre les chemins joignant les côtés de l'enceinte à chacun des deux yeux. Les images du bas sont les images rétiniennes de chacun des deux yeux. La position des objets lointains (murs, plots éloignés) est similaire dans les deux images, contrairement à celle de l'enceinte, celle-ci étant proche du point d'observation. La disparité entre les deux images (visible notamment en comparant la position de l'enceinte par rapport à son ombre ou aux plots dans chacune des deux images) est due à l'écartement des yeux.

Localisation en présentation audiovisuelle

Lorsqu'un stimulus visuel et un stimulus sonore sont présentés simultanément mais avec une disparité spatiale, il est commun que notre système perceptif altère les caractéristiques conflictuelles des stimuli perçus (Kohlrausch et van de Par, 2005) pour les faire correspondre l'un à l'autre. Ces interactions intermodales donnent lieu à deux phénomènes, les biais intersensoriels et l'effet ventriloque, agissant sur la localisation audiovisuelle.

Les biais intersensoriels : Ces biais représentent une déviation de la localisation auditive due à la présence d'un stimulus visuel positionné à un endroit différent (Weerts et Thurlow, 1971; Warren, 1979; Bertelson et Radeau, 1981). Ces biais sont révélés lors de tâches de localisation dans lesquelles les participants doivent indiquer la direction de provenance de sons présentés avec un stimulus visuel de direction incongruente. Le décalage de la direction perçue du stimulus sonore est alors proportionnel à la disparité spatiale entre les stimuli sonore et visuel. Ainsi, pour une séparation de 7° entre les deux stimuli, Bertelson et Radeau (1981) ont relevé un décalage de 4° (soit 57% du décalage introduit entre les stimuli) de la direction perçue du stimulus sonore. Pour une séparation de 25° entre les deux stimuli, les auteurs ont relevé un décalage de 8° (soit 32% du décalage introduit entre les stimuli) de la direction perçue.

L'hypothèse de modalité pertinente, avancée par Welch et Warren (Welch et Warren, 1980; Welch, 1999), atteste que pour une tâche donnée, un participant établit sa décision sur la base du

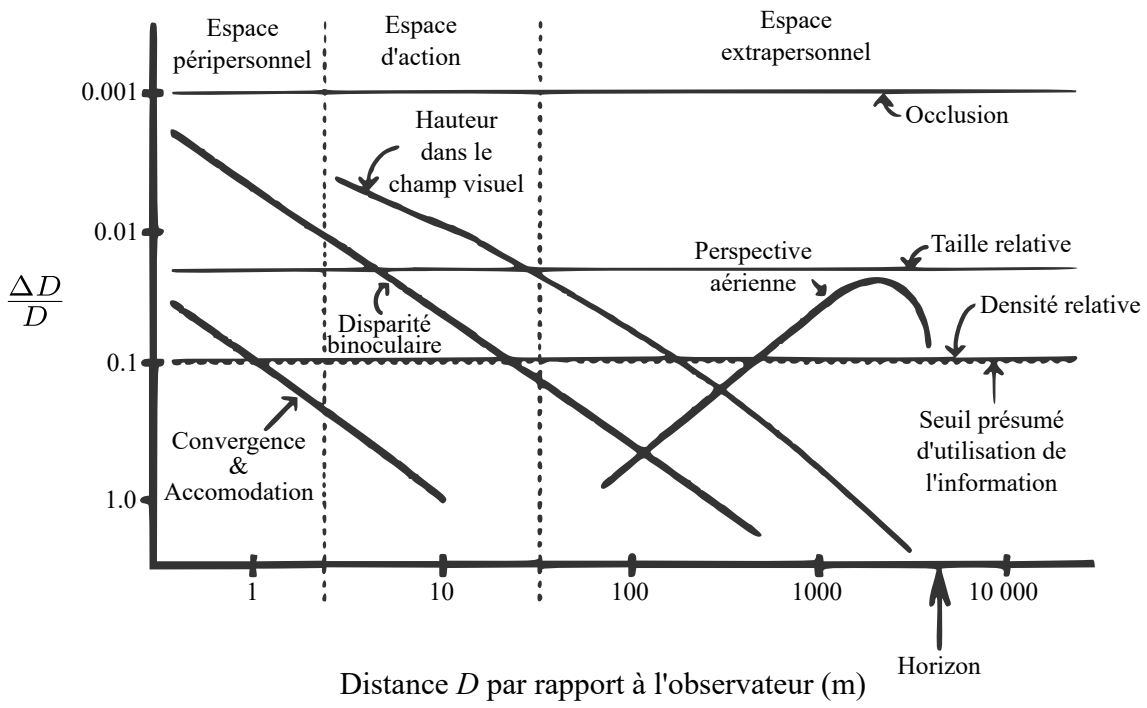


FIGURE 1.18 – Seuil différentiel de profondeur en fonction du logarithme de la distance entre l'objet et l'observateur pour différents indices de distance. D'après Cutting et Vishton (1995). Certains indices ne figurent pas dans la figure. Pour les auteurs, les perspectives linéaire et de texture sont une combinaison d'autres indices, tels que la taille et densité relatives.

sens qui fournit l'information la plus précise. Lors de tâches de localisation en présentation multimodale, l'information fournie par la vision est plus précise que celle fournie par l'audition. Elle dominerait alors cette dernière et attirerait la direction perçue du son en direction du stimulus visuel. Il est alors possible de modéliser, d'après les performances de localisation visuelle et auditive en présentation unimodale (Ernst et Banks, 2002; Alais et Burr, 2004), la position perçue d'un stimulus audiovisuel disparate. Ce modèle de « maximum de vraisemblance » détermine, pour un stimulus visuel perçu à une position m_V avec une variance σ_V^2 et un stimulus sonore perçu à une position m_A avec une variance σ_A^2, une position m_{AV} où sera localisé le stimulus audiovisuel telle que :

$$m_{AV} = \frac{\sigma_V^2}{\sigma_V^2 + \sigma_A^2} m_A + \frac{\sigma_A^2}{\sigma_V^2 + \sigma_A^2} m_V \tag{1.5}$$

La position m_{AV} perçue du stimulus audiovisuel est, d'après ce modèle, dépendante de la perception unimodale de chacun des deux sens. En conditions normales, le stimulus audiovisuel sera perçu plus proche du stimulus visuel, puisque la localisation visuelle est plus précise que la localisation auditive. Si l'on dégrade expérimentalement les indices visuels de localisation (par exemple en rendant les indices visuels flous) de sorte qu'ils deviennent moins précis que les indices auditifs, il est alors possible d'observer l'effet inverse où le stimulus audiovisuel est localisé à une position proche de celle du stimulus sonore (Alais et Burr, 2004).

L'effet ventriloque : Tandis que les biais intersensoriels déterminent les variations de localisation visuelle, auditive ou audiovisuelle en présentation multimodale, l'effet ventriloque (Thurlow et Jack, 1973; Hendrickx *et al.*, 2015) désigne le fait que les personnes perçoivent un stimulus audiovisuel unique né de la fusion entre les stimuli visuel et sonore malgré leur disparité spatiale. Cet effet est observé lors de tâches de discrimination (et non lors de tâches de localisation) dans lesquelles les participants indiquent s'ils perçoivent que les stimuli visuel et sonore proviennent ou non de la même position.

L'effet ventriloque a été étudié surtout dans le plan horizontal, dans une moindre mesure en distance, et très peu en élévation (pour une revue détaillée, voir Hendrickx (2015) et Andre (2013)). En azimut, l'effet ventriloque est de moins en moins prononcé à mesure que l'écart angulaire entre les stimuli sonore et visuel augmente, avec des seuils allant de 3° (Lewald *et al.*, 2001) à 20° (Komiyama, 1989) selon plusieurs facteurs potentiels dont la disparité temporelle entre les stimuli sonore et visuel, le réalisme du stimulus audiovisuel formé par les deux stimuli, l'utilisation de sources rélles ou virtuelles, l'attention et l'expérience du sujet (Hendrickx, 2015). En élévation, puisque les performances auditives sont moindres qu'en azimut, l'effet ventriloque est d'autant plus important. Hendrickx *et al.* (2015) ont mesuré des seuils de fusion de stimuli sonore et visuel (respectivement un enregistrement vocal et vidéo d'un locuteur) quatre fois plus importants pour des sources sonores situées dans le plan vertical que dans le plan horizontal.

L'effet ventriloque en profondeur a été décrit en tant qu'« effet de proximité » par Gardner (1968). L'auteur décrit une expérience dans laquelle cinq haut-parleurs visibles étaient placés à des distances croissantes des participants dans une salle anéchoïque. Les haut-parleurs étaient placés l'un derrière l'autre, rendant les haut-parleurs lointains masqués par le haut-parleur le plus proche. Un enregistrement vocal était diffusé sur le haut parleur le plus lointain (à environ 9 m du point d'écoute), et les participants percevaient systématiquement le son comme provenant du haut-parleur le plus proche (à environ 1 m du point d'écoute). Zahorik (2001) a mis en place un dispositif similaire à celui de Gardner (1968), à la différence que chacun des cinq haut-parleurs était visible par les participants et que l'environnement dans lequel l'expérience se déroulait était semi-réverbérant. Un enregistrement vocal était diffusé sur l'un des cinq haut-parleurs, et les participants devaient évaluer la distance de la source sonore, la moitié d'entre eux ayant les yeux bandés et l'autre moitié voyant le dispositif de haut-parleurs. Ses résultats contredisent ceux de Gardner, puisque les estimations de distances de la source sonore étaient plus précises pour les participants voyant les haut-parleurs que pour ceux ayant les yeux bandés. Cette absence d'« effet de proximité » peut être due au fait que plusieurs sources probables étaient présentées aux auditeurs. Puisque ceux-ci disposaient d'indices acoustiques de distances précis par le biais du champ réverbéré, la vision les a probablement aidés à affiner leur estimation de distance en déterminant parmi les haut-parleurs visibles lequel était le plus proche de la distance à laquelle ils percevaient le stimulus sonore. En effet, les estimations de distance sont plus précises, pour des sources sonore et visuelle congruentes, en présentation audiovisuelle qu'en présentation audio seule (Anderson et Zahorik, 2014). Lorsqu'une seule source visuelle est proposée (Mershon *et al.*, 1980; Zahorik, 2003), l'« effet de proximité » a été observé même en conditions non-anéchoïques. Mershon *et al.* (1980) ont également observé une asymétrie de cet effet selon laquelle la capture

visuelle (attraction de localisation du stimulus sonore vers celle du stimulus visuel) est plus importante lorsque la source sonore est située derrière le haut-parleur que lorsqu'elle est plus proche que ce dernier. Cette asymétrie a été également observée par Paquier *et al.* (2016) dans un environnement virtuel.

La relation entre biais intersensoriels et effet ventriloque, étudiée par Bertelson et Radeau (1981) ainsi que par Wallace *et al.* (2004), montre que les biais de localisation auditive en direction du stimulus visuel sont plus importants lorsque les participants observent une fusion des stimuli sonore et visuel, bien qu'ils soient existants même lorsque les participants perçoivent deux stimuli distincts.

1.3 Perception d'intensité sonore

Cette troisième partie décrit les différents mécanismes du système auditif qui aboutissent à la sensation de niveau sonore perçu. Cette partie évoquera dans un premier temps les grandeurs psychophysiques représentant la sensation d'intensité sonore ainsi que les différents facteurs jouant sur cette sensation, qu'ils impliquent une ou deux oreilles. La modélisation de cette sensation, basée essentiellement sur des données physiologiques et psychophysiques, sera ensuite exposée. Ces données sont pour la plupart recueillies en conditions de laboratoire dans lesquelles les stimuli sont définis et maîtrisés. La façon dont la sensation de niveau sonore perçu évolue lorsque les stimuli sonores et leur mode de restitution se rapprochent de conditions d'écoute naturelles sera ensuite évoquée.

1.3.1 Sonie et isosonie : le sone et le phone

La sonie désigne la sensation de force sonore qu'un auditeur associe à un son, laquelle pouvant aller de très faible à très forte. La sonie est définie d'après Scharf (1978) comme étant l'intensité subjective d'un son qui ne peut ainsi pas être mesurée avec des outils de mesures physiques. Des estimations doivent être effectuées par un panel d'auditeurs lors de tests perceptifs pour la quantifier et, pour un ensemble de conditions expérimentales données, la sonie est mesurée d'après la convergence des estimations réalisées par ce panel (Florentine, 2011).

L'unité communément utilisée pour désigner cette grandeur psychoacoustique est le sone. Une sonie de 1 sone correspond à la sensation de volume sonore produite par un sinus de fréquence 1 kHz dont le niveau de pression est 40 dB SPL, pour un auditeur l'entendant avec ses deux oreilles et faisant face à sa source, en champ libre. Si la sensation de volume sonore double (le son est perçu comme étant deux fois plus fort), alors la sonie double. Un son perçu comme étant deux fois plus fort qu'un sinus de 1 kHz à 40 dB SPL aura alors une sonie de 2 sones.

Il est commun de comparer la sonie de sons ayant des propriétés physiques différentes. Pour faciliter ces comparaisons, la sonie de chaque son est exprimée en terme de niveau de pression sonore d'un son de comparaison ayant une sonie similaire. Ces niveaux sont appelés niveaux d'isosonie. Si deux stimuli sont au même niveau d'isosonie, leur sonie est égale et les deux sons sont perçus aussi forts l'un que l'autre. Si ces deux sons ont des propriétés physiques différentes, il est probable qu'au même niveau d'isosonie, les niveaux physiques des deux sons soient différents. À titre d'exemple, si au même niveau de pression sonore, un son A est perçu plus fort qu'un son B et que l'on ajuste les niveaux des deux sons pour qu'ils atteignent un même niveau d'isosonie, le niveau de pression sonore du son A sera inférieur à celui du son B et les deux seront perçus aussi forts l'un que l'autre. L'unité définissant ces niveaux d'isosonie est le phone, et prend pour son de comparaison un sinus de fréquence 1 kHz entendu de façon binaurale, frontale et en champ libre. Ainsi, si le niveau d'isosonie d'un son est de 40 phones, la sensation de volume sonore qu'il engendre est la même que celle d'un sinus de 1 kHz entendu à 40 dB SPL, soit une sonie de 1 sone. Si la sonie d'un son dépend évidemment de son niveau de pression sonore, de nombreux autres facteurs, qu'ils impliquent ou non l'utilisation couplée de nos deux oreilles, peuvent la

faire varier. Le niveau d'isosonie permet alors de ramener la sonie de chaque son, quelle que soit sa complexité, à une représentation simple impliquant uniquement le niveau de pression sonore d'un son pur.

1.3.2 Principaux facteurs physiques influant sur la sonie monaurale

Si la caractéristique physique d'un son ayant le lien le plus évident avec la sonie est le niveau de pression sonore, de nombreux autres facteurs moins directs jouent également un rôle dans cette sensation. Les relations entre la sonie et chacun de ces facteurs n'est pas linéaire. Cette dépendance multiple et complexe de la sonie envers de nombreux paramètres physiques signifie donc que la sonie d'un son ne peut pas être décrite de façon directe à partir d'une caractéristique physique quelle qu'elle soit (Fletcher et Munson, 1933). Cette partie décrit les relations partagées entre différentes caractéristiques physiques données et la sonie qu'elles engendrent. Ces relations sont chacune valables pour le stimulus particulier au moyen duquel elles ont été étudiées. Par exemple, un son pur basse fréquence et un bruit à bande étroite de fréquence centrale élevée peuvent avoir une sonie différente malgré le fait que leur niveau de pression sonore est identique, puisque des caractéristiques autres que le niveau sonore (ici la largeur de bande et la fréquence centrale) entrent en jeu dans la sonie globale.

Niveau sonore

Plus le niveau sonore d'un son est élevé, plus la sonie de ce son est susceptible de l'être. La fonction psychophysique liant ces grandeurs physique et psychoacoustique est appelée fonction de sonie. L'une des premières approches pour représenter la sonie fut celle de Fechner *et al.* (1966), qui décrit la sonie sous forme d'une loi logarithmique. Sous cette forme, la quantité perçue p dépend de la grandeur du stimulus S par la relation $p = k \cdot \log(\frac{S}{S_0})$, où k est une constante et S_0 une grandeur de référence du stimulus. Si la quantité perçue p désigne la sonie et la grandeur du stimulus S la pression sonore de l'onde, la sonie pourrait alors être représentée en dB. Cependant, l'expérience montre que le dB ne parvient pas à représenter correctement les sensations relatives produites par des sons de niveaux différents. Ainsi, comme le cite Churcher (1935), le bruit d'un moteur mesuré à 80 dB au-dessus du seuil d'audition paraît bien plus que deux fois plus fort qu'un bruit identique mesuré à 40 dB au-dessus de ce même seuil. Dans la première partie du XXe siècle, Fletcher et Munson (1933) proposent une nouvelle approche du problème en utilisant le niveau sonore d'un *son de référence*, un sinus de fréquence 1 kHz, auquel comparer le niveau sonore d'un son quelconque dont la sonie doit être évaluée. La première fonction de sonie utilisant le sone fut établie par Stevens (1936). Cette fonction de sonie fut révisée au milieu du siècle dernier (Stevens, 1955, 1956), fournissant une fonction simple, sous forme de fonction puissance, liant la pression d'un sinus de fréquence 1 kHz à sa sonie :

$$S = k \cdot P^{\alpha} \tag{1.6}$$

où P est la pression du son pur en µPa, S sa sonie en sone, k et α des constantes de valeur $k = 0.01$ et $\alpha = 0.6$. Cette loi décrit correctement la sonie pour des niveaux supérieurs à 40 phones, pour

lesquels une augmentation de 10 phones correspond à un doublement de sonie. À ces niveaux, il est alors possible d'établir une relation directe $S = 2^{\frac{N-40}{10}}$ entre la sonie exprimée en sone (S) et le niveau d'isosonie du stimulus exprimée en phone (N). La valeur de l'exposant α peut connaitre de grandes variations interindividuelles (Stevens, 1955; Hellman, 1991).

Pour les niveaux inférieurs à 40 phones, la pente de la fonction de sonie devient plus importante. Différentes modélisations de la fonction de sonie à ces niveaux ont été proposées par de nombreux auteurs dans la deuxième moitié du XXe siècle, comme en témoignent les revues de Humes et Jesteadt (1991), Hellman (1997) et Buus *et al.* (1998). La figure 1.19 présente plusieurs de ces modélisations en fonction du niveau de sensation d'un son pur de fréquence 1 kHz (en dB SL, correspondant à la différence entre niveau de pression du stimulus et niveau de pression au seuil d'audition de l'auditeur). La valeur de la sonie au seuil d'audition constitue un élément majeur des discussions, certains la considérant égale à zéro et d'autres y faisant correspondre une valeur positive. Glasberg et Moore (2006) considèrent par exemple dans leur modèle (Moore *et al.*, 2016) qu'un son au seuil absolu d'audition a une sonie de 0.003 sone, correspondant à un niveau de 2.2 phones.

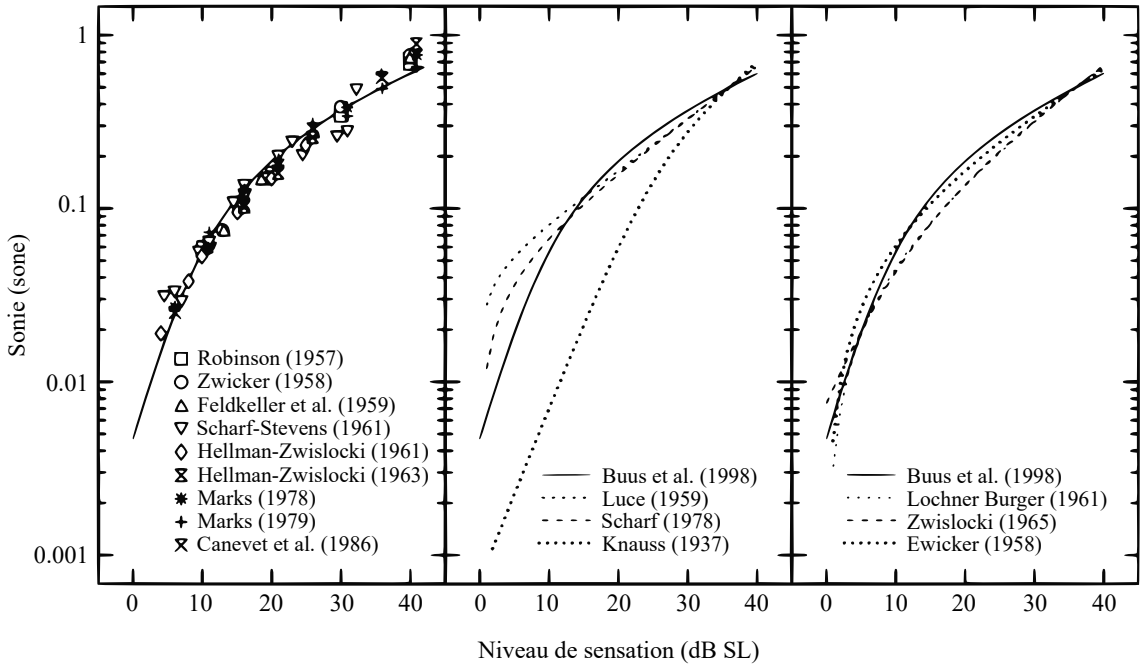


FIGURE 1.19 – Fonctions de sonie d'un sinus de fréquence 1 kHz à niveaux faibles. La fonction de Buus *et al.* (1998) apparaît dans chacun des trois panneaux à titre de référence. D'après Jesteadt et Leibold (2011).

À très fort niveau (au delà d'approximativement 100 phones), la pente de la fonction de sonie semble également augmenter (Viemeister et Bacon, 1988; Buus *et al.*, 1997). Cependant, il est difficile d'étudier la fonction de sonie à de tels niveaux en laboratoire. En effet, la présentation de sons à ces forts niveaux représente un risque de dégradation de l'audition pour les participants d'études psychoacoustiques.

Enfin, la fonction InEx (pour *Inflected Exponential law*) définie par Florentine et Epstein (2006) d'après les modifications apportées par Buus et Florentine (2002) à la fonction de sonie sous forme de loi de puissance introduit une compression sur la sonie de niveaux moyens, également observable dans Epstein et Florentine (2005) et Marozeau *et al.* (2006). La figure 1.20 représente cette fonction de sonie, accompagnée de plusieurs autres fonctions couramment utilisées dans différents standards.

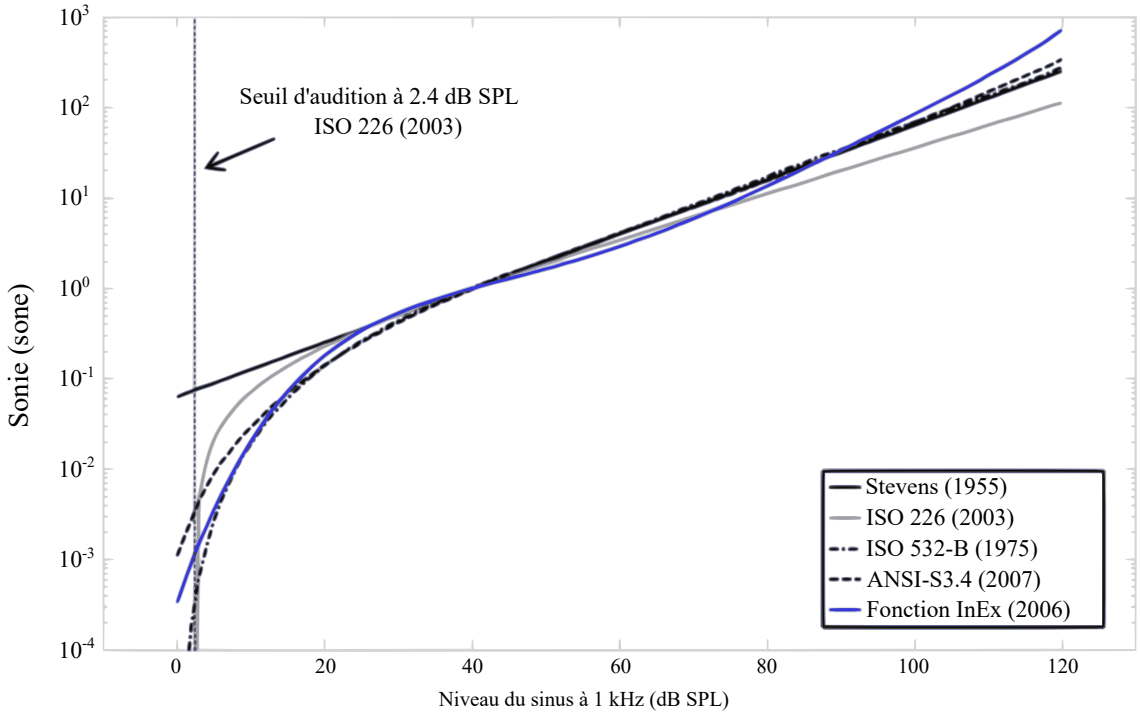


FIGURE 1.20 – Fonctions de sonie d'un sinus de fréquence 1 kHz définies dans plusieurs standards de sonie, dont Stevens (1955); ISO 532-B (1975); ISO 226 (2003); Florentine et Epstein (2006); ANSI S3.4 (2017). Chacune de ces fonctions est normalisée d'après la définition du sone et passe donc par le point d'abscisse 40 dB SPL et d'ordonnée 1 sone. D'après Vannier (2015).

Fréquence

La chaîne physiologique comporte une fonction de transfert dont le module dépend de la fréquence, notamment aux niveaux de l'oreille externe et de l'oreille moyenne, comme l'évoque la section 1.1. En conséquence, la sonie d'un son pur dépendra non seulement de son niveau sonore, mais également de sa fréquence. La méthode la plus commune pour décrire cette dépendance est d'établir des lignes d'isosonie, qui définissent le niveau sonore à fournir pour conserver une sonie constante en fonction de la fréquence d'un son pur. La figure 1.21 représente l'allure de ces courbes, chacune correspondant à un niveau d'isosonie donné en phone. Puisque le phone est défini par rapport au niveau de pression sonore d'un son pur de fréquence 1 kHz, le niveau de pression correspondant à la courbe de niveau d'isosonie de n phones est n dB SPL au point d'abscisse 1 kHz. L'allure globale de ces courbes est largement expliquée par les effets joints de

l'oreille externe et moyenne. Par exemple, les lignes isosoniques présentent un creux aux fréquences allant de 2 kHz à 5 kHz, que l'on retrouve sur la fonction de transfert de l'oreille externe (voir figure 1.2) et qui témoigne d'une grande sensibilité à ces fréquences. Les premières courbes isosoniques étaient basées sur des données obtenues au casque, notamment par Kingsbury (1927) et Fletcher et Munson (1933), dans des expériences où les participants comparaient la sonie d'un son pur d'une fréquence donnée avec celle d'un sinus de fréquence 1 kHz diffusé à un niveau de référence. Plus tard, Robinson et Dadson (1956) ont établi de nouvelles courbes selon une procédure similaire, mais en restituant les sons sur enceintes en champ libre. Cette technique a été la plus utilisée dans les études ayant alimenté les standards actuels (voir Suzuki et Takeshima (2004) pour une revue des études ayant été utilisées par la norme ISO 226 (2003)). Là aussi, plusieurs normes différentes sont en vigueur puisque la norme ANSI S3.4 (2017) propose des courbes isosoniques légèrement différentes, basées sur le modèle décrit par Glasberg et Moore (2006). Suzuki et Takeshima (2004) proposent une analyse des différences entre ces deux modélisations de courbes isosoniques.

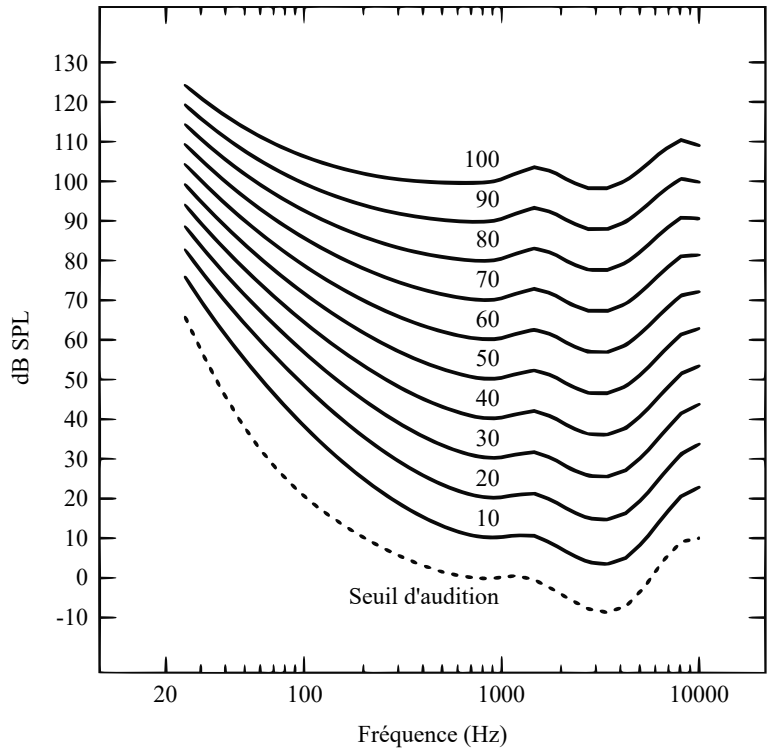


FIGURE 1.21 – Courbes isosoniques telles que définies par la norme ISO 226 (2003). Chaque ligne correspond au niveau à fournir, pour une fréquence donnée, pour obtenir une sonie égale dont la valeur est notée en phone.

Durée

L'intégration temporelle de sonie désigne l'augmentation de la sonie avec la durée pour des sons de courte durée. Les études s'intéressant à cet effet l'ont mesuré à l'aide de différentes

méthodes. Miller (1948) a par exemple utilisé une méthode d'ajustement, dans laquelle les participants ajustaient le niveau d'un son afin qu'il leur paraisse aussi fort qu'un son de durée différente. Munson (1947) et Florentine *et al.* (1996) ont utilisé des procédures adaptatives, dans lesquelles deux sons (un *son de référence* et un *son test*) de durées différentes étaient présentés aux participants, qui devaient indiquer lequel des deux sons leur paraissait être le plus fort. Le niveau du *son test* était alors modifié en fonction de leur réponse (s'il avait été jugé plus fort, son niveau était réduit et vice versa) afin de converger vers un niveau, appelé point d'égalité subjective, auquel le *son test* était perçu aussi fort que le *son de référence*. Enfin, Epstein et Florentine (2006) ont utilisé des méthodes d'estimation directe de sonie dans lesquelles les participants assignaient un nombre, dont l'amplitude devait correspondre à la sonie évoquée, à des sons de durées différentes. Dans une variante de cette méthode (Epstein et Florentine, 2005), qualifiée de méthode cross-modale, les participants n'utilisaient pas un nombre pour quantifier la sonie mais devaient couper une ficelle à une longueur correspondant à cette sensation.

Les résultats de ces études reportent que des sons de très courtes durée (de l'ordre de 5 ms) sont perçus moins forts que des sons de durée plus longue (par exemple 200 ms). Cependant, certaines observations semblent à première vue contradictoires entre les différentes études. La différence de niveau requise pour atteindre une sonie égale entre un stimulus très bref et un stimulus de durée plus longue est plus importante pour les sons diffusés à niveaux moyens que pour les sons diffusés à niveaux faibles et forts. Pourtant, l'hypothèse de Florentine *et al.* (1996) et les résultats d'Epstein et Florentine (2005, 2006) montrent que l'augmentation de sonie provoquée par une augmentation de la durée du stimulus est constante selon le niveau du stimulus. Ces différences sont expliquées par le fait que les différentes méthodes évaluent deux grandeurs différentes :

- L'augmentation de sonie ΔS provoquée par l'augmentation de la durée du son est obtenue lors de méthodes d'estimation directe. Pour un niveau de pression sonore donné, les estimations de sonie obtenues pour des stimuli de durées différentes sont comparées et le rapport entre ces estimations donne la grandeur ΔS.
- La différence de niveau ΔL nécessaire pour que la sonie d'un son d'une durée donnée soit égale à celle d'un son de durée différente est obtenue par des méthodes d'ajustement.

La dépendance de ΔL et l'indépendance de ΔS vis-à-vis du niveau est alors expliquée par l'allure des fonctions de sonie des stimuli étudiés (voir figure 1.22).

L'intégration de sonie comporte une durée critique au-delà de laquelle la sonie n'évolue plus avec la durée. Cette durée critique a été observée à des valeurs très variables selon les études, allant de 15 ms (observée par Small *et al.* (1962) sur des salves de bruit à moyens niveaux) à 500 ms (observée par Ekman *et al.* (1966) sur des sons purs de fréquence 1 kHz à faibles niveaux). Ces variations peuvent d'une part provenir des différences entre les stimuli utilisés, mais pourraient également venir de biais systématiques liés à la difficulté qu'ont les participants à comparer des stimuli de durées différentes (Epstein et Florentine, 2006). Les courbes d'isosonie présentées dans la partie précédente sont établies pour des sons de durée supérieure à cette durée critique (de l'ordre de la seconde).

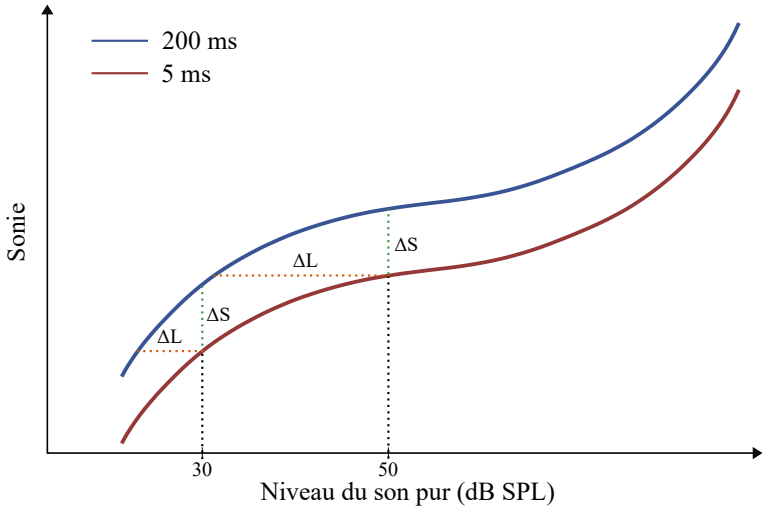


FIGURE 1.22 – Représentation schématique de fonctions de sonie faisant apparaître l'intégration temporelle de sonie. L'augmentation de sonie ΔS provoquée par l'augmentation de la durée du son pur est constante quel que soit le niveau du son. En revanche, la différence de niveau ΔL nécessaire pour atteindre une sonie égale entre les deux stimuli est plus importante à moyens niveaux du fait de la non-linéarité de la fonction de sonie.

Lorsqu'un son est présenté pendant une longue durée (de l'ordre de la minute), un phénomène inverse peut apparaître donnant lieu à une chute de la sonie. Ce phénomène d'adaptation intervient principalement à faible niveau (Hellman *et al.*, 1997) et en hautes fréquences (Miśkiewicz *et al.*, 1993). L'origine de l'adaptation de sonie est encore inconnue. Les mesures d'adaptation du nerf auditif montrent que ce dernier s'adapte à des niveaux élevés, à l'encontre du phénomène d'adaptation en sonie. Voir Scharf (1983) pour une revue complète.

Largeur de bande

Bien que les sons purs permettent de mettre en évidence, en laboratoire, des relations simples entre les paramètres évoqués ci-dessus et la sonie, ce sont des sons auxquels le système auditif est très peu confronté dans la vie réelle. Les sons présents dans notre environnement quotidien ont un spectre plus large qui comporte de nombreuses composantes fréquentielles dont l'étendue définit la largeur de bande.

Lorsque l'onde sonore est transmise à l'oreille interne, elle fait vibrer la membrane basilaire en des zones dépendant de son contenu fréquentiel (voir la tonotopie cochléaire définie en section 1.1). La cochlée fonctionne sur le principe d'une batterie de filtres passe-bandes qui sélectionnent des bandes du spectre du son pour les analyser. Les fréquences de coupure de ces filtres cochléaires (plus communément appelés bandes critiques) ne sont pas figées mais dépendent de la composition fréquentielle du son en présence. Lors du processus de sommation de sonie, une sonie dite *spécifique* est calculée pour chaque bande critique (Zwicker, 1960) puis la sonie totale correspond à la somme des sonies spécifiques (Zwicker et Scharf, 1965; Moore *et al.*, 1997; Fastl et Zwicker, 2007). Zwicker *et al.* (1957) ont réalisé une étude visant à déterminer la largeur de ces

bandes critiques. Dans cette étude, des participants avaient pour tâche de faire correspondre la sonie de sons complexes de diverses largeurs de bande et la sonie d'un son pur. Lorsque la largeur de bande du son complexe est inférieure à une bande critique, la sonie totale est égale à la sonie spécifique de cette seule bande critique. En résultat, la sonie est indépendante de la largeur de bande. En revanche, lorsque la largeur de bande du stimulus est supérieure à une bande critique (le signal « déborde » sur les bandes critiques voisines), la sonie globale devient la somme des sonies spécifiques de la bande critique centrale et des bandes critiques voisines. Cela se traduit par une augmentation de la sonie lorsque les sons sont présentés à des niveaux relativement élevés. De nombreuses études plus récentes ont confirmé l'augmentation de sonie observée par Zwicker *et al.* (1957) (e.g. Florentine *et al.* (1978); Leibold *et al.* (2007)). À niveaux faibles (<20 phones), la sonie est quasiment indépendante de la largeur de bande puisque l'énergie débordant dans les bandes critiques adjacentes est susceptible d'être infraliminaire.

Historiquement, les bandes critiques étaient représentées en bark, unité proposée par Zwicker (1961) définissant 24 bandes critiques juxtaposées représentant une longueur constante d'environ 1.3 mm sur la cochlée, couvrant la majorité du spectre audible. Plus récemment, elles ont été représentées sous forme d'ERB [1] par Moore et Glasberg (1983), dont la largeur est définie par Glasberg et Moore (1990) :

$$\mathrm{ERB} = 24.7(4.37 \cdot 10^{-3} \cdot f + 1) \tag{1.7}$$

où f est la fréquence centrale de l'ERB en Hz. Une ERB correspond approximativement à une longueur constante d'environ 0.9 mm sur la cochlée (Moore, 2012a).

La largeur de la zone d'excitation de la membrane basilaire dépend également du niveau auquel le son est présenté. Lorsque la membrane basilaire vibre, elle entraîne une excitation définie par Moore et Glasberg (1983) comme étant la sortie des filtres auditifs en fonction de leur fréquence centrale. À niveau moyen, les filtres auditifs semblent être relativement symétriques sur une échelle fréquentielle linéaire. Cette excitation, représentée sous forme de « pattern d'excitation », peut alors être dérivée à partir de la sortie de chacun des filtres auditifs (voir figure 1.23). Lorsque le niveau augmente, les filtres auditifs deviennent asymétriques. Tandis que la pente de ces filtres reste relativement insensible au niveau du côté des hautes fréquences, la pente du côté des basses fréquences devient plus douce lorsque le niveau augmente. Ainsi, plus un son pur est fort, plus il excitera de filtres de fréquences centrales supérieures. En résultat, le pattern d'excitation s'étendra sur les hautes fréquences d'autant plus que le niveau du stimulus est fort, comme le montre la figure 1.24.

1. Pour *Equivalent Rectangular Bandwidth*. L'ERB représentant un filtre F correspond à la largeur de bande d'un filtre rectangulaire dont la transmission est égale à la transmission maximale de F et qui transmet la même énergie de bruit blanc que F (Moore, 2012b).

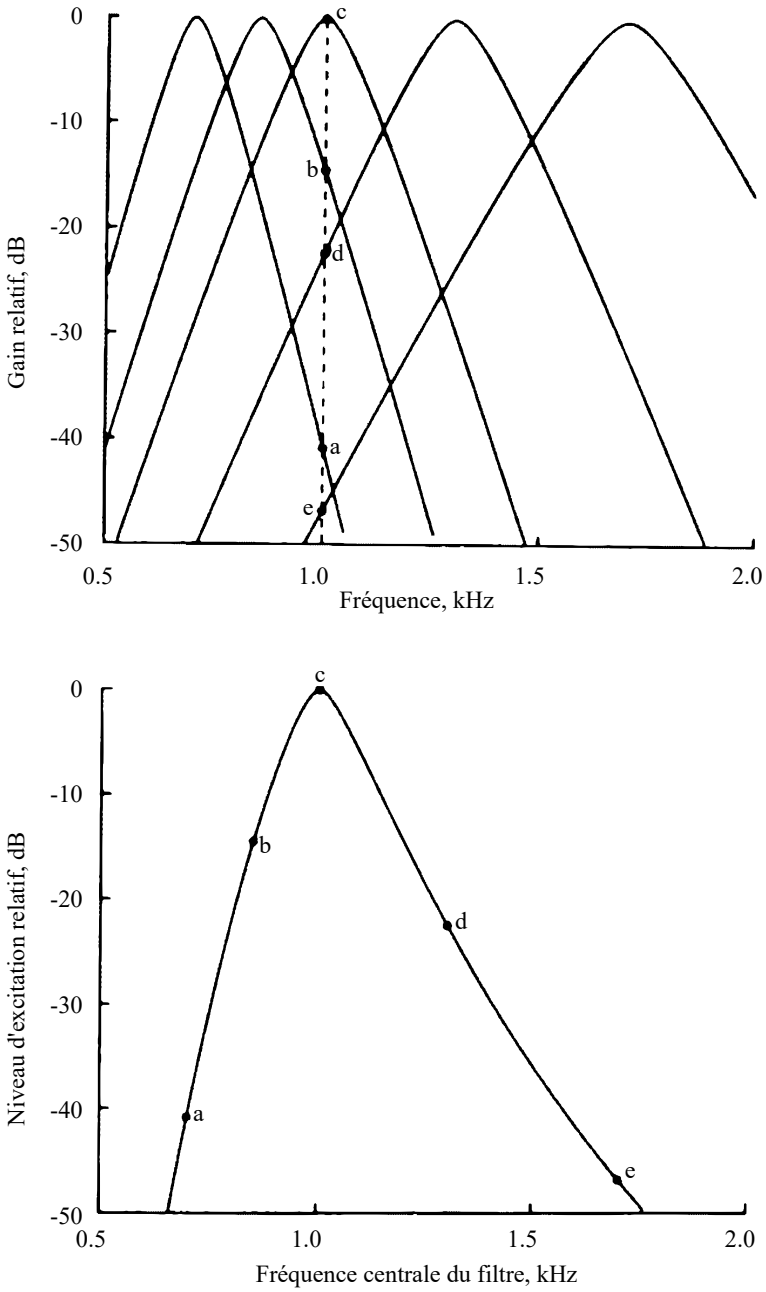


FIGURE 1.23 – Pattern d'excitation provoqué par un son pur de fréquence 1 kHz, déterminé à partir de la sortie du filtre auditif centré sur cette fréquence (déterminée par le point c) et des filtres voisins (déterminées par les points a, b, d et e). La partie du haut montre cinq filtres auditifs distincts centrés sur des fréquences centrales différentes. La partie du bas montre le pattern d'excitation ainsi créé. D'après Moore et Glasberg (1983).

1.3.3 Sonie de sons non-stationnaires

La partie précédente a évoqué des relations entre certains paramètres d'un son et la sonie. Ces relations ont été étudiées sur des sons stationnaires (qui n'évoluent pas dans le temps), pour lesquels il est simple d'isoler et de manipuler un paramètre défini. Or, la plupart des sons auxquels

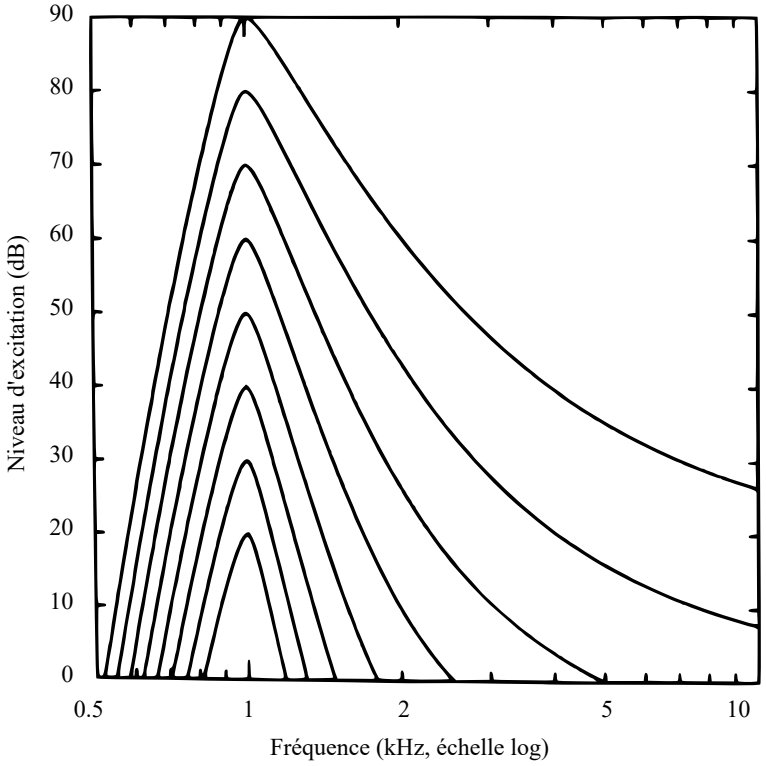


FIGURE 1.24 – Patterns d'excitation de sons purs de fréquence 1 kHz de niveaux allant de 20 dB SPL à 90 dB SPL par pas de 10 dB. L'excitation est exprimée en niveau d'excitation ($L_E = 10\log_{10}\frac{E}{E_0}$) relatif à E_0, l'excitation produite par un sinus de fréquence 1 kHz à 0 dB SPL. D'après Moore (2012a).

nous sommes confrontés dans notre environnement varient au cours du temps : ils sont non-stationnaires. Certains paramètres pouvant affecter la sonie peuvent alors varier avec le temps, donnant lieu à une évolution de la sonie au cours du temps. Une modulation d'amplitude peut notamment résulter en une fluctuation de la sonie si la fréquence de modulation est suffisamment faible. Pour des modulations d'amplitude dont la fréquence est supérieure à 20 Hz, la sonie devient constante et les fluctuations sont perçues comme une composante perturbante et déplaisante du son (Terhardt, 1974). Namba (1974) a comparé la sonie de sons stationnaires à celle de sons de même amplitude maximale et de même durée, mais dont l'amplitude augmentait progressivement dans le temps. Ses résultats indiquent que la sonie des sons non-stationnaires était inférieure à celle des sons stationnaires. Cette observation est expliquée par le fait que, pour ces stimuli, l'énergie totale des sons non-stationnaires était inférieure à celle des sons stationnaires. Lorsque la durée du son stationnaire était manipulée de sorte que les énergies totales des deux stimuli étaient similaires, la différence de sonie précédemment observée n'était plus systématique. Les résultats de Kuwano *et al.* (2007) montrent que le niveau d'énergie total semble être un bon indicateur de la sonie globale de sons de courte durée, indépendamment de la façon dont leur amplitude varie au cours du temps. Pour des sons de plus longue durée, la manière dont l'amplitude varie au cours du temps peut affecter la sensation globale de sonie. Susini et McAdams (2008) ont révélé que le son produit par l'accélération d'une voiture a une sonie plus élevée que celui produit par un son

de décélération. Les auteurs expliquent ce résultat par une influence du niveau d'arrivée d'un son d'amplitude croissante sur la sonie. L'évolution instantanée de la sonie de sons non-stationnaires dépend également du sens de variation de l'amplitude. La sonie spécifique instantanée peut augmenter relativement rapidement lorsque l'amplitude d'un signal sonore croît, mais prend plus de temps à diminuer lorsque l'amplitude de ce signal décroît. Cette différence peut correspondre, d'après Moore *et al.* (2016), à la persistance de l'activité neuronale à certains stades du système auditif. Pour une revue détaillée sur la sonie de sons non-stationnaires, voir Kuwano et Namba (2011).

1.3.4 Sonie binaurale

Le système auditif comporte deux oreilles et donc deux cochlées sur lesquelles les sons qui nous parviennent sont captés. Les paramètres physiques évoqués ci-dessus vont ainsi affecter la sonie monaurale perçue par chacune de nos deux oreilles. La sonie globale est ensuite déterminée d'après la mise en commun de ces deux sonies monaurales dans un procédé dit de sommation de sonie binaurale.

Gain binaural et rapport de sonie binaural-à-monaural

Il est communément admis qu'un son est globalement perçu plus fort lorsqu'il est présenté aux deux oreilles plutôt que sur une seule des deux. En revanche, la proportion dans laquelle la sonie augmente lorsque l'on passe d'une écoute monaurale à une écoute binaurale n'est pas systématique et dépend de certains paramètres. L'écoute binaurale peut se séparer en deux catégories, à savoir l'écoute diotique (dans laquelle les sons perçus par les deux oreilles sont strictement identiques) et l'écoute dichotique (dans laquelle ces sons présentent des différences interaurales). Il est alors possible de décrire la sommation de sonie à l'aide de deux propriétés :

- Le gain binaural, noté g, représente le gain à appliquer à un son présenté de façon monaurale pour que sa sonie soit la même qu'un son identique présenté de façon binaurale. Il est mesuré à l'aide de méthodes de comparaison telles que l'ajustement ou les procédures adaptatives.
- Le rapport de sonie binaural-à-monaural, noté $\frac{S_\mathrm{b}}{S_\mathrm{m}}$, correspond au rapport entre la sonie d'un son présenté de façon binaurale et la sonie de ce même son présenté de façon monaurale. Il est déterminé à partir d'estimations directes de sonie de sons binauraux et monauraux.

La relation entre gain binaural et rapport de sonie binaural-à-monaural dépend de la pente de la fonction de sonie (comparable par analogie aux ΔL et ΔS, respectivement, de la figure 1.22 représentant les fonctions de sonie pour des sons purs de durées différentes).

Les valeurs de g varient selon les études (e.g. de 6 dB à 8 dB pour Scharf (1969), de 9 dB à 17 dB pour Mulligan *et al.* (1985)). Ces différences semblent s'expliquer par les différences de stimuli et de protocoles utilisés. Le gain binaural semble toutefois être dépendant du niveau (Whilby *et al.*, 2006). Il est relativement faible pour les niveaux faibles, augmente progressivement

pour atteindre son maximum entre 50 dB SPL et 60 dB SPL, puis chute de nouveau à forts niveaux.

Un rapport binaural-à-monaural de 2 décrit une sommation de sonie « parfaite », pour laquelle un son perçu avec les deux oreilles est perçu deux fois plus fort qu'un son perçu avec une seule oreille. Cette valeur de rapport de sonie binaural-à-monaural a été observée par Marks (1978), lors d'une étude dans laquelle les participants réalisaient des estimations libres de sonie pour des sons purs délivrés de façon monaurale, binaurale diotique ou dichotique (avec ILD). Les estimations de sonie des sons binauraux correspondaient alors à la somme des deux estimations de sonie monaurales. Il est cependant plus commun d'observer des sommations de sonie « moins que parfaites », avec des rapports de sonie binaural-à-monaural de 1.7 (Scharf et Fishken, 1970), 1.5 (Algom *et al.*, 1989; Zwicker et Zwicker, 1991) voire 1.3 (Marozeau *et al.*, 2006).

Le rapport de sonie binaural-à-monaural a été premièrement étudié par Reynolds et Stevens (1960), qui ont déterminé les sonies monaurale $S_\mathrm{m} = k_\mathrm{m} \cdot P_\mathrm{m}^{\alpha_\mathrm{m}}$ et binaurale $S_\mathrm{b} = k_\mathrm{b} \cdot P_\mathrm{b}^{\alpha_\mathrm{b}}$ de différents sons. Les auteurs ont trouvé des différences dans les exposants α_m et α_b, désignant un rapport $\frac{S_\mathrm{b}}{S_\mathrm{m}}$ dépendant du niveau. Ces différences étaient toutefois relativement faibles, les auteurs reportant les valeurs $\alpha_\mathrm{m} = 0.54$ et $\alpha_\mathrm{b} = 0.6$. D'autres études plus récentes présentent une sommation de sonie parfaite (Hellman et Zwislocki, 1963; Marks, 1978) ou non (Zwicker et Zwicker, 1991; Marozeau *et al.*, 2006) en constituant l'hypothèse d'un rapport binaural-à-monaural indépendant du niveau (*binaural equal-loudness-ratio hypothesis*). Les résultats de Marozeau *et al.* (2006) renforcent notamment le fait de concevoir la fonction de sonie sous une forme autre qu'une fonction puissance (telles que les fonctions plus récentes présentées figure 1.20), expliquant la dépendance du gain binaural au niveau observée par Whilby *et al.* (2006), comparable par analogie aux observations faites sur l'effet de la durée des stimuli (voir figure 1.22).

La plupart des études s'intéressant à la sommation de sonie binaurale ont employé des restitutions au casque permettant de contrôler précisément les sons envoyés à chacune des deux oreilles et de créer des situations purement monaurales, diotiques ou dichotiques (e.g. Zwicker et Zwicker (1991); Marozeau *et al.* (2006)). Cependant, des études récentes observent des sommations de sonie différentes selon le mode de restitution des sons (voir section 1.3.6).

La sommation de sonie binaurale dépend de peu de paramètres lorsque l'écoute binaurale est diotique. Elle a été observée comme étant indépendante de la durée du stimulus (Whilby *et al.*, 2006), du niveau (Marozeau *et al.*, 2006) et de la fréquence (Marks, 1978) pour des sons purs. Le rapport de sonie binaural-à-monaural semble cependant varier lorsque le spectre des stimuli devient plus étendu (Algom *et al.*, 1989), sans pour autant dépendre de la zone spectrale contenant l'énergie (Zwicker et Zwicker, 1991). Le nombre de variables mises en jeu lors de telles expériences rend toutefois difficiles les conclusions générales, chaque différence entre diverses observations pouvant être due à une combinaison particulière de nombreux paramètres. La sommation de sonie binaurale semble en revanche dépendre considérablement des différences interaurales dans le cas d'une écoute dichotique. Ces dépendances seront présentées dans la section 1.4.

Moore et Glasberg (2007) supposent que l'inhibition controlatérale, définissant le fait qu'un son présenté sur une oreille peut réduire la sonie d'un son présent sur l'autre oreille (Scharf, 1969; Gigerenzer et Strube, 1983; Glasberg et Moore, 2010), pourrait être la raison pour laquelle la sommation de sonie binaurale est généralement moins que parfaite. Selon leur modélisation, la sonie monaurale du son perçu par chaque oreille est réduite par le son perçu à l'autre oreille dans le processus de sommation de sonie, rendant la sonie binaurale inférieure à la somme des sonies monaurales avant inhibition. Cette inhibition controlatérale a une fréquence d'accord large. Ainsi, un son d'une fréquence donnée présenté à une oreille pourra inhiber les sons perçus par l'autre oreille sur une large bande de fréquences (Scharf, 1969).

1.3.5 Modéliser la sonie

Dans certaines situations, il peut être intéressant d'évaluer la sonie évoquée par un son donné. Cette tâche n'est pas simple, puisque la sonie résulte de processus complexes dont les mécanismes ne sont pas entièrement compris à ce jour. De plus, la sonie ne peut pas être directement mesurée au moyen d'outils de mesures physiques. Il n'existe d'ailleurs pas de valeur « vraie » de la sonie d'un son donné, puisque la sonie mesurée peut varier d'un auditeur à l'autre, ou même pour un même auditeur lors de deux présentations différentes d'un même son. Une solution est d'évaluer la sonie moyenne évoquée par ce son en la restituant à un panel d'auditeurs puis en leur demandant de reporter la sensation évoquée. C'est un processus relativement lourd, qu'il faudrait répéter dès lors que le son à évaluer ou que les conditions dans lesquelles il est restitué ou perçu changent. La deuxième solution est de modéliser la sonie d'un son en fonction de différents paramètres physiques quantifiables. En se basant sur des données récoltées lors d'expériences psychoacoustiques, il est alors possible d'analyser le son, tel qu'il est présenté à l'auditeur, et d'en déduire une estimation globale de la sonie qu'il évoque. Les modélisations ainsi réalisées peuvent être plus ou moins complexes. Les fonctions de sonie, telles que celle établie par Stevens (1955, 1956), peuvent déjà être considérées comme des modèles de sonie, au sens où elles font correspondre une sonie à une mesure physique des paramètres d'un son. Elles restent cependant fondamentalement simples puisque, pour un son pur d'une fréquence donnée, elles ne prennent en compte que le niveau sonore dans leur estimation et omettent de ce fait de nombreux autres paramètres ayant une influence sur la sonie (voir section 1.3.2). D'autres modèles, plus exhaustifs, ont pour objectif d'analyser en détail les stimuli à évaluer et de prendre en compte la majorité des phénomènes psychoacoustiques connus jouant un rôle dans la sonie pour réaliser des estimations les plus fidèles possibles aux observations expérimentales.

Ces modèles sont structurés sous forme de blocs représentant chacun un maillon de la chaîne menant un son à la sonie. À titre d'exemple, la figure 1.25 présente le schéma de fonctionnement du modèle TVL (pour *Time Varying Loudness*) de Moore *et al.* (2016). Ce modèle, forme évoluée de leur premier modèle (Moore et Glasberg, 1996), permet d'estimer la sonie binaurale de sons non-stationnaires en présentations diotique et dichotique. Concrètement, le modèle se présente sous la forme d'une boîte noire prenant en entrée un signal audio ainsi que certains paramètres supplémentaires, indiquant notamment le niveau physique de restitution du signal. En sortie, le

modèle fournit une estimation de la sonie à court terme et à long terme. La première correspond à l'évolution continue de la sonie, et pourrait décrire la sonie provoquée par une syllabe ou un court mot. La seconde permet d'estimer efficacement la sonie globale perçue d'un son. Ces deux grandeurs peuvent différer pour des stimuli non-stationnaires. Dans l'étude de Susini *et al.* (2007) par exemple, les participants évaluaient de deux façons différentes la sonie de sons dont le niveau variait dans le temps. Les évaluations étaient soit continues (les participants ajustaient un curseur en temps-réel selon la sonie évoquée par le stimulus), soit globales (les participants reportaient la sonie globale provoquée par le stimulus). Les résultats des estimations continues peuvent être assimilés à la sonie à court-terme et ceux des estimations globales à la sonie à long-terme. Cette étude révèle une asymétrie des estimations de sonie globale selon que le niveau des stimuli était croissant ou décroissant, ce qui n'était pas le cas pour les estimations de sonie continue.

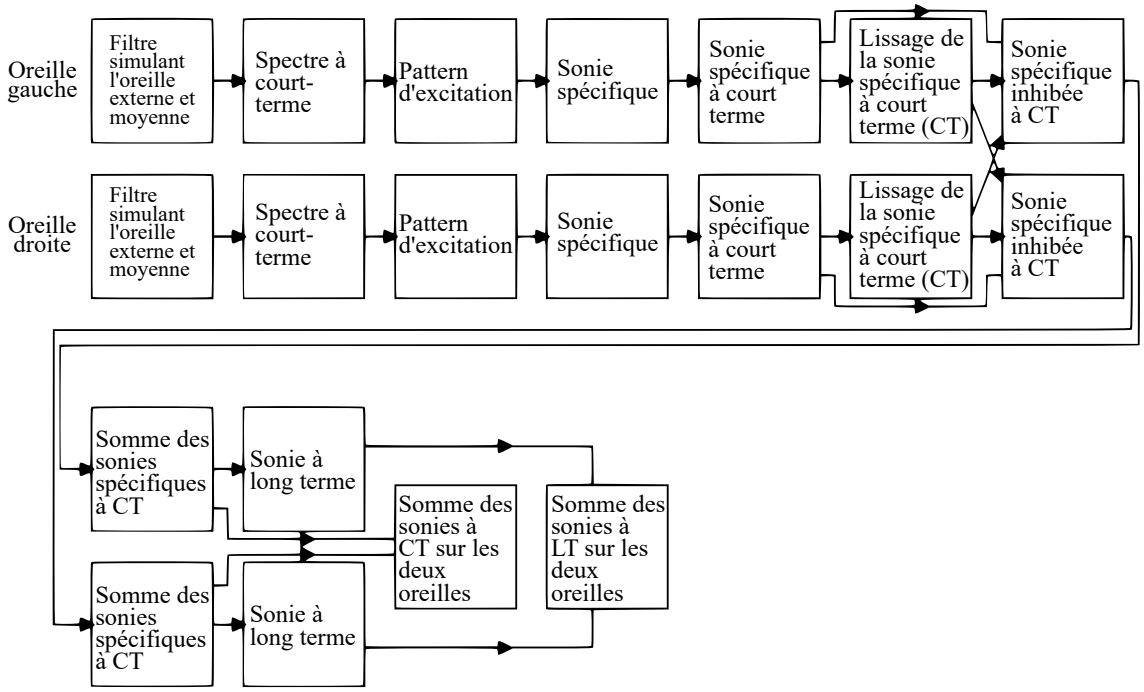


FIGURE 1.25 – Schéma de fonctionnement du modèle de sonie binaural TVL pour des sons non-stationnaires incorporant l'inhibition controlatérale.

Le fonctionnement général de chacun des blocs représentés dans la figure 1.25, correspondant au fonctionnement d'un maillon de la chaîne auditive aboutissant à la sonie, est expliqué ci-dessous :

· **Filtre simulant l'oreille externe et moyenne** : Il faut indiquer au modèle quelles modifications spectrales le son va subir avant d'atteindre les tympans de l'auditeur. À titre d'exemple, si le son que l'on injecte au modèle est une prise de son binaurale effectuée avec une tête artificielle dont les capsules sont en fond de CAE, il a déjà été filtré par l'oreille externe du mannequin et le modèle ne doit pas réitérer cette étape. Le choix se porte entre trois filtres. Le premier, le filtre champ libre, correspond au filtrage intervenant entre un son mesuré au centre de la tête de l'auditeur en son absence et ce même son

lorsqu'il atteint l'oreille interne de l'auditeur, en champ libre et en incidence frontale. Il correspond aux actions combinées des HRTF et de l'oreille moyenne. Le second filtre, le filtre champ diffus, correspond à une moyenne de filtres champ libre pour de nombreuses directions d'incidence. Enfin, le filtre « plat au tympan » ne prend en compte que l'action de l'oreille moyenne.

- **Spectre à court terme** : Six transformées de Fourier à fenêtre glissante, dont les fenêtres temporelles sont de tailles différentes, sont effectuées simultanément chaque milliseconde pour calculer le spectre du son avec une résolution fréquentielle suffisante en basses fréquences (fenêtres longues) et une résolution temporelle suffisante en hautes fréquences (fenêtres courtes).
- **Pattern d'excitation** : Un « pattern d'excitation », correspondant à la sortie des filtres auditifs, est calculé d'après le spectre à court terme chaque milliseconde d'après la méthode décrite par Glasberg et Moore (1990).
- **Sonie spécifique** : La sonie spécifique instantanée N' est calculée chaque milliseconde pour chaque ERB, selon une fenêtre glissante de largeur d'un quart d'ERB (introduisant un recouvrement, afin d'augmenter la précision du calcul). N' est proportionnelle à l'effet interne évoqué par l'excitation et introduit une non-linéarité compressive (Robles et Ruggero, 2001) supposée provenir du mécanisme actif d'amplification de la cochlée (Yates, 1995) et de la transduction entre vibration de la membrane basilaire et activité neuronale (Yates, 1990). Trois équations différentes liant la sonie spécifique N' à l'excitation du signal E_{SIG} sont définies, chacune correspondant à une différente gamme de niveaux :

$$N' = C \cdot [(G \cdot E_{\mathrm{SIG}} + A)^{\alpha} - A^{\alpha}] \tag{1.8}$$

pour les excitations moyennes ($E_{\mathrm{THRQ}} < E_{\mathrm{SIG}} < 10^{10}$, avec E_{THRQ} l'excitation produite par un signal au seuil d'audition THRQ, pour *THReshold in Quiet*), où G est un terme représentant la dépendance en fréquence du mécanisme d'amplification actif de la cochlée, où $\alpha < 1$ et où A, dont les valeurs dépendent de G, définissent la pente de la fonction de sonie. C est une constante d'ajustement permettant d'obtenir une sonie de 1 sone pour un sinus de 1 kHz à 40 dB SPL. Aux niveaux d'excitation infraliminaires ($E_{\mathrm{SIG}} < E_{\mathrm{THRQ}}$), où l'activité neuronale évoquée par le signal devient plus faible que le bruit neuronal inhérent du système auditif (Moore *et al.*, 1997), l'équation devient :

$$N' = C \cdot (\frac{2E_{\mathrm{SIG}}}{E_{\mathrm{SIG}} + E_{\mathrm{THRQ}}})^{1.5} \cdot [(G \cdot E_{SIG} + A)^{\alpha} - A^{\alpha}] \tag{1.9}$$

N' diminue alors plus rapidement que le prédirait l'équation 1.8. Enfin, lorsque les sons sont présentés à très fort niveau ($E_{\mathrm{SIG}} > 10^{10}$), le mécanisme actif d'amplification de la cochlée n'a plus d'effet et N' devient indépendant de G. De plus, la pente de la fonction de transfert de la membrane basilaire augmente. L'équation devient alors :

$$N' = C \cdot (\frac{E_{\mathrm{SIG}}}{1.04 \cdot 10^{6}})^{0.5} \tag{1.10}$$

Le modèle incorpore également une méthode de traitement permettant de déterminer la

sonie spécifique d'un signal présenté dans du bruit avec un raisonnement similaire, où un terme additionnel est pris en compte, E_{THRN}, l'excitation produite par le signal à son seuil de masquage (THRN, pour *THReshold in Noise*).

- **Sonie spécifique à court terme** : Les sonies spécifiques calculées jusqu'à présent sont des sonies dites instantanées, qui ne sont pas perceptibles. Ces sonies sont ensuite lissées dans le temps pour aboutir à la sonie spécifique à court terme, correspondant à la sonie provoquée par une syllabe ou un court mot et permettant entre autres de prendre en compte les effets d'inhibition controlatérale. La sonie spécifique à court terme est calculée depuis une moyenne glissante de la sonie spécifique instantanée, séparément pour chaque ERB, dans un processus comprenant un temps d'attaque T_a et un temps de relâchement T_r. En définissant S'_n, l'évaluation à court-terme glissante de la sonie spécifique au temps correspondant à la fenêtre n, S_n la sonie spécifique instantanée de la fenêtre n et S'_{n-1} la sonie spécifique à court terme de la fenêtre $n-1$, on peut exprimer S'_n :

$$S'_n = \alpha_{a/r} \cdot S_n + (1 - \alpha_{a/r}) \cdot S'_{n-1} \tag{1.11}$$

où $\alpha_{a/r}$ désigne soit α_a, une constante liée à T_a, si $S_n > S'_{n-1}$ (correspondant à une attaque) soit α_r, une constante liée à T_r, si $S_n < S'_{n-1}$ (correspondant à un relâchement), avec $\alpha_a > \alpha_r$, signifiant que la sonie spécifique à court terme peut augmenter relativement rapidement lorsqu'un son est émis, mais chute plus lentement lorsque le son est éteint, par correspondance avec la persistance de l'activité neuronale à certains stades du système auditif.

- **Lissage de la sonie spécifique à court terme** : L'inhibition controlatérale, représentant le fait que le son perçu sur une oreille puisse réduire la sonie du son perçu à l'autre oreille, intervient sur une large plage fréquentielle autour de la fréquence du son inhibiteur. La sonie spécifique à court terme est alors lissée en fréquence avant de calculer l'inhibition controlatérale. Si l'on note $N'_L(i)$ et $N'_R(i)$ les valeurs des sonies spécifiques à court-terme évoquées aux oreilles droite et gauche, respectivement, de la $i^{ème}$ ERB, la sonie spécifique à court terme de l'oreille gauche est lissée ainsi :

$$N'_{\mathrm{L}}(i)_{\text{lissée}} = \sum_{D_i=-18}^{D_i=+18} N'_L \cdot (i - D_i) \cdot \mathrm{e}^{-(B \cdot D_i)^2} \tag{1.12}$$

où D_i est la déviation depuis le i donné et évolue par pas de 0.25. B est un paramètre déterminant le degré d'étalement spectral de l'inhibition. La sonie spécifique à court terme est lissée de la même manière pour l'oreille droite.

- **Sonie spécifique inhibée à court terme** : Un facteur INH(i) est calculé pour chaque oreille, par lequel la sonie spécifique à court terme à l'ERB i est réduite. Ce facteur est déterminé ainsi :

$$\mathrm{INH}_{\mathrm{IPSI}}(i) = \frac{2}{1 + [\mathrm{sech}(\frac{N'_{\mathrm{CONTRO}}(i)_{\text{lissée}}}{N'_{\mathrm{IPSI}}(i)_{\text{lissée}}})]^{\theta}} \tag{1.13}$$

où sech est la fonction sécante hyperbolique, CONTRO désigne l'oreille controlatérale et IPSI l'oreille ipsilatérale. Ainsi, la valeur $N'_L(i)$ est divisée par $\mathrm{INH_L}$ et $N'_R(i)$ est divisée par $\mathrm{INH_R}$. La constante θ est fixée à $\theta = 1.598$. Ainsi, la fonction $[\mathrm{sech}(1)]^{\theta}$=0.5, indiquant un rapport binaural-à-monaural de 1.5 dans le cas d'une présentation diotique d'après ce modèle.

- **Somme des sonies spécifiques à court terme** : La sonie à court-terme est calculée pour chacune des deux oreilles en sommant les sonie spécifiques inhibées de chaque ERB. Ensuite, la sonie à court terme binaurale équivaut à la somme des sonies à court terme des deux oreilles.
- **Sonie à long terme** : Cette sonie est déterminée en lissant temporellement la sonie à court terme de chacune des oreilles. Si, pour une oreille donnée, S'_n désigne la valeur de la sonie à court terme au temps correspondant à la fenêtre n et S''_n la sonie à long terme à ce même temps, alors :

$$S''_n = \alpha_{al/rl} \cdot S'_n + (1 - \alpha_{al/rl}) \cdot S''_{n-1} \tag{1.14}$$

Où $\alpha_{al/rl}$ désigne soit α_{al}, une constante dépendant du temps d'attaque, dans le cas d'une attaque ($S'_n > S''_{n-1}$) soit α_{rl}, une constante dépendant du temps de relâchement, dans le cas d'une relâchement ($S'_n < S''_{n-1}$). Ici aussi, $\alpha_{al} > \alpha_{rl}$, traduisant une augmentation relativement rapide de la sonie à long terme et une chute plus lente. La sonie à long terme binaurale est obtenue en sommant la sonie à long terme des deux oreilles.

D'autres modèles de sonie binaurale existent, tel que le modèle de Sivonen et Ellermeier (2008) permettant l'estimation de sonie binaurale de sons stationnaires, dont le fonctionnement est globalement similaire au modèle de Moore *et al.* (2016) mais qui comporte quelques différences, notamment sur le stade auquel se fait la sommation binaurale. Dans leur modèle, les niveaux d'intensité des deux oreilles sont sommés avant le stade perceptif, s'éloignant d'une représentation basée sur la physiologie du système auditif. La modélisation est ainsi plus simple, réduisant le nombre d'hypothèses mises en jeu. Il est notable que ce modèle prédit un rapport de sonie binaural à monaural d'environ 1.2 en cas de présentation diotique, plus faible que dans le modèle de Moore *et al.* (2016).

Outre leur fonction de prédiction, les modèles de sonie peuvent avoir une utilité dans la vérification d'hypothèses sur les maillons de la chaîne de l'audition. Admettons que le réflexe stapédien déclenché par un signal de fort niveau sur oreille atténue également le signal perçu par l'oreille controlatérale. Il est alors possible de modéliser cette hypothèse (dans notre cas en réduisant la sonie de l'oreille controlatérale lorsque l'oreille ipsilatérale est stimulée par de forts niveaux) et de comparer les estimations établies par le modèle modifié avec les données récoltées lors d'expériences psychoacoustiques pour supporter ou non l'hypothèse émise.

1.3.6 Sonie dans la vie de tous les jours

Cette section a présenté la sonie ainsi que différents facteurs susceptibles de la modifier. Les interactions entre ces facteurs et la sonie sont pour la plupart déterminés lors d'études psychoacoustiques en laboratoire au cours desquelles les paramètres (niveau, fréquence...) de stimuli particuliers (sons purs, bruits...) sont modifiés de manière isolée. Bien que ce procédé permette d'établir clairement le lien entre chacun de ces paramètres et la sonie, il implique des conditions d'écoute s'éloignant grandement des conditions dans laquelle nous percevons les sons de notre environnement. Premièrement, les sons purs et bruits utilisés en laboratoire ne sont pas courants dans la vie de tous les jours, constituant des stimuli non familiers. Dans notre environnement quotidien, il est usuel d'entendre des sons que notre cerveau connaît et identifie, pouvant potentiellement introduire des phénomènes de plus haut niveau jouant sur la sonie que l'on en perçoit. De plus, ces études sont couramment réalisées en champ libre ou au casque afin de contrôler précisément les stimuli atteignant les oreilles des auditeurs. Ces conditions d'écoutes s'éloignent également de la réalité dans laquelle de nombreuses parois réfléchissent les ondes sonores, où de nombreuses sources sonores peuvent diffuser des sons simultanément et où les sons peuvent être perçus dans du bruit.

Nature du stimulus

Les sons auxquels nous sommes exposés dans la vie de tous les jours sont généralement vecteurs d'information. Une sirène de voiture, par exemple, sert à avertir les auditeurs et est généralement diffusée à fort niveau. Si l'on entend une sirène de voiture à faible niveau, il est plus probable qu'elle provienne d'une source sonore puissante mais lointaine que d'une source faible et proche. Si l'on prend l'exemple de la voix, que le système auditif humain traite de façon particulière (Moore, 2012d), il est aisément possible de distinguer une voix chuchotée comme étant moins puissante qu'une voix criée, même si les deux voix différentes sont présentées à un niveau physique égal. La sonie qui en découle pourrait alors être perturbée par cette distinction. Pollack (1952) a observé que la différence d'intensité nécessaire pour atteindre un doublement ou une réduction de moitié de la sonie était bien plus élevée pour de la voix que pour des bruits ou des sons purs. Il établit un rapprochement avec la « constance de taille », phénomène visuel selon lequel il est possible d'identifier la taille réelle d'un objet malgré les variations de la taille de son image sur la rétine. L'auteur justifie cette « constance de sonie » en prenant l'exemple d'un locuteur qui se déplace dans une pièce tout en parlant. Sa voix ne semblera pas varier considérablement pour ses auditeurs, malgré le fait que le niveau sonore à leurs oreilles puisse subir d'importantes variations du fait du mouvement du locuteur. Warren (1973) a obtenu des estimations de sonie pour de la voix diffusée avec différents niveaux de réverbération, certains étant trop élevés pour que la voix reste intelligible. À fort DRR, lorsque la voix est pleinement intelligible, l'auteur observe une différence de niveau de 12 dB nécessaire à un doublement de la sonie. Lorsque le niveau du champ réverbéré est fort et que la voix n'est plus intelligible, cette différence de niveau n'est plus que de 6 dB. Ces résultats font alors état d'une constance de sonie uniquement lorsque la voix est intelligible. L'utilisation de la puissance de la source dans

l'estimation de sonie va également dans le sens des résultats d'Epstein et Florentine (2009), qui ont observé un rapport de sonie binaural-à-monaural plus faible pour de la voix que pour des sons purs. Si, pour la voix, la sonie était – au moins en partie – dépendante de l'effort vocal perçu – lequel serait identique dans les conditions de présentation monaurale et binaurale car basé sur le timbre de la voix –, le rapport de sonie binaural-à-monaural serait plus faible que pour un son pur, pour lequel la sonie globale serait entièrement définie par la somme des sonies perçues par les deux oreilles.

Interactions multisensorielles

Les interactions multisensorielles, notamment celles faisant intervenir la vue et l'ouïe, sont couramment présentes dans notre environnement naturel. Il a été évoqué dans la section 1.2.5 que la vue pouvait largement influencer la localisation auditive. La sonie peut également être modifiée par l'information visuelle, comme en témoigne l'étude d'Odgaard *et al.* (2004), qui révèle une nette augmentation de sonie pour des bruits diffusés à faible niveau lorsqu'ils sont présentés de façon synchrone avec des flashs lumineux de même durée (40 ms). L'information visuelle peut également modifier le sens du stimulus auditif et en faire ainsi varier la sonie. Suzuki *et al.* (2000) ont par exemple observé que les auditeurs jugeaient un bruit blanc comme étant moins fort lorsqu'il était accompagné de l'image d'une cascade. La sonie serait également dépendante de la couleur de l'image associée au son. Un train de couleur rouge paraîtrait ainsi produire un son plus fort qu'un train de couleur verte (Patsouras *et al.*, 2002). Cet effet pourrait cependant être associé à des références culturelles des auditeurs, puisqu'il a été observé sur des auditeurs allemands et japonais (Rader *et al.*, 2004) mais n'a pas été observé sur des auditeurs français (Parizet et Koehl, 2011). Enfin, l'information visuelle peut renseigner l'auditeur sur la puissance de la source sonore et ainsi en modifier la sonie. Rosenblum et Fowler (1991) ont obtenu des estimations de sonie pour de la voix ou des claps de mains plus ou moins puissants, présentés conjointement à une vidéo représentant leur production. Les productions de voix ou de claps présentés visuellement pouvaient ne pas correspondre aux sons diffusés (le son d'un clap faible pouvait par exemple être diffusé avec une vidéo d'un clap fort). Bien que les participants avaient pour consigne d'estimer la sonie des stimuli présentés en se basant uniquement sur ce qu'ils entendaient, les auteurs ont observé une relation entre les estimations et l'information visuelle de puissance.

Contexte d'écoute

En laboratoire, les sons sont généralement restitués de manière contrôlée. Les résultats obtenus lors d'une expérience psychoacoustique ne sont alors valables que dans les conditions dans lesquelles ils ont été obtenus. L'utilisation de chambre anéchoïque permet par exemple de s'affranchir du champ réverbéré et de définir précisément les relations entre un paramètre isolé et la sonie. Cette relation ainsi observée ne sera alors véridique que dans les conditions dans lesquelles elle a été observée, à savoir en champ libre, et sera généralement extrapolée à des conditions d'écoute plus naturelles. De même, il est commun de restituer les sons au casque lors de tests

psychoacoustique afin de manipuler précisément les stimuli perçus par chaque oreille des auditeurs. Bien que l'utilité d'une telle pratique soit incontestable, elle implique une écoute non naturelle pour l'auditeur, pouvant faire intervenir des paramètres à même de modifier leurs estimations. Le rapport de sonie binaural-à-monaural peut par exemple être évalué en estimant la sonie d'un son diffusé sur un seul canal ou sur les deux canaux d'un casque. Epstein et Florentine (2009) ont comparé les rapports de sonie binaural-à-monaural obtenus de cette manière et lors d'une diffusion sur enceinte (l'écoute monaurale était alors effectuée en bouchant une oreille des auditeurs), pour de la voix parlée en direct, de la voix enregistrée et des sons purs. Le rapport de sonie binaural-à-monaural était alors systématiquement plus faible pour les sons diffusés sur haut-parleurs qu'au casque, donnant naissance à une « constance de sonie binaurale », particulièrement prononcée lorsque la source sonore était visible (Epstein et Florentine, 2012). La sonie ne dépend alors plus uniquement des signaux présentés aux oreilles des auditeurs, mais également de la validité écologique des conditions dans lesquelles les sons sont perçus.

1.4 Interactions entre position et sonie

Après avoir passé en revue les différents aspects de la localisation auditive et de la sonie, cette quatrième et dernière partie de l'état de l'art s'intéresse aux relations que partagent ces deux branches de l'audition. La sonie d'un son peut dépendre de la position de sa source de plusieurs manières, certaines liées directement aux transformations physiques liées à la position (ITD, ILD, DRR...) et d'autres étant issues de processus de plus haut niveau liés à la localisation de la source sonore.

1.4.1 En azimut et en élévation

Sonie directionnelle

La sonie directionnelle désigne la relation entre la direction d'incidence de l'onde sonore et la sonie évoquée chez l'auditeur. Robinson et Whittle (1960) ont dans un premier temps récolté des estimations de sonie pour des bruits à bande étroite diffusés depuis douze directions dans chacun des plans transversal, frontal et sagittal (voir figure 1.9) en champ libre. Les estimations étaient ensuite comparées à une modélisation réalisée à partir de mesures physiques du niveau sonore à l'entrée des conduits auditifs externes pour chaque position de la source et basée sur la « règle des 6 dB », correspondant à un gain binaural de 6 dB. Cette loi revient à exprimer le niveau L_{mon} d'un signal monaural de la forme $L_{\text{mon}} = 6 \cdot \log_2(2^{\frac{L_{\text{L}}}{6}} + 2^{\frac{L_{\text{R}}}{6}})$ pour que sa sonie soit égale à celle d'un signal binaural défini par les niveaux aux oreilles gauche L_{L} et droite L_{R}. Les bruits utilisés dans cette expérience étaient centrés sur six fréquences différentes, allant de 1.6 kHz à 10 kHz, suffisamment élevées pour que l'ombre acoustique du corps de l'auditeur ne soit pas négligeable et filtre efficacement les ondes sonores. Conformément à leurs attentes, les estimations de sonie obtenues dépendaient largement de la direction d'incidence des sons. Leur modélisation, basée uniquement sur le niveau sonore à chaque oreille, suffisait à expliquer convenablement le phénomène de sonie directionnelle observé. Dans ces conditions, la sonie directionnelle semblait alors dépendre des niveaux sonores aux oreilles plutôt que de la localisation des sons. Plus précisément dans le plan horizontal, la sonie augmentait lorsque le son était positionné sur le côté de l'auditeur, et ce dans de plus grandes proportions pour les fréquences les plus élevées jusqu'à atteindre une différence de 15 dB pour des fréquences supérieures à 8 kHz. Peu d'études se sont ensuite penchées sur ce phénomène (Remmers et Prante (1991); Jørgensen (2002) en sont des exemples, mais leurs études comptent moins de conditions expérimentales que celle de Robinson et Whittle (1960) et observent une sonie directionnelle moins marquée). Keen (1972) a observé des niveaux d'isosonie entre des stimuli de référence diotiques (des sons purs, bruits ou signaux triangulaires) et ces mêmes signaux comportant une ILD non nulle. Plus précisément, le niveau sur une oreille était fixé 2 dB à 8 dB au-dessus du niveau du stimulus de référence, et les auditeurs devaient ajuster le niveau à l'autre oreille pour que les sonies des deux stimuli soient égales. L'auteur a observé différentes fonctions d'isosonie pour chacun des stimuli. Pour les sons purs (pour lesquels la fonction d'isosonie était indépendante de la fréquence), certains participants jugeaient le stimulus à égaliser comme étant plus fort que la référence diotique même en l'absence de son à l'autre oreille lorsque le niveau sur une oreille était fixé 6 dB au-delà

du niveau du son diotique. Cette observation, en plus de témoigner d'une sommation de sonie binaurale inférieure à 6 dB, peut laisser imaginer une augmentation de sonie de sons provenant du côté. En effet, le niveau atteint sur l'oreille ipsilatérale pour ces sons latéraux est largement supérieur au niveau qui serait atteint en incidence frontale (voir section 1.2.1).

Sivonen et Ellermeier (2006) ont réalisé une deuxième étude se penchant de façon détaillée sur la sonie directionnelle, dans laquelle les participants comparaient la sonie de bruits à bande étroite (centrés sur 0.4, 1 et 5 kHz) provenant de sources situées à 2 m dans diverses directions en azimut (0°, 30°, 60°, 90° et 135°) et en élévation (45° et 90°) avec celle d'un son de référence dont la source était située en face d'eux. Les auteurs ont également mesuré les HRTF de chacun des participants afin d'étudier les proportions dans lesquelles les modifications physiques apportées aux ondes sonores expliquent la sonie directionnelle observée dans chacune des conditions expérimentales, en se focalisant uniquement sur les spectres d'amplitude témoignant de variations de niveau sonore. Cette démarche est similaire à celle de Robinson et Whittle (1960) mais, à l'époque de cette étude, le concept de HRTF et les méthodes actuelles permettant de les mesurer n'étaient pas définis. Les mesures de niveau physique se limitaient alors à une mesure à chaque oreille pour chaque fréquence discrète utilisée lors de l'expérience. Conformément aux résultats de Robinson et Whittle (1960), Sivonen et Ellermeier (2006) ont observé une sonie directionnelle plus marquée et plus dépendante de la direction de la source en hautes fréquences, pour lesquelles le corps atténue de façon plus importante les ondes sonores. Les auteurs reportent que les résultats obtenus sont convenablement prédits par une loi de sommation de sonie binaurale de 3 dB, inférieure à celle de Robinson et Whittle (1960). Globalement, la sensibilité directionnelle en sonie observée suivait les HRTF mesurées, comportant d'importantes variations interindividuelles. Cependant, certaines différences interindividuelles persistaient même après la prise en compte des HRTF individualisées. La sensibilité directionnelle en sonie observée dans cette étude ne serait alors pas exclusivement expliquée par les spectres d'amplitude des HRTF des sujets. Cette observation pourrait témoigner de phénomènes de plus haut niveau intervenant dans la sommation de sonie binaurale. L'un des questionnements des auteurs porte sur le phénomène de constance de sonie. Pour certains auditeurs, la sonie des sons aurait pu rester constante (ou du moins varier dans de moins grandes proportions) malgré les variations physiques qu'impliquent les changements de position de leur source. Ce phénomène de constance de sonie sera abordé plus en détail dans la section 1.4.2. De plus, les sensibilités directionnelles de sonie n'ont été comparées qu'au spectre d'amplitude des HRTF ne prenant pas en compte les différences interaurales de temps, utiles en localisation en azimut aux plus basses fréquences étudiées. Lors des égalisations réalisées sur des sources de différents azimuts et pour l'un de leurs huit participants en particulier, la modélisation de sommation binaurale basée sur les niveaux aux oreilles menait à une sensibilité directionnelle en sonie bien plus faible qu'observée expérimentalement à la fréquence étudiée la plus basse (400 Hz). Moore et Glasberg (2007) ont également utilisé les stimuli de Sivonen et Ellermeier (2006) en entrée de leur modèle, qui a prédit une sensibilité directionnelle en sonie plus faible qu'observée expérimentalement. Cependant, les niveaux aux oreilles ne sont pas les seuls paramètres qui dépendent de la position de la source sonore. En basses fréquences spécifiquement, la latéralisation de la source sonore engendre des ITD utiles à la localisation auditive.

Peu d'études se sont intéressées aux relations entre ITD et sonie. Deux des études les plus anciennes rapportent des résultats surprenants, témoignant d'asymétries ou de sensibilité en sonie moindre pour les sons latéralisés que frontaux : Mulligan *et al.* (1985), dans un premier temps, ont comparé les niveaux requis pour une sonie égale pour des sons purs de fréquence 250 Hz comportant une IPD nulle ou non-nulle. Les auteurs ont relevé une différence de 2.4 dB, favorisant la sonie du son dichotique, lorsqu'une IPD de 59° était introduite et retardait l'oreille droite. Étonnamment, les auteurs ont relevé une différence de 2.4 dB dans l'autre sens (le son diotique était perçu plus fort que le son dichotique) lorsque la même IPD retardait l'oreille gauche. Algom *et al.* (1988) ont étudié la relation entre la sonie de clics diffusés à différents niveaux allant de 20 dB SL à 49 dB SL ainsi que leur latéralisation en demandant à des participants d'estimer la sonie et la latéralisation pour différentes ITD. Les estimations de sonie et de latéralisation évoluaient alors de façon réciproque, les sons ayant été estimés les plus latéralisés ayant également été estimés comme étant les moins forts.

Plus récemment, Koehl et Paquier (2015) ont effectué une étude plus exhaustive sur le comportement de la sonie de sons purs basses fréquences avec l'ITD. En effectuant des comparaisons de sonie, les auteurs ont observé la différence de niveau nécessaire pour que des sons purs de fréquences 200 et 400 Hz engendrent la même sonie qu'un son pur diotique de référence lorsque différents retards interauraux y étaient appliqués (±386 µs, ±669 µs et ±772 µs, correspondant à des angles d'incidence de ±30°, ±60° et ±90°, respectivement, d'après le modèle basses fréquences de Kuhn (1977)). Les sons dichotiques étaient ajustés sur des sons diotiques de référence à deux niveaux différents, 40 et 70 phones. Les auteurs rapportent un effet de l'ITD sur la sonie pour le plus faible des niveaux étudiés (40 phones) seulement, pour lequel la sonie des sons purs augmentait proportionnellement à l'ITD qu'ils comportaient et ce de la même manière selon que le retard était appliqué sur l'oreille droite ou gauche. Les sons purs dichotiques paraissaient alors aussi forts que leurs équivalents diotiques lorsque leur niveau était inférieur d'environ 1.25 dB, différence faible mais néanmoins supérieure aux seuils différentiels d'intensité à ces fréquences, qui sont de l'ordre de 1 dB à 200 Hz à 30 phones et décroissent encore lorsque le niveau sonore augmente (Fletcher, 1953).

Koehl *et al.* (2015) ont ensuite défini le rôle de la localisation auditive dans l'effet de l'ITD sur la sonie. En effet, puisque l'ITD sert d'indice de localisation, elle induit une sensation de latéralisation des sons et l'augmentation de sonie observé pour les sons comportant une ITD peut être causée soit par l'ITD elle-même, soit par la sensation de latéralisation qu'elle induit. Dans cette étude, les auditeurs effectuaient des ajustements de sonie de sons purs comportant une ITD de 772 µs sur des sons purs de référence diotiques à un niveau de 40 phones, dans des cas où une ILD de 2.5 ou 5 dB menait soit du même côté que l'ITD, soit sur l'autre côté. Les auteurs ont observé un effet de l'ITD sur la sonie indépendant de l'ILD, favorisant l'hypothèse que l'effet de l'ITD sur la sonie (et, par extension, la sonie directionnelle en basses fréquences) était dû aux indices de localisation plutôt qu'à la localisation qu'ils induisent. Les auteurs ont également réalisé une étude des seuils absolus d'audition à différentes fréquences (125, 200, 250, 400 et 500 Hz) pour différentes valeurs d'ITD (0, 386, 669 et 772 µs). Cette étude a été réalisée en suivant une

hypothèse selon laquelle l'augmentation de sonie avec l'ITD serait due à une distinction du signal dichotique par rapport au bruit interne inhérent au système auditif, lequel serait partiellement corrélé (Yost, 1988) et comporterait essentiellement du contenu basses fréquences (Shaw et Piercy, 1962). Les auteurs n'ont cependant observé aucune différence significative entre les seuils obtenus pour les différentes ITD. Toutefois, cette absence d'effet de l'ITD au seuil pourrait s'expliquer par le fait que les seuils différentiels de niveau sonore augmentent lorsque le niveau se rapproche du seuil (Fletcher, 1953) et pourraient devenir supérieurs à la taille de l'effet de l'ITD sur la sonie.

Dans une dernière étude, Koehl *et al.* (2017) ont observé l'évolution de cet effet pour différentes fréquences plus élevées que celles étudiées précédemment. Les stimuli à égaliser étaient ici des sons purs de fréquence 500, 707, 1000, 1414 et 2000 Hz et comportaient une ITD de 0, 607, 662, 717, 772 et 827 µs, 662 µs correspondant à l'ITD d'un son dont la source est localisée à un azimut de 90° d'après le modèle de Woodworth (1954). Les auditeurs avaient pour tâche d'égaliser ces stimuli en sonie sur des références diotiques de fréquence égale. Les auteurs ont alors observé une augmentation de la sonie due à l'ITD à la fréquence étudiée la plus basse (500 Hz) uniquement. L'augmentation de sonie était d'un ordre similaire à celle observée précédemment.

Démasquage binaural

Le démasquage représente la capacité de notre système auditif à extraire un signal présenté dans un bruit masquant. Lorsque les stimuli (signal à détecter et bruit masquant) sont binauraux, différents seuils de masquage (ou seuil masqué, différence de niveau entre le signal et le bruit nécessaire pour que le bruit soit démasqué) peuvent être observés selon les propriétés binaurales respectives des deux stimuli. En particulier, les seuils de masquage peuvent être réduits lorsque le signal est perçu sur le côté en comparaison à une présentation frontale. Grange et Culling (2016) ont étudié ce phénomène pour de la voix. Dans leur étude, un haut-parleur frontal diffusait de la voix en présence de bruit masquant. Les auditeurs avaient pour consigne d'orienter leur tête dans le but d'extraire au mieux le signal vocal. Les participants tournaient alors spontanément la tête pour s'éloigner d'une incidence frontale. L'augmentation de l'intelligibilité dans cette nouvelle orientation de la tête est due d'une part à un phénomène d'« écoute à la meilleure oreille » (*better ear listening*), selon lequel l'augmentation du niveau sur l'une des deux oreilles permet une meilleur intelligibilité de la voix (Edmonds et Culling, 2006). D'autre part cette augmentation de l'intelligibilité vient de phénomènes de démasquage binaural, où les disparités interaurales du signal ou du bruit peuvent modifier les seuils de masquage. Ces phénomènes ont été étudiés avec des cas plus fondamentaux, où les seuils de masquage sont étudiés en fonction de la corrélation interaurale ρ du signal et du bruit. La présentation d'un signal d'une corrélation ρ_S donnée dans un bruit de corrélation ρ_N donnée est communément représentée de la forme $\mathrm{N}_{0/\pi/\mathrm{u}/\mathrm{m}}\mathrm{S}_{0/\pi/\mathrm{m}}$, N désignant la corrélation interaurale du bruit et S celle du signal. L'indice $_0$ signifie que le son (bruit ou signal) est diotique (les signaux aux deux oreilles sont corrélés, $\rho = 1$), $_\pi$ qu'il est en opposition de phase entre les deux oreilles (les signaux aux deux oreilles sont anticorrélés, $\rho = -1$), $_\mathrm{m}$ qu'il est monaural et $\mathrm{N_u}$ signifie que le bruit est décorrélé entre les deux

oreilles (les bruits diffusés sur l'oreille gauche et droite sont différents, $\rho = 0$). Par exemple, si un signal binaural dont les composantes envoyées aux oreilles droite et gauche sont anticorrélées est diffusé conjointement à un bruit masquant diotique, la condition sera notée N_0S_π. Hirsh (1948) a dans un premier temps mis en évidence que les seuils de masquage étaient plus faibles lorsque la corrélation interaurale du signal et celle du bruit étaient différentes. La différence entre les niveaux du signal au seuil de masquage dans les conditions N_0S_0 et N_0S_π, notée MLD (pour *Masking Level Differences*, différences de niveau de masquage), peut atteindre 15 dB en basses fréquences (autour de 500 Hz) et est plus négligeable pour des fréquences plus hautes lorsque le bruit masquant est large bande (2 dB à 3 dB à 1.5 kHz) d'après Durlach (1978). McFadden *et al.* (1972) ont cependant observé d'importantes différences de seuils en fonction de la corrélation interaurale du signal à 1 kHz en utilisant des bruits masquants à bande étroite. Ainsi, si l'on diffuse un signal diotique juste en-dessous de son seuil de masquage dans un bruit diotique (N_0S_0), le fait d'inverser la phase du signal reçu par l'une des deux oreilles (N_0S_π) permet à l'auditeur d'entendre le signal aux alentours de 15 dB SL en basses fréquences (voir figure 1.26 (a) et (b)). Plus étonnant encore, si l'on se place juste en-dessous du seuil de masquage d'un signal monaural lorsqu'un bruit monaural est diffusé sur la même oreille (N_mS_m) et que l'on rajoute du bruit sur l'autre oreille de façon à ce que le bruit masquant devienne diotique (N_0S_m), le signal peut être entendu (voir figure 1.26 (c) et (d)). Bien que le niveau global du bruit masquant résultant du processus de sommation de sonie binaurale soit plus élevé, le fait que le signal soit dichotique et le bruit diotique permet au cerveau de plus facilement extraire le signal du bruit. L'amplitude des MLD est dépendante du niveau et décroît lorsque le niveau du bruit masquant est faible (Hirsh, 1948).

Diercks et Jeffress (1962) ont observé un seuil d'audition absolu plus faible pour des signaux anticorrélés que pour des signaux diotiques. Cette observation a amené les auteurs à formuler l'hypothèse selon laquelle le seuil d'audition absolu correspondrait au seuil de masquage dans le bruit interne inhérent au système auditif. Cependant, la nature exacte de ce bruit interne n'est pas à ce jour pleinement définie. D'après Bernstein et Trahiotis (2008), ce bruit serait composé de deux sources. L'une correspondrait au bruit physiologique du corps humain, tel que mesuré par Shaw et Piercy (1962), serait indépendante du stimulus et aurait une corrélation interaurale faible et négative. L'autre source serait due à une dégradation de l'information du stimulus entrant dans le processus de traitement binaural ou à des erreurs créées lors du traitement binaural des sons. Certains auteurs soumettent cependant que la corrélation interaurale du bruit interne est faible et positive, les différences avec les observations de Bernstein et Trahiotis (2008) pouvant s'expliquer par les différences de fréquence des stimuli étudiés. Le fait que l'amplitude des MLD décroisse lorsque le niveau d'un bruit masquant diotique décroît (Hirsh, 1948) pourrait s'expliquer, d'après Yost (1988), par le fait qu'à faible niveau le bruit masquant se mêle au bruit interne. Dans l'hypothèse où le bruit interne est au moins partiellement décorrélé, le bruit total résultant de la somme du bruit diotique et du bruit interne aurait une corrélation interaurale inférieure à 1, menant ainsi à une MLD plus faible (Robinson et Jeffress, 1963). La corrélation interaurale de ce bruit total a été modélisée par McFadden (1968) pour la condition où le bruit externe est diotique. Dans cette modélisation, le bruit interne est composé de trois composantes,

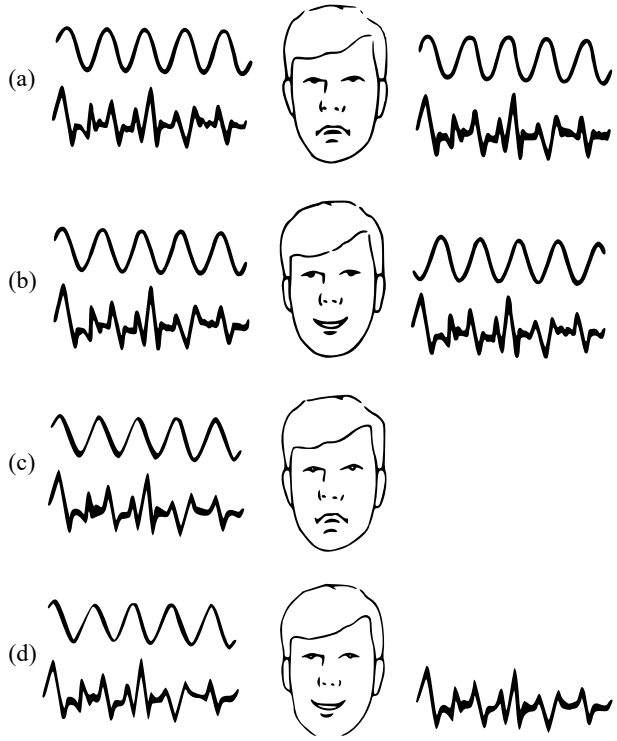


FIGURE 1.26 – Démasquage binaural. Dans les conditions (a) et (c), les relations entre les deux oreilles sont similaires pour le signal et pour le bruit, rendant le démasquage du signal difficile. Dans les conditions (b) et (d), le bruit est diotique, ce qui n'est pas le cas du signal, rendant son démasquage plus simple. D'après Moore (2012c).

l'une notée f_1 correspondant à la partie commune aux deux oreilles (partie corrélée du bruit interne) et les deux autres, f_2 et f_3, correspondant aux parties respectives des oreilles droite et gauche et étant décorrélées l'une à l'autre. La corrélation interaurale du bruit interne ρ_I, d'après cette modélisation, est déterminée par le rapport :

$$\rho_I = \frac{\sigma_{CI}^2}{\sigma_{CI}^2 + \sigma_{UI}^2} \tag{1.15}$$

où σ_{CI}^2 désigne la puissance de f_1 et σ_{UI}^2 les puissances de f_2 et f_3, présumées égales. Expérimentalement, les auteurs déterminent une corrélation $\rho_I = +0.35$ d'après les données de Diercks et Jeffress (1962) et Robinson et Jeffress (1963). Le facteur σ_{CI}^2 a alors été fixé à 1.0 dans la modélisation, et la valeur de σ_{UI}^2 en a été déduite de façon relative à ρ_I et à σ_{CI}^2. Lorsqu'un bruit externe diotique est ajouté, la corrélation interaurale du bruit total ρ_{xy} devient :

$$\rho_{xy} = \frac{\sigma_{CI}^2 + a^2 \cdot \sigma_e^2}{\sigma_{CI}^2 + \sigma_{UI}^2 + a^2 \cdot \sigma_e^2} \tag{1.16}$$

où σ_e^2 désigne la puissance du bruit externe diotique et a^2 est un facteur d'atténuation utilisé par les auteurs pour comparer leur modélisation aux résultats expérimentaux lorsque différents niveaux de bruit externe étaient ajoutés.

Corrélation interaurale	MLD (dB)
N_uS_π	3
N_uS_0	4
$N_\pi S_m$	6
N_0S_m	9
$N_\pi S_0$	13
N_0S_π	15

Table 1.2 – Valeur de la MLD pour différentes relations interaurales du signal et du bruit. Les résultats sont donnés pour des bruits masquants large bande et des signaux basses fréquences. D'après Moore (2012c).

Lorsque le rapport signal-à-bruit augmente et que le niveau du signal devient supraliminaire, la sonie du signal peut être affectée par le rapport entre sa corrélation interaurale et celle du bruit, bien que des résultats contradictoires aient été observés. Townsend et Goldstein (1972) ont relevé que les différences de corrélation interaurale entre signal et bruit pouvaient affecter la sonie d'un son pur de fréquence 250 Hz jusqu'à un niveau de 20 dB SL, le plus haut niveau que les auteurs aient testé. À la même fréquence, Zwicker et Henning (1991) ont montré que l'IPD pouvait faire varier la sonie du signal jusqu'à des niveaux de 30 dB SL à 40 dB SL. Cependant, Soderquist et Shilling (1990) n'ont pas observé d'interaction entre les corrélations interaurales du signal et du bruit et la sonie à des niveaux supérieurs à 10 dB SL à 200 Hz.

La théorie d'égalisation-annulation (EC, pour *Equalization-Cancellation theory*), proposée par Durlach (1963) et soutenue récemment par Culling (2007), émet l'hypothèse que le système auditif extrait le signal du bruit masquant en réalisant deux étapes. La première, l'étape d'égalisation, consiste à appliquer un certain nombre de transformations aux signaux perçus par les deux oreilles pour que les bruits masquants des deux oreilles soient identiques. La deuxième consiste en une soustraction des signaux modifiés, permettant la suppression du bruit masquant. Cette modélisation permet d'expliquer convenablement les différentes MLD répertoriées par Moore (2012c) et représentées tableau 1.2. Dans le cas N_0S_π menant à la plus grande MLD, l'étape d'égalisation n'est pas nécessaire et l'étape d'annulation mène à une amplitude nulle pour le bruit et au double de l'amplitude initiale pour le signal, comme dans le cas $N_\pi S_0$ après l'étape d'égalisation, correspondant ici à un déphasage de π du son perçu par l'une des deux oreilles. Dans les cas où le signal et monaural, l'égalisation-annulation mène également à un résultat nul pour le bruit, mais où l'amplitude du signal est identique à l'amplitude initiale, menant à une MLD moindre. Enfin, lorsque le bruit est décorrélé entre les deux oreilles, la théorie EC ne permet pas d'éliminer efficacement le bruit masquant, menant à une faible MLD.

1.4.2 En distance

Un phénomène de constance de sonie, selon lequel la sonie d'un son reste constante malgré les variations de niveau induites par son éloignement, a été observé lors de différentes études dont les résultats varient selon les paramètres mis en jeu. Cette constance de sonie est souvent comparée à la constance de taille observée en perception visuelle. Ce phénomène sera évoqué dans un premier temps pour ensuite introduire par analogie la constance de sonie.

Lorsqu'un objet s'éloigne du point d'observation, la taille de l'image qu'il forme sur la rétine est de plus en plus petite. La constance de taille décrit le phénomène selon lequel il est possible d'identifier la taille réelle d'un objet situé à différentes distances malgré les différences de tailles dans le champ de vision (Kilpatrick et Ittelson, 1953). Le dessin *Terror Subterra* présent dans le livre de Shepard (1990) permet d'illustrer clairement cet effet (voir figure 1.27 (a) et son adaptation en (b)). Chacun des deux dessins présente deux personnages situés à des distances différentes dans un tunnel. On souhaite étudier la perception de leur taille réelle (la taille des personnages) ainsi que leur taille apparente (la taille de leurs images). Dans le cas (a), les deux personnages ont la même taille sur l'image et notre cerveau identifie le personnage du premier plan comme étant plus petit que le personnage du second plan. Dans le cas (b), la taille sur l'image du personnage du premier plan est plus grande que celle du personnage du second plan et notre cerveau identifie les deux personnages comme étant de tailles identiques. Si l'on supprime le tunnel et que l'on représente les personnages seuls (voir figure 1.28), il est plus difficile de tirer des conclusions sur les tailles « réelles » des personnages, que le cerveau ne peut alors déduire que depuis la taille de leurs images. Globalement, les personnages semblent avoir la même taille dans le cas (a), puisque la taille de leurs images sont identiques. Dans le cas (b), l'observateur peut réaliser différentes interprétations basées sur sa familiarité avec l'objet dont la taille est à estimer, laquelle joue un rôle important dans la constance de taille perçue (Bolles et Bailey, 1956). L'observateur peut dans un premier temps se baser uniquement sur la taille des images et juger le personnage de gauche comme étant plus grand que celui de droite. Une autre interprétation est possible, basée sur le fait que les personnages dessinés ont une forme humanoïde relativement familière et que leurs proportions sont identiques. L'observateur peut, d'après ce constat, déduire qu'ils ont la même taille « réelle » et qu'ils sont situés à des distances différentes, l'absence d'indice de distance (autre que la hauteur dans le champ visuel) permettant une interprétation libre de cette variable. Si l'on reproduit ces scènes avec des objets non familiers, par exemple des formes géométriques quelconques, l'observateur ne peut plus déduire les dimensions des objets par identification. Dans la figure 1.29 (a), les indices de profondeur fournis par le tunnel permettent encore de se représenter la forme ovale comme étant plus grande que la forme carrée, bien que les dimensions de leurs images soient similaires. Dans le cas (b) cependant, aucun indice ne permet de déduire les tailles des deux objets, et l'observateur ne peut baser son jugement que sur la taille des images. Les relations entre sonie et distance de la source dépendent de façon similaire de différents paramètres comparables par analogie avec ceux évoqués ci-dessus.

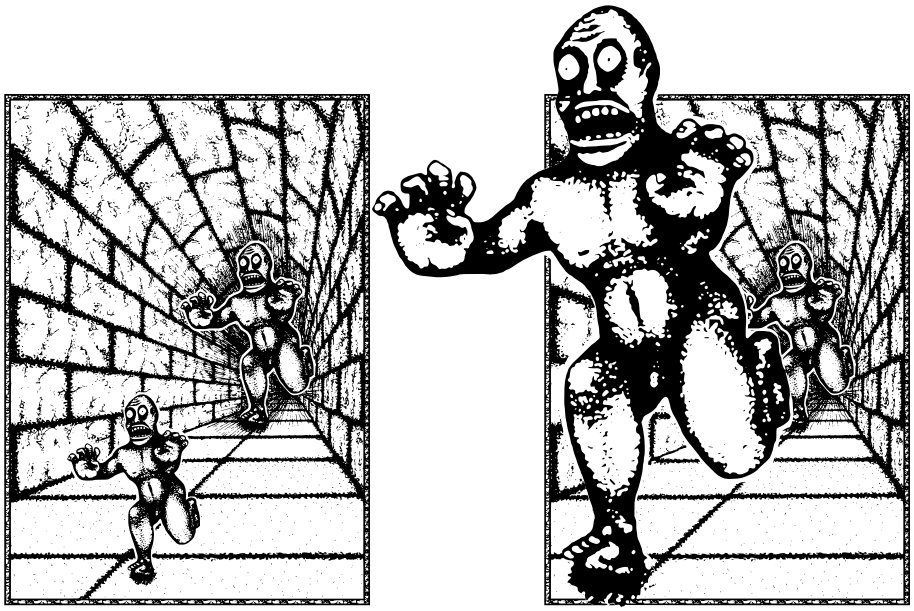


FIGURE 1.27 – Constance de taille visuelle. Dans le dessin (a), la taille des deux personnages est identique sur l'image et notre cerveau perçoit le personnage au premier plan comme étant plus petit que le personnage au second plan. Dans le dessin (b), la taille de l'image du personnage au premier plan est plus grande que celle du personnage au second plan, et notre cerveau perçoit les deux personnages comme étant de tailles identiques. Adapté de *Terror Subterra*, tiré de Shepard (1990).

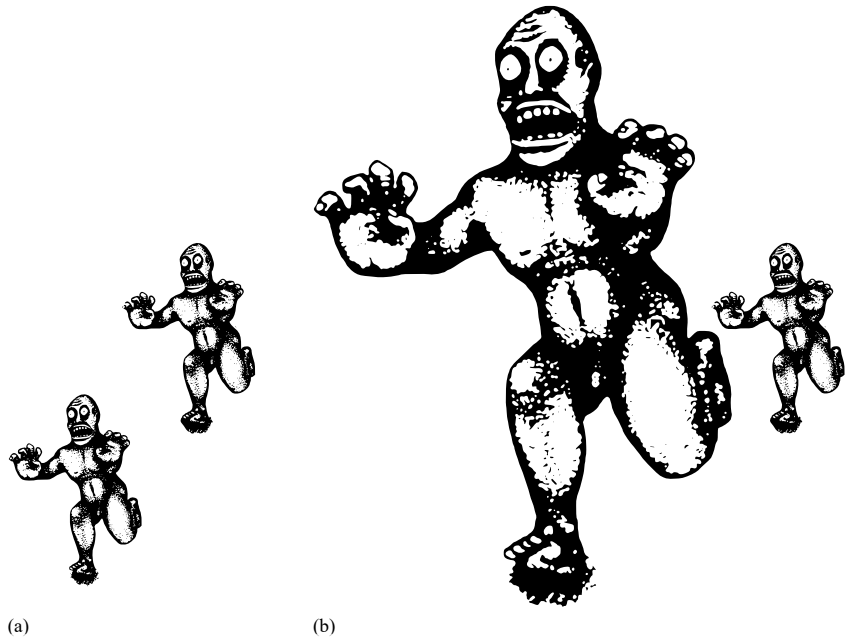


FIGURE 1.28 – Les personnages sont les mêmes que dans la figure 1.27, représentés sans les indices de profondeur apporté par le tunnel. Le personnage au second plan du cas (b) a été légèrement décalé pour supprimer l'indice d'occlusion. Dans le cas (a), les deux personnages ont et la même taille sur l'image et le cerveau les perçoit comme étant de la même taille perçue. Dans le cas (b), l'image du personnage au premier plan est plus grande que celle de l'autre personnage et plusieurs interprétations sont possibles. Adapté de *Terror Subterra*, tiré de Shepard (1990).

Lorsqu'une source sonore s'éloigne de l'auditeur, le niveau sonore à ses oreilles décroît (voir section 1.2.3), de la même manière que la taille de l'image d'un objet décroît lorsqu'il s'éloigne de l'observateur. Un phénomène de constance de sonie, selon lequel la sonie d'un son reste constante malgré les variations de niveau induites par son éloignement, a été reporté par plusieurs auteurs. Cette constance de sonie n'est pas systématique et dépend de plusieurs paramètres, dont la quantité d'information de distance (telle qu'apportée par le tunnel dans les figures 1.27 et 1.29) ainsi que la familiarité de la source sonore (différenciant les sources connues et identifiables, comme par analogie les personnages dans les figures 1.27 et 1.28, des sources non-identifiables, comme par analogie les formes de la figure 1.29). Les différentes études rapportent également des résultats différents selon les instructions données aux participants :

- Il est possible d'orienter l'attention de l'auditeur sur la force sonore au niveau de la source (*loudness at the source* (Sivonen et Ellermeier, 2011)). Dans ce cas, la tâche des auditeurs est d'indiquer à quel point le son est fort *au niveau de la source*. Ces estimations sont similaires à des estimations de puissance de la source sonore. Elles peuvent être comparées par analogie avec la perception visuelle à l'estimation de la taille « réelle » d'un objet. Dans la suite de ce manuscrit, ce type de sonie sera décrit sous le terme de « sonie *à la source* ».

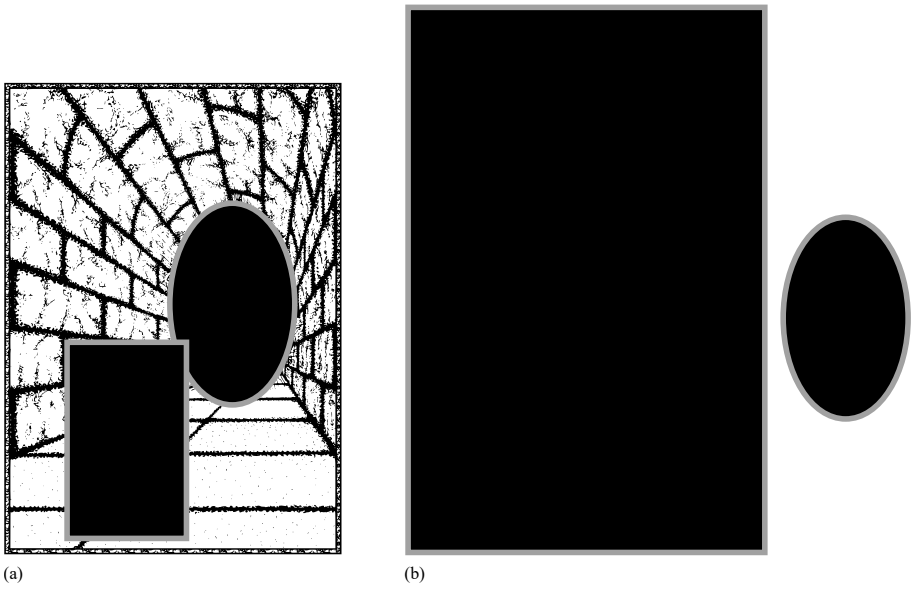


FIGURE 1.29 – Les personnages ont été remplacés par des formes, rendant moins évidentes les conclusions sur leur taille « réelle ». Sans indice de distance (b), le cerveau ne peut déterminer la taille réelle qu'à partir de la taille des images. Adapté de *Terror Subterra*, tiré de Shepard (1990).

- Il est également possible d'orienter l'attention de l'auditeur sur la force sonore au niveau de ses oreilles (*loudness at the ear* (Mershon *et al.*, 1981)). Dans ce cas, la tâche des auditeurs est d'indiquer à quel point le son est fort *au niveau de leurs oreilles*. Ces estimations reviennent à décrire la sensation de niveau sonore « entendu » provoquée par le son. Elles peuvent être comparées par analogie avec la perception visuelle à l'estimation de la taille de l'image d'un objet sur la rétine. Dans la suite de ce manuscrit, ce type de sonie sera décrit sous le terme de « sonie *aux oreilles* ».
- Enfin, il est possible de ne pas préciser à l'auditeur sur quoi porter son attention, en lui demandant simplement d'estimer la force sonore qu'il perçoit du son, en lui demandant par exemple de spécifier à quel point le son « lui paraît fort ». Si l'on compare par analogie avec la perception visuelle, cela reviendrait à demander à un observateur à quel point un objet « lui paraît grand ». L'auditeur (ou l'observateur) peut alors interpréter librement la consigne.

L'étude consacrée à ce phénomène la plus souvent rapportée dans la littérature a été réalisée par Zahorik et Wightman (2001). Dans cette étude, les participants réalisaient des estimations de sonie *à la source* et de distance selon un protocole d'estimation de grandeur sans référence pour des bruits diffusés au casque. Les bruits étaient virtuellement placés à douze distances différentes du point d'écoute par l'intermédiaire de convolutions avec des fonctions de transfert binaurales de salle (BRIR, pour *Binaural Room Impulse Response*) d'une salle semi-réverbérante et diffusés à différents niveaux (-30, -10, -5, 0, $+5$ et $+10$ dB par rapport à un niveau de référence fixé à 68 dB SPL à 30 cm de la source). En effectuant une régression des résultats sur des fonctions puissance de la forme $L = k \cdot r^b$, où L est la sonie *à la source* perçue, r la distance physique

de la source, k une constante, et b l'exposant de la fonction de sonie ainsi obtenue, les auteurs ont relevé un exposant b proche de 0 témoignant d'une forte constance de sonie avec la distance. Il est important de préciser que les participants avaient pour consigne de baser leur estimation sur la sonie *à la source*, revenant à estimer la puissance des sources sonores indépendamment du niveau à leurs oreilles. Deux raisonnements différents peuvent mener à ce jugement pour une telle présentation des sons. La première serait d'utiliser les différents indices de distance disponibles pour estimer la distance de la source, pour ensuite ramener le niveau perçu aux oreilles au niveau de la source. La deuxième, plus directe, se base uniquement sur l'exploitation du niveau du champ réverbéré. Puisque ce dernier est approximativement constant avec la distance (les auteurs ont mesuré une décroissance d'environ 1 dB par doublement de distance), il ne dépend approximativement que de la puissance de la source et en constitue un indice absolu direct. Les auteurs supportent cette deuxième hypothèse d'après deux constats. Le premier est que les estimations de distance des sujets montraient un écart considérable avec les distances réelles. Le deuxième est qu'au plus faible niveau de diffusion étudié, pour lequel une certaine partie de l'énergie du champ réverbéré se situait sous la courbe de seuil des auditeurs, l'exposant b de la fonction de sonie était de 0.35, témoignant d'une absence de constance de sonie dans le cas où le champ réverbéré n'était plus entièrement audible.

Mohrmann (1939), dans une étude décrite par Brunswik (1956), a demandé à des auditeurs d'ajuster le niveau de sources visibles ou non, diffusant des sons familiers (voix et musique) et non familiers (sons purs et bruits), situées à différentes distances dans une cabine d'enregistrement peu réverbérante afin d'égaliser leur sonie en se basant soit sur la sonie *à la source*, soit sur la sonie *aux oreilles*. Les égalisations ainsi réalisées étaient alors plus précises dans le cas de la sonie *à la source* lorsque les stimuli étaient familiers et lorsque les sources étaient visibles. Ces résultats confortent l'hypothèse d'une prévalence naturelle de la sonie *à la source* sur la sonie *aux oreilles* pour les sons familiers. Une telle observation n'a cependant pas été faite sur les résultats obtenus pour des sources de sons non familiers, surtout lorsque les sources n'étaient pas visibles. Selon une procédure similaire effectuée en environnement extérieur peu réverbérant et avec des sources visibles, Shigenaga (1965) a observé une constance de sonie pour des sons familiers (voix et musique) et non familiers (sons purs) lorsque les auditeurs avaient pour consigne de baser leurs égalisation sur la sonie *à la source*. Dans cet environnement peu réverbérant, limitant l'utilisation de l'énergie du champ réverbéré dans l'estimation de sonie *à la source*, les auditeurs auraient alors réalisé des estimations constantes à l'aide de la distance perçue visuellement des sources sonores.

Les études dans lesquelles les auditeurs ont pour tâche de baser leur jugement sur la sonie (sans précision) ou sur la sonie *aux oreilles* relèvent généralement une constance de sonie plus faible, voire inexistante. Stevens et Guirao (1962) ont récolté des estimations de sonie, douceur (*softness*) et distance apparentes pour des bruits et sons purs présentés au casque. Les estimations étaient faites selon plusieurs méthodes et ont mené à une relation réciproque entre les jugements de sonie et de distance, témoignant d'une absence de constance de sonie. Il est à noter que les estimations de distance étaient uniquement basées sur le niveau sonore puisqu'aucun autre indice n'était disponible. Une absence de constance de sonie similaire a été obtenue par Mershon *et al.*

(1981), qui ont également étudié les relations entre distance et sonie dans le cas où la distance est définie à l'aide d'indices visuels. Les estimations réalisées étaient, d'après les auteurs, des estimations d'*apparent loudness*. Ces estimations de sonie ont été réalisées pour des bruits diffusés à un niveau fixe dans des salles anéchoïque et semi-réverbérante par un haut-parleur caché situé à une position fixe, tandis qu'un faux haut-parleur visible était placé à différentes distances des auditeurs. Les estimations de sonie ainsi obtenues étaient alors plus grandes lorsque le faux haut-parleur était situé loin des auditeurs, conformément avec le phénomène de constance de sonie, et ce particulièrement dans l'environnement semi-réverbérant, où des indices auditifs de distance viennent pourtant s'opposer à la distance perçue visuellement. Dans la salle anéchoïque, les résultats n'étaient pas si clairs et les estimations de sonie devenaient plus faibles pour la plus grande distance testée du faux haut-parleur. Enfin, Altmann *et al.* (2012) ont obtenu des estimations de sonie couplées à des mesures par magnétoencéphalographie (MEG) lorsque leurs participants entendaient des bruits large bande présentés via des écouteurs et diffusés conjointement à un signal lumineux situé à différentes distances. Bien que les réponses des mesures faites par MEG étaient significativement plus fortes pour les stimuli visuels distants que proches, aucune différence significative n'a été observée sur les estimations de sonie.

De manière générale, la constance de sonie est davantage prononcée lorsque la tâche est d'effectuer des estimations de sonie *à la source* que pour des estimations de sonie *aux oreilles*. Elle semble également être plus forte lorsque les sons sont familiers, lorsque les sources sont visibles et que l'environnement dispose d'un champ réverbéré. Il est en effet plus simple d'identifier la puissance d'un son familier, tel que de la voix (un chuchotement est identifiable comme étant faible, un cri comme étant fort), que celle d'un son non familier (au delà du niveau perçu, il n'est pas possible d'identifier le niveau de diffusion d'un son pur ou un bruit diffusé à fort ou faible niveau). Ce phénomène reste néanmoins difficile à étudier expérimentalement puisqu'il nécessite de diffuser des sons depuis des sources situées à différentes distances, pouvant poser un certain nombre de problèmes : le protocole de Mershon *et al.* (1981), par exemple, impose une incohérence entre la distance du stimulus sonore et celle du stimulus visuel. Il est également compliqué de manipuler rapidement et discrètement la position de la source sonore, comme en témoigne le dispositif expérimental de Calcagno *et al.* (2012) dont l'objectif était de déterminer les relations entre perceptions auditive et visuelle de distance. Ce dernier comportait un haut-parleur mobile déplacé par un rail. Puisque le bruit du rail était susceptible d'informer les participants sur le déplacement du haut-parleur, des haut-parleurs supplémentaires diffusaient du bruit à fort niveau entre chaque essai pour masquer le son du rail, imposant aux participants d'entendre du bruit à fort niveau susceptible de les distraire et ainsi d'affecter la perception des stimuli sonores de l'expérience. Contrairement aux études portées sur la sonie directionnelle, il n'est pas évident de placer différentes sources à différentes positions puisque les sources les plus éloignées se retrouveraient masquées par les sources les plus proches. Bien que les études au casque (Zahorik et Wightman, 2001) permettent de manipuler aisément les positions de sources sonores, elles posent un problème de non-externalisation des sources, où les auditeurs perçoivent les sources à l'intérieur de la tête plutôt qu'à l'extérieur (Hendrickx *et al.*, 2017b). Les relations entre distance perçue et externalisation des sources sonores ne sont pas clairement définies à ce jour. Kolarik

et al. (2016) reporte que l'externalisation est un pré-requis pour percevoir la distance d'une source sonore. Il est ainsi envisageable que la notion de distance apparente soit différente selon qu'un source sonore est perçue à l'intérieur de la tête ou qu'elle est perçue à l'extérieur de la tête.

Première partie

Sonie et azimut

Chapitre 2

Étude de l'effet de l'ITD sur la sonie de sons purs BF à différents niveaux

Cette étude a été présentée à la conférence *Acoustics '17 Boston* (Berthomieu *et al.*, 2017).

2.1 Introduction

La section 1.4 a présenté différentes propriétés de la sonie témoignant de sa dépendance à la position de la source sonore. Cette sensation est généralement plus élevée lorsque le son vient du côté de l'auditeur que s'il est en incidence frontale. En basses fréquences, l'ITD augmente significativement la sonie pour des sons purs, seulement lorsque ceux-ci sont diffusés à faible niveau (Koehl et Paquier, 2015; Koehl *et al.*, 2015). La première étude de cette thèse a été mise en place dans le but d'analyser jusqu'à quelle fréquence et jusqu'à quel niveau l'ITD augmente la sonie. Ainsi, l'effet de l'ITD a été étudié sur des sons purs de deux basses fréquences différentes, 100 et 200 Hz, restitués à sept niveaux d'isosonie allant de 30 phones à 90 phones par pas de 10 phones.

2.2 Matériel et méthode

Dans cette expérience, la tâche des participants était de comparer la sonie de sons purs diotiques et dichotiques (auquel cas ils comportaient une ITD non nulle) à la sonie de *sons de référence* diotiques de même fréquence. Les sons purs étaient diffusés à différents niveaux de référence dans le but d'étudier l'évolution de l'effet de l'ITD sur la sonie avec le niveau.

2.2.1 Stimuli

D'après les résultats de Koehl et Paquier (2015), l'effet de la sonie sur l'ITD semble plus prononcé à 200 Hz qu'à 400 Hz. Les stimuli de la présente étude étaient des sons purs de fréquence 100 Hz et 200 Hz. La valeur de l'ITD appliquée aux stimuli dichotiques a été choisie pour correspondre à l'ITD d'un stimulus localisé à un azimut $\theta = 90°$. Cette valeur a été déterminée d'après le modèle d'ITD pour basses fréquences établi par Kuhn (1977), valide pour les fréquences inférieures à 500 Hz :

	100 Hz	200 Hz
30 phones	56.76	44.78
40 phones	64.37	53.41
50 phones	71.61	61.72
60 phones	78.65	69.86
70 phones	85.60	77.92
80 phones	92.48	85.92
90 phones	99.33	93.89

Table 2.1 – Équivalence entre niveau de pression sonore (en dB SPL) et niveau d'isosonie (en phone) à 100 Hz et 200 Hz d'après la norme ISO 226 (2003).

$$\mathrm{ITD}_{\mathrm{BF}} = \frac{3r}{c_0} \sin(\theta) \tag{2.1}$$

où $a = 8.75\,\mathrm{cm}$ est le rayon moyen d'une tête humaine, $c_0 = 340\,\mathrm{m} \cdot \mathrm{s}^{-1}$ est la vitesse du son dans l'air et θ est l'azimut de la source, avec $\theta = 90°$ pour une source localisée sur le côté de l'auditeur. Ce modèle donne une ITD de 772 µs (i.e. le son était présenté sur une oreille 772 µs plus tôt que sur l'autre oreille). Les sons dichotiques à égaliser étaient systématiquement présentés en avance sur la même oreille (l'oreille gauche) puisque l'effet de l'ITD sur la sonie est symétrique d'après Koehl et Paquier (2015). Ces stimuli devaient être égalisés en sonie sur des stimuli de référence diotiques dont les niveaux étaient fixés à 30, 40, 50, 60, 70, 80 et 90 phones. Les niveaux de pression sonore à délivrer pour atteindre ces niveaux d'isosonie à 100 Hz et 200 Hz ont été calculés d'après la norme ISO 226 (2003) et sont donnés dans le tableau 2.1. Les sons étaient générés par une carte son RME Babyface Pro et le système a été étalonné en plaçant le casque audio utilisé lors du test (Sennheiser HD 650, casque circumaural ouvert) sur une tête artificielle (Neumann KU100), laquelle était étalonnée à l'aide d'un pistonphone (Brüel & Kjær Type 4231) à 1 kHz. Chaque stimulus durait 1.6 s et comportait des rampes d'attaque et de relâchement de 100 ms (comme les stimuli utilisés par Koehl et Paquier (2015)).

2.2.2 Procédure

Vingt-deux participants (huit femmes et quatorze hommes âgés de vingt à vingt-cinq ans) ont pris part à cette étude et étaient rémunérés pour leur participation. Chacun d'eux avait des seuils d'audition normaux ($\leq$ 10 dB HL) d'après un audiogramme passé dans le mois précédant l'expérience. Aucun d'entre eux n'avait d'expérience en test d'écoute en laboratoire.

Les stimuli était égalisés en utilisant une procédure adaptative à choix forcé à deux intervalles et deux alternatives (2I2AFC) suivant une règle « 1-up 1-down » (Levitt, 1971). Cette procédure était la même que celle utilisée dans les études précédentes s'étant intéressées à la sonie directionnelle (Sivonen et Ellermeier, 2006) et aux effets de l'ITD sur la sonie (Koehl et Paquier, 2015; Koehl *et al.*, 2015). Dans chaque essai, un *son test* (diotique ou dichotique) et un *son de référence* diotique étaient présentés aux participants dans un ordre aléatoire, séparés par une pause de 500 ms. La tâche des participants était d'indiquer lequel du premier ou du second son était le plus fort, indépendamment de toute autre différence perçue et sans autre possibilité de

réponse. Les comparaisons entre *sons test* diotiques et *sons de référence* diotiques (qui étaient les mêmes stimuli) ont été mises en place pour obtenir des données de contrôle permettant d'évaluer l'erreur liée à la procédure seule.

Les sons purs étaient délivrés aux oreilles des auditeurs par un casque sans autre apport d'information extérieure. Selon toute vraisemblance, les auditeurs effectuaient ainsi des comparaisons basées uniquement sur la sonie *aux oreilles* des stimuli. Les participants répondaient en cliquant à l'aide de la souris ou du clavier de l'ordinateur de test sur un bouton d'une interface graphique programmée sur Max/MSP. Ces instructions étaient indiquées oralement et sous forme écrite. Les participants réalisaient quelques essais avant de commencer l'expérience afin de se familiariser avec l'interface et la procédure. Les *sons test* étaient présentés à un niveau initial réglé à ± 10 dB du niveau de référence des sons diotiques (excepté pour les niveaux de références extrêmes, 30 et 90 phones, pour lesquels les niveaux initiaux des *sons test* étaient fixés à $+10$ et -10 dB, respectivement, afin de ne pas atteindre de niveaux trop faibles ou trop forts). À la suite de la réponse du participant pour un essai donné, le niveau du *son test* était diminué si le participant l'avait jugé plus fort que le *son de référence* et augmenté si le participant avait jugé la référence plus forte que le *son test*. Le pas d'ajustement était de ± 4 dB pour les premiers essais et diminuait à ± 1 dB après la deuxième inversion dans la courbe d'égalisation. La figure 2.1 représente un exemple de courbe d'égalisation, résultant de la séquence adaptative d'un *son test* donné. Après la huitième inversion dans la courbe d'égalisation, les valeurs du niveau du *son test* aux six dernières inversions étaient moyennées pour donner le point d'égalité subjective (PSE, pour *Point of Subjective Equality*) représentant le niveau auquel le *son test* paraissait aussi fort que le *son de référence*.

L'expérience était partagée en deux sessions séparées par une pause de cinq minutes. Afin de ne pas présenter consécutivement des sons de niveaux trop éloignés, l'une de ces sessions était consacrée aux sons à égaliser sur les *sons de référence* de niveaux 30, 40 et 50 phones tandis que l'autre comportait les *sons de référence* de niveaux 60, 70, 80 et 90 phones. Chaque participant a pris part aux deux sessions dans un ordre aléatoire. Les séquences adaptatives, correspondant à chaque suite d'essais menant à la courbe d'égalisation d'un *son test*, étaient entrelacées de sorte que deux essais consécutifs n'appartenaient pas à une même séquence. Par exemple, l'essai suivant le premier essai de la séquence du *son test* 200 Hz/90 phones/772 µs n'était pas le second essai de cette même séquence, mais pouvait par exemple être le premier essai de la séquence du *son test* 100 Hz/70 phones/0 µs. De cette manière, les participants ne pouvaient pas identifier les *sons de référence* (qui restaient inchangés tout au long des séquences adaptatives). La session apparaissait donc aux participants comme une suite de comparaisons par paire sans relation les unes avec les autres.

Les participants étaient assis dans une cabine audiométrique et avaient pour consigne de prendre leur temps pour placer confortablement le casque sur leur tête afin de ne pas avoir à y toucher pendant toute la durée de la session, puisque les changements de position de casque audio sont susceptibles d'affecter les stimuli perçus de façon audible (Paquier et Koehl, 2015).

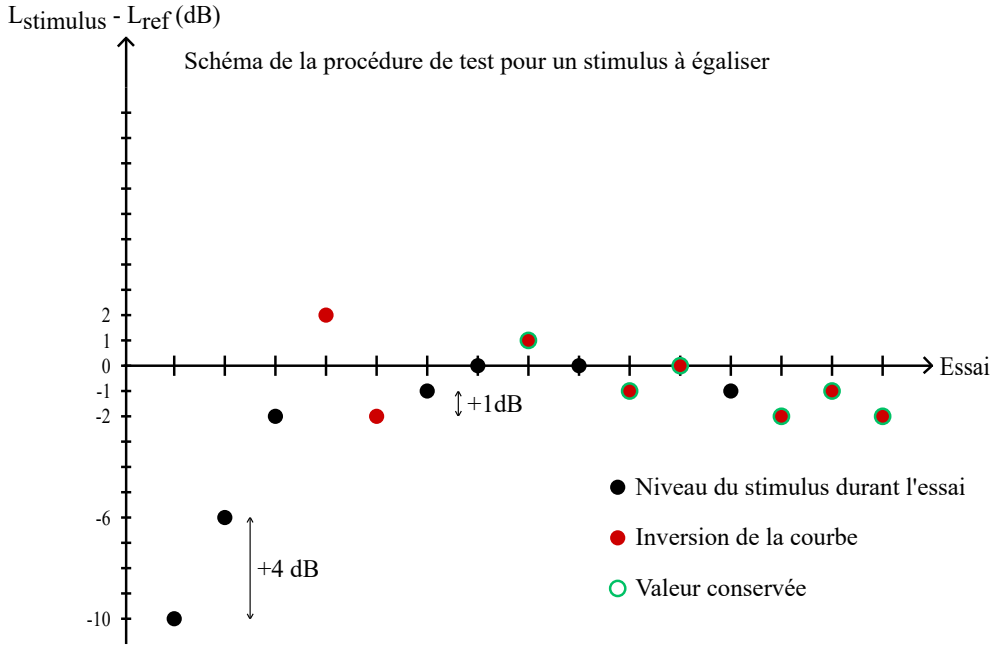


FIGURE 2.1 – Courbe d'égalisation pour un *son test* dans une procédure 2I2AFC. Le *son test* démarre ici à un niveau de −10 dB par rapport au niveau de référence (0 dB). Le participant a jugé, lors du premier essai, que le *son de référence* était plus fort. Le niveau du *son test* lors du deuxième essai est donc rehaussé de 4 dB, atteignant ainsi un niveau de −6 dB par rapport au niveau du *son de référence*. Le participant prend ensuite la même décision jusqu'au quatrième essai, lors duquel il juge le *son test* comme étant plus fort que le *son de référence*, formant la première inversion dans la courbe d'égalisation (point rouge dans la figure). Après la deuxième inversion dans la courbe d'égalisation (survenant au cinquième essai dans cet exemple), le pas d'ajustement du *son test* est réduit à ±1 dB. Après la huitième inversion (ici au quinzième essai), la séquence est terminée et le point d'égalité subjective vaut la moyenne des niveaux du *son test* aux six dernières inversions (cerclées en vert dans la figure).

L'expérience durait approximativement une heure pour chaque participant (20 min pour la session à faibles niveaux et 30 min pour la session à forts niveaux, plus les temps consacrés aux instructions, aux essais pré-test et à la pause). La génération des sons et l'application des gains suivant la procédure adaptative était réalisée par un programme Max/MSP.

2.3 Résultats & Discussion

Pour des raisons de clarté, les résultats de cette partie seront présentés sous forme de la différence entre le niveau du *son test* et celui du *son de référence* au PSE. Ainsi, si un *son test* a été égalisé sur un *son de référence* de niveau 53.41 dB SPL et que le PSE obtenu est de 51.41 dB SPL, les résultats présenteront le PSE sous la forme de la différence de ces deux niveaux, soit −2 dB. Dorénavant, le terme PSE désignera cette différence. Les PSE tels qu'ils ont été obtenus signifient que pour un PSE donné, admettons −2 dB, le *son test* était perçu aussi fort que le *son de référence* lorsque son niveau en était inférieur de 2 dB. En d'autres mots,

cela revient à dire que le *son test* aurait était perçu 2 dB plus fort que le *son de référence* si les deux sons avaient été présentés au même niveau. Pour cette raison, les PSE seront représentés en inversant l'axe des ordonnées pour la suite de ce document.

Le protocole expérimental de cette étude comportait une comparaison de *sons test* diotiques sur des références diotiques, revenant à égaliser les *sons de référence* sur eux-même. Un t-test avec correction de Bonferroni a révélé qu'aucun des PSE obtenus pour les *sons test* diotiques n'était significativement différent de 0, n'indiquant pas de biais expérimental.

Les PSE des signaux dichotiques obtenus dans les deux sessions ont été soumis à une analyse de variance à mesures répétées (RM ANOVA) comportant deux facteurs :

- Fréquence (deux niveaux) : 100 Hz, 200 Hz.
- Niveau sonore (sept niveaux) : 30 phones, 40 phones, 50 phones, 60 phones, 70 phones, 80 phones, 90 phones.

Les résidus du modèle linéaire étaient normalement distribués. L'analyse a révélé un effet significatif de la fréquence et du niveau sans révéler d'interaction significative entre ces deux facteurs (voir résultats tableau 2.2).

Source	**SS**	**DF**	**MS**	**F**	**Sig.p**
Fréquence	51.685	1	51.685	10.603	**.004**
Niveau	157.583	3.927*	40.133	9.643	**< .001**
Fréquence × Niveau	10.598	6	1.766	.630	.706

* Le degré de liberté a été ajusté par une correction de Greenhouse-Geisser suite à une violation de l'hypothèse de sphéricité des résultats.

Table 2.2 – Résultats de l'ANOVA conduite sur les PSE obtenus.

L'effet simple de la fréquence, représenté figure 2.2 (a), montre que l'effet de l'ITD sur la sonie est significativement plus prononcé à 200 Hz qu'à 100 Hz.

L'effet simple du niveau, représenté figure 2.2 (b), montre que l'augmentation de la sonie avec l'ITD décroît progressivement à mesure que le niveau augmente.

Des t-tests avec correction de Bonferroni ont été réalisés sur les PSE obtenus à chacune des deux fréquences afin d'évaluer à quels niveaux l'ITD a significativement augmenté la sonie. Ces tests ont indiqué que les PSE étaient significativement inférieurs à zéro pour les stimuli de niveau 40 phones à 100 Hz ($p = .047$) et pour les stimuli de niveaux 30 phones ($p < .001$), 40 phones ($p < .001$), 50 phones ($p < .001$) et 60 phones ($p = .045$) à 200 Hz. Ces PSE sont représentés figure 2.3.

Les résultats de cette expérience sont en accord avec ceux obtenus par les précédentes études liées à ce phénomène (Koehl et Paquier, 2015; Koehl *et al.*, 2015, 2017) et montrent une augmentation de la sonie de sons purs basses fréquences par l'ajout d'une ITD de 772 µs lorsqu'ils sont diffusés à faible niveau. Cet apport en sonie décroît à mesure que le niveau de diffusion augmente, allant dans le sens de l'hypothèse avancée par Koehl et Paquier (2015) selon laquelle cet effet

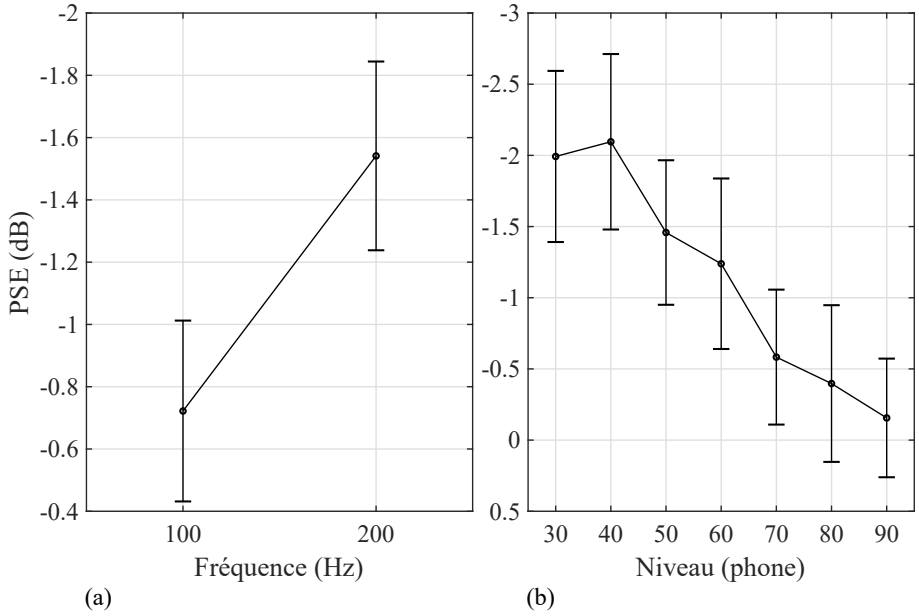


FIGURE 2.2 – PSE en fonction de la fréquence (a) et du niveau (b) des stimuli, avec intervalles de confiance à 95%. Les axes des ordonnées sont inversés.

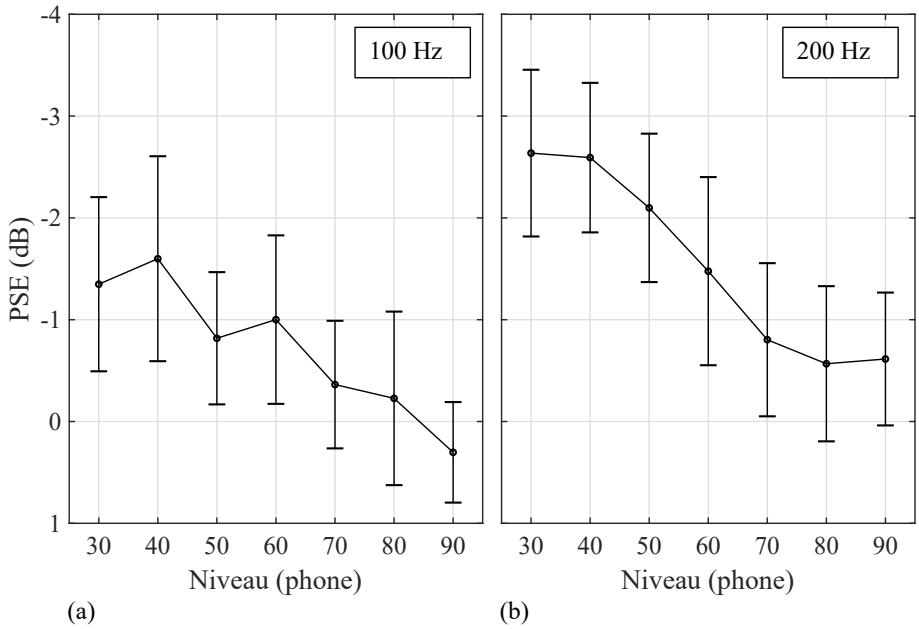


FIGURE 2.3 – Évolution des PSE en fonction du niveau pour les stimuli de fréquence 100 Hz (a) et 200 Hz (b), avec intervalles de confiance à 95%. Les axes des ordonnées sont inversés.

pourrait être dû au bruit interne inhérent au système auditif. Cette hypothèse a été formulée en raison des similarités entre l'effet de l'ITD sur la sonie et la dépendance de la sonie vis-à-vis de la corrélation interaurale du signal en présence de bruit masquant (voir section 1.4.1). L'augmentation de sonie observée dans cette étude reste cependant significative jusqu'à des niveaux

largement supraliminaires (donc largement supérieurs au niveau du bruit interne), les seuils d'audition se trouvant à 100 Hz et 200 Hz à 26.5 dB SPL et 14.4 dB SPL, respectivement, d'après la norme ISO 389-7 (2005) donnant les seuils absolus d'audition en champ libre.

Chapitre 3

Influence d'un bruit externe sur l'effet de l'ITD sur la sonie

Cette étude a été présentée à la conférence *Acoustics '17 Boston* (Berthomieu *et al.*, 2017).

3.1 Introduction

Koehl et Paquier (2015) ont formulé une hypothèse selon laquelle l'augmentation de sonie avec l'ITD pourrait être due au bruit interne inhérent au système auditif. D'après Diercks et Jeffress (1962), le seuil absolu d'audition correspondrait au seuil de masquage dans le bruit interne. Lorsqu'un son pur est diffusé à faible niveau, l'ajout d'un décalage temporel entre les deux oreilles pourrait aider le système auditif à extraire ce son pur du bruit interne, aboutissant à une augmentation de la sonie pour ce signal. Cette augmentation aurait alors lieu à condition que le rapport signal-à-bruit (SNR) ne soit pas trop élevé. Si tel était le cas, le système auditif parviendrait déjà à extraire convenablement le signal diotique du bruit. L'ajout d'une ITD ne permettrait alors pas d'améliorer cette extraction et n'augmenterait pas la sonie du signal. Cette condition pourrait expliquer l'absence d'effet de l'ITD sur la sonie de sons purs diffusés à forts niveaux décrite dans le chapitre 2 et en accord avec les résultats de Koehl et Paquier (2015).

Zwicker et Henning (1991) ont observé une augmentation de sonie due à l'IPD pour des sons purs de fréquence 250 Hz jusqu'à 30 dB SL voire 40 dB SL lorsque les sons étaient restitués en présence d'un bruit diotique, formant une condition N_0S_π (voir section 1.4.1 pour une explication de cette nomenclature). Au-delà de tels SNR (définis par les niveaux de 30 dB SL et 40 dB SL), la sonie des sons diotiques n'est vraisemblablement plus affectée par le bruit masquant. L'IPD n'apporte alors plus de bénéfice dans la perception du signal et n'en augmente plus la sonie.

Dans l'étude présentée dans le chapitre 2, les conditions n'étaient pas identiques à celles de Zwicker et Henning (1991) puisque le bruit interne n'est vraisemblablement pas diotique et qu'un décalage de 772 µs à 200 Hz représente un déphasage d'environ $\frac{\pi}{3}$. Une augmentation significative de la sonie de sons purs comportant une ITD a été observée à des niveaux allant jusqu'à 60 phones, soit un niveau largement supraliminaire.

L'objectif de la présente étude était d'observer le lien entre démasquage du signal dans le bruit et augmentation de la sonie du signal avec son ITD. Pour cela, le SNR était réduit par l'ajout d'un bruit externe.

En présence d'un bruit externe, la corrélation interaurale du bruit total (composé des bruits interne et externe) est fonction de la corrélation interaurale de chacun des deux bruits ainsi que de leur niveau relatif (voir équation 1.16). L'ajout d'un bruit externe diotique aura donc pour effet d'une part d'augmenter le niveau global de bruit et d'autre part d'augmenter la corrélation interaurale de l'ensemble dans l'hypothèse où la corrélation ρ_I du bruit interne est inférieure à 1. Si le gain en sonie dû à l'ITD était expliqué par la présence du bruit interne, l'augmentation du niveau du bruit global par l'ajout d'un bruit externe provoquerait une augmentation de l'effet de l'ITD sur la sonie. La procédure expérimentale était la même que celle utilisée lors de l'expérience précédente. Le test était découpée en deux sessions, chacune correspondant à un niveau de bruit externe ajouté.

3.2 Matériel et méthode

Le dispositif expérimental de la présente étude était similaire à celui de l'expérience précédente, la tâche des participants étant toujours d'égaliser la sonie de sons purs diotiques et dichotiques (comportant une ITD fixe) sur la sonie de *sons de référence* diotiques de même fréquence. Cependant, un bruit externe était restitué aux participants tout au long de l'expérience. Le niveau de ce bruit était fixé tout au long d'une session et variait d'une session à l'autre. Le niveau des stimuli à égaliser variait au cours de chacune des sessions, aboutissant à une multitude de SNR différents.

3.2.1 Stimuli

Puisque l'expérience précédente a relevé une augmentation de sonie due à l'ITD plus importante à 200 Hz qu'à 100 Hz, la présente expérience s'est exclusivement consacrée à des sons purs de fréquence 200 Hz. La largeur de bande du bruit externe a été fixée à une ERB (Edmonds et Culling, 2009) centrée sur 200 Hz, soit une largeur de bande de 46 Hz, de sorte que toute l'énergie du bruit soit contenue dans la même bande critique que le signal (voir section 1.3.2). Ce bruit a été généré avec une durée de 10 min sur MATLAB et était ensuite bouclé et diffusé en temps réel sur Max/MSP pendant l'expérience. Le bruit était diffusé à des niveaux de 10 phones ou 20 phones selon la session afin de générer différents SNR. A 200 Hz, les niveaux de pression sonore évoquant un niveau d'isosonie de 10 phones et 20 phones sont respectivement 24.8 dB SPL et 35.5 dB SPL d'après la norme ISO 226 (2003). L'effet de l'ITD sur la sonie était, d'après l'étude précédente, significatif jusqu'à 60 phones. Dans la présente étude, où les signaux pouvaient être présentés dans un bruit externe de niveau 20 phones, l'effet de l'ITD sur la sonie a été étudié jusqu'à 80 phones. Les niveaux des *sons de référence* étaient fixés à 30, 40, 50, 60, 70 et 80 phones. Le SNR variait alors d'environ 9 dB (entre le son pur de niveau 30 phones et le bruit de 20 phones) à environ 61 dB (entre le son pur de niveau 80 phones et le bruit de niveau

10 phones) d'après la norme ISO 226 (2003) définissant les lignes isosoniques. Yost (1988) a estimé le niveau du bruit interne dans la bande de tiers d'octave centrée sur 200 Hz en observant l'évolution des seuils de masquage en fonction de la corrélation interaurale (MLD, pour *Masking Level Difference*, voir section 1.4.1) d'un sinus de 200 Hz restitué à différents niveaux sonores. Les estimations de l'auteur font état d'un bruit interne de niveau 29 dB SPL, compris entre les deux niveaux de bruit externe ajoutés lors de cette expérience.

3.2.2 Procédure

Vingt-deux participants (six femmes et seize hommes âgés de vingt à vingt-cinq ans) ont pris part à l'expérience. Leurs seuils d'audition ont été mesurés d'après la procédure ANSI S3.21 (2004) en utilisant le bruit à bande étroite constituant le bruit externe de l'expérience. Ces seuils ont été mesurés pour vérifier que les auditeurs entendaient le bruit externe ajouté lors de l'expérience. Pour chaque participant, le seuil mesuré était inférieur à 10 dB HL.

La procédure utilisée dans cette expérience était la même que celle de l'expérience précédente. Les deux sessions (l'une correspondant à l'ajout d'un bruit externe à un niveau de 10 phones et l'autre d'un bruit externe à un niveau de 20 phones) étaient ordonnées aléatoirement pour chacun des sujets, de même que les douze combinaisons de conditions expérimentales contenues dans chacune des deux sessions (correspondant à chaque combinaison des deux valeurs d'ITD et des six niveaux de diffusion des signaux sonores). Hellman *et al.* (1997) ont observé qu'en basses fréquences, la sonie de sons de longue durée décroissait rapidement dans les trois premières minutes de présentation et continuait de décroître plus doucement ensuite. Dans la présente expérience, le bruit externe ajouté avait pour objectif de réhausser le niveau sonore de bruit total et ainsi de se mélanger au bruit interne. Ce bruit externe était alors envoyé au casque avant même que les participants ne le placent sur leurs oreilles de façon à ce qu'une certaine forme d'adaptation auditive soit déjà présente au début de l'expérience. Les participants n'étaient pas informés explicitement de la présence d'un bruit externe diffusé dans le casque, et certains ont reporté à la fin de l'expérience ne pas s'être rendu compte de la présence de ce bruit dans la session où il était présenté à un niveau de 10 phones (alors qu'ils étaient en capacité de le percevoir d'après les mesures de seuil effectuées en début de session). Les instructions étaient données une fois que les auditeurs avaient le casque sur leurs oreilles, sous formes orale et écrite. Ensuite, les participants effectuaient un pré-test de six comparaisons incluant chacun des six niveaux de référence présentés dans un ordre aléatoire de façon à se familiariser avec les stimuli et la procédure. Pour tous les niveaux de référence supérieurs à 30 phones, les niveaux de départ des *sons test* étaient aléatoirement fixés à ± 10 dB de leurs niveaux de référence. Les stimuli à égaliser sur un *son de référence* de niveau 30 phones démarraient systématiquement à un niveau de $+10$ dB afin de ne pas faire intervenir de niveaux trop faibles, susceptibles d'être masqués par le bruit externe. Le test durait approximativement une heure, incluant deux sessions d'une vingtaine de minutes, deux pré-tests (un au début de chaque session) ainsi que les temps nécessaires aux instructions et à une pause de cinq minutes effectuée entre les deux sessions.

3.3 Résultats & Discussion

Comme dans la section 2.3, un PSE négatif signifie qu'à niveau égal, le *son test* était perçu plus fort que le *son de référence*. Les PSE seront donc représentés en inversant l'axe des ordonnées.

De même que lors de l'expérience précédente, une partie de cette étude consistait à comparer des *sons test* diotiques sur des *sons de référence* également diotiques, revenant à égaliser ces sons sur eux-même. Des t-tests avec correction de Bonferroni ont révélé que dans l'une des conditions (lorsque le stimulus à égaliser sur un niveau de référence de 30 phones était présenté dans du bruit de niveau 10 phones) le PSE était significativement supérieur à zéro ($t(21) = 3.850, p = .011$) avec une moyenne de 2.4 dB. Ce résultat semble témoigner d'une erreur systématique due au protocole (puisque le stimulus de référence et le stimulus à égaliser étaient identiques) et pourrait provenir de l'asymétrie du niveau de départ, fixé systématiquement à +10 dB pour ce stimulus dont le niveau de référence était 30 phones (voir section 3.2). La figure 3.1 représente les PSE obtenus en présence du bruit externe de niveau 10 phones pour les stimuli diotiques (a) et dichotiques (b). Le décalage du PSE obtenu à 30 phones lors de l'égalisation du *son test* diotique est visible dans la partie (a), et semble également être présent sur le *son test* dichotique (b), pour lequel le niveau de départ était également fixé à +10 dB. Si l'on considérait que l'erreur liée au protocole pour les *sons test* dichotiques était la même que celle mesurée sur les *sons test* diotiques (soit un décalage du PSE de 2.4 dB), il serait possible de compenser cette erreur (en ajoutant 2.4 dB). Cependant, rien ne permet d'affirmer que l'erreur liée au protocole expérimental était identique pour les *sons test* diotiques et dichotiques et effectuer une correction pourrait être source d'erreur supplémentaire. De plus, cette erreur pour le *son test* de niveau 30 phones n'a pas été observée lorsque le bruit externe était présenté à un niveau de 20 phones (alors que le niveau de départ était également fixé à +10 dB). En conséquence, le choix a été fait de ne pas prendre en compte les résultats obtenus au niveau de référence 30 phones dans la suite de l'analyse pour les deux niveaux de bruit.

Les PSE ont ensuite été soumis à une ANOVA à mesures répétées comportant deux facteurs :

- Niveau sonore du bruit externe (deux niveaux) : 10 phones, 20 phones.
- Niveau sonore du stimulus (cinq niveaux) : 40 phones, 50 phones, 60 phones, 70 phones, 80 phones.

Les résidus du modèle linéaire étaient normalement distribués. Les résultats de l'analyse sont présentés dans le tableau 3.1. Ces résultats indiquent un effet simple significatif du niveau du stimulus, sans révéler d'effet significatif du niveau de bruit externe ou d'interaction entre niveau du bruit externe et niveau du stimulus. La figure 3.2 représente cet effet simple du niveau du stimulus. Les résultats de cette analyse révèlent que l'effet de l'ITD sur la sonie était similaire selon que le bruit externe était présenté à un niveau de 10 phones ou 20 phones.

Afin d'évaluer l'apport du bruit externe sur l'effet de l'ITD, les PSE obtenus dans la présente étude ont été comparés à ceux obtenus dans le silence lors de l'expérience précédente, réalisée par un groupe de participants différent. Une ANOVA mixte a été effectuée. Cette analyse prenait

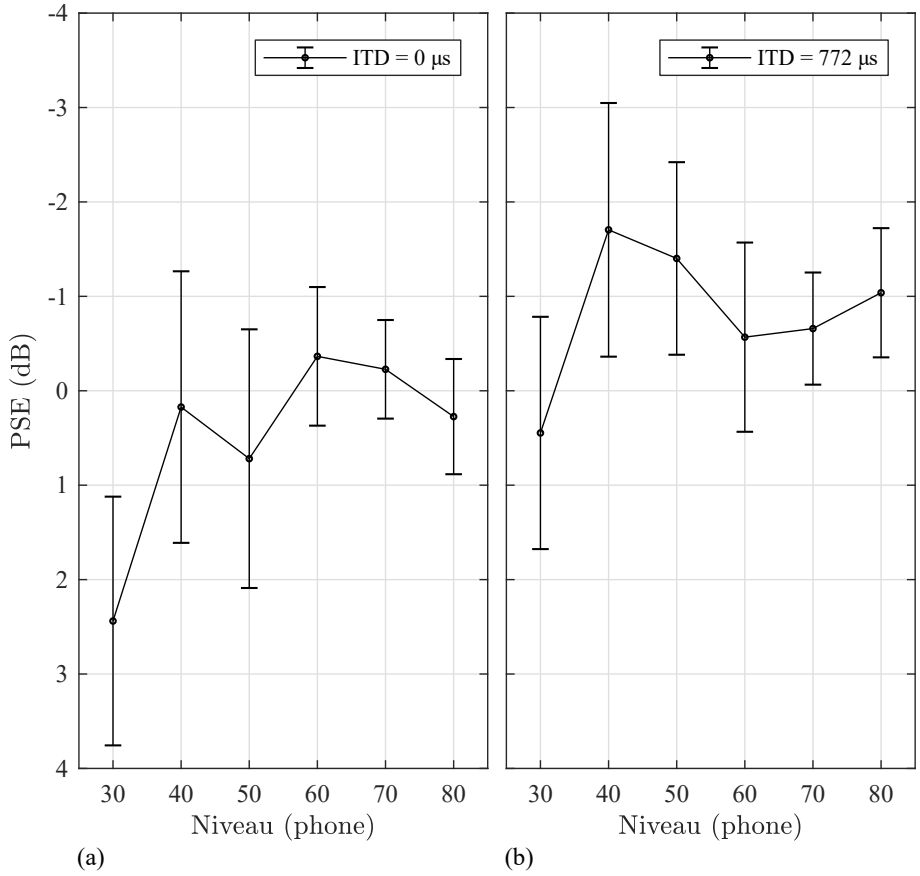


FIGURE 3.1 – PSE en fonction du niveau, avec intervalles de confiance à 95%, montrant l'erreur observée sur l'égalisation du *son test* diotique (a) à un niveau de 30 phones lorsqu'un bruit externe de niveau 10 phones était présenté aux participants. La partie (b) représente les PSE des *sons test* dichotiques en fonction du niveau lorsqu'ils étaient présentés dans un bruit externe de niveau 10 phones. Les axes des ordonnées sont inversés.

Source	**SS**	**DF**	**MS**	**F**	**Sig.p**
S	63.281	2.755*	22.973	2.974	**.043**
B	.603	1	.603	.127	.725
S × B	3.035	2.68*	1.132	.198	.878

* Le degré de liberté a été ajusté par une correction de Greenhouse-Geisser suite à une violation de l'hypothèse de sphéricité des résultats.

TABLE 3.1 – Résultats de l'ANOVA conduite sur les PSE. S désigne le niveau du stimulus et B le niveau de bruit externe ajouté.

comme unique facteur intra-sujet le niveau du stimulus et comme facteur inter-sujets la présence ou non de bruit externe :

- Niveau du stimulus (cinq niveaux, facteur intra-sujets) : 40 phones, 50 phones, 60 phones, 70 phones, 80 phones.
- Bruit externe (deux niveaux, facteur inter-sujets) : avec, sans.

Les résidus du modèle linéaire étaient normalement distribués. Les résultats de cette analyse

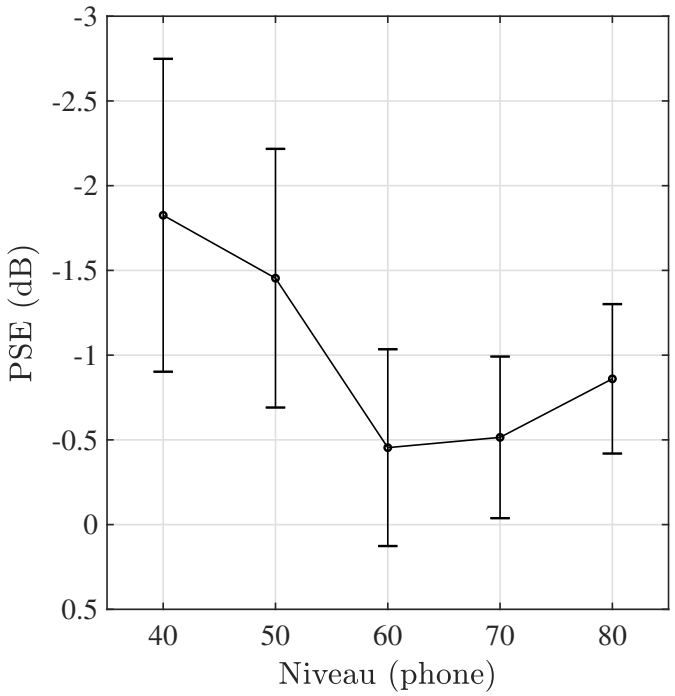


FIGURE 3.2 – PSE en fonction du niveau des stimuli comportant une ITD de 772 µs diffusés dans des bruits de niveaux 10 phones et 20 phones (les résultats sont moyennés sur les deux niveaux de bruit), avec intervalles de confiance à 95%. L'axe des ordonnées est inversé.

sont présentés dans le tableau 3.2. Ces résultats **n'indiquent pas d'effet significatif impliquant la présence ou l'absence de bruit externe ni d'interaction entre niveau du stimulus et présence de bruit externe**.

L'ANOVA a révélé un effet simple significatif du niveau du stimulus ($p < .001$), représenté figure 3.3. Cette analyse a été suivie de tests post-hoc LSD de Fisher avec correction de Bonferroni indiquant que les PSE obtenus aux niveaux 40 et 50 phones n'étaient pas significativement différents, de même que ceux obtenus aux niveaux 60, 70 et 80 phones. Parmi ces deux groupes, seule la comparaison entre les PSE obtenus aux niveaux 50 et 60 phones n'a pas révélé de différence significative ($p = .193$). Enfin, des t-tests avec correction de Bonferroni ont été réalisés sur ces résultats. Ces t-tests indiquent que tous les PSE obtenus de cette manière étaient significativement différents de zéro (40 phones $t(65) = -6.337$, $p < .001$, 50 phones $t(65) = -5.983$, $p < .001$, 60 phones $t(65) = -3.208$, $p = .010$, 70 phones $t(65) = -3.092$, $p = .015$, 80 phones $t(65) = -4.028$, $p < .001$), témoignant d'une augmentation de la sonie de sons purs basses fréquences par l'ajout d'une ITD de 772 µs en présence ou en absence de bruit externe.

En résumé, les résultats de cette expérience indiquent :

- Que l'ITD affecte la sonie de sons purs à 200 Hz de la même manière lorsque les stimuli sont présentés avec un bruit externe de niveau 10 phone ou de niveau 20 phones.

Source	SS	DF	MS	F	Sig.p
S	112.243	3.276*	34.259	7.248	**< .001**
B	17.315	1	17.315	3.1	.083
S × B	15.222	4	3.806	.983	.417

* Le degré de liberté a été ajusté par une correction de Greenhouse-Geisser suite à une violation de l'hypothèse de sphéricité des résultats.

Table 3.2 – Résultats de l'ANOVA mixte à mesures répétées conduite sur les PSE. S désigne le niveau du stimulus (facteur intra-sujets) et B la présence ou l'absence d'un bruit externe (facteur inter-sujets).

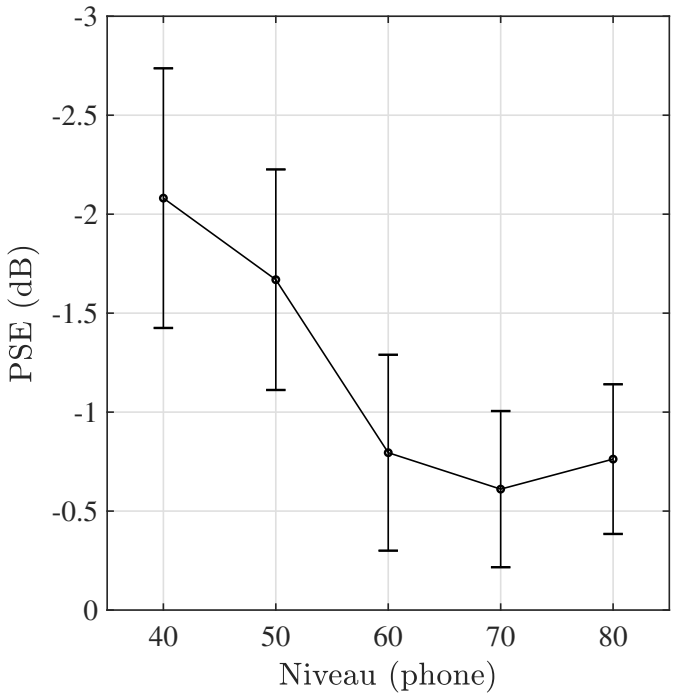


Figure 3.3 – PSE en fonction du niveau des stimuli pour toutes conditions de bruit externe (absent, présent) confondues, avec intervalles de confiance à 95%. Les axes des ordonnées sont inversés.

- Que l'augmentation de sonie due à l'ITD en présence de bruit externe diffusé à faible niveau (10 phones ou 20 phones) n'est pas significativement différente de celle obtenue dans le silence.
- Que la sonie de sons purs basses fréquences est significativement plus élevée avec ITD que sans ITD pour tous les niveaux testés (de 40 phones à 80 phones) et pour toutes conditions de bruit externe confondues.
- Que l'augmentation de sonie avec l'ITD est plus importante à faibles niveaux (de 40 phones à 50 phones) qu'à forts niveaux (de 60 phones à 80 phones).

D'après les résultats de cette expérience, l'effet de l'ITD sur la sonie de sons purs basses fréquences semble identique quel que soit le SNR. Ce résultat peut paraître étonnant lorsqu'il est comparé aux augmentations de sonie habituellement observées pour des signaux dichotiques

en présence de bruit masquant telles qu'observées par Zwicker et Henning (1991) à des niveaux atteignant 30-40 dB SL. Cependant, de telles augmentations de sonie ont été observées dans des conditions N_0S_π, où l'IPD était de π et donc supérieure à l'IPD produite par une ITD de 772 µs sur un son pur de fréquence 200 Hz (laquelle vaut environ $\frac{\pi}{3}$), expliquant possiblement une augmentation de sonie moins prononcée et s'estompant plus rapidement à niveau supraliminaire.

L'influence de la corrélation interaurale du bruit externe sur les seuils de masquage décroît aux niveaux faibles. D'après Yost (1988), cette observation s'explique par le fait que le niveau du bruit externe masquant devient comparable à celui du bruit interne, entraînant un bruit total dont la corrélation interaurale est inférieure à celle du bruit externe seul. D'après l'évolution des seuils de masquage sur un sinus de 200 Hz, l'auteur a estimé que le niveau du bruit interne était de 29 dB SPL dans la bande de tiers d'octave centrée sur 200 Hz.

Dans la présente expérience, le bruit externe était présenté soit à un niveau de 24.8 dB SPL (inférieur au niveau du bruit interne estimé par Yost (1988)), soit à un niveau de 35.5 dB SPL (supérieur au niveau du bruit interne estimé par Yost (1988)). Le niveau et la corrélation interaurale du bruit total (composé du bruit interne et du bruit externe) dépendait donc du niveau auquel le bruit externe était présenté. L'absence d'influence significative du bruit externe sur les résultats ne permet donc pas de favoriser l'hypothèse impliquant le bruit interne dans l'augmentation de sonie des sons purs basses fréquences avec l'ITD.

Le recours à l'électroencéphalographie (EEG) permettrait de comparer les potentiels évoqués lors de la présentation de sons comportant ou non une ITD afin de comprendre à quel stade de la chaîne auditive l'ITD provoque une augmentation de la sonie. La fréquence de décharge de neurones présents dans plusieurs noyaux du tronc cérébral est par exemple dépendante de l'ITD chez les mammifères (Grothe *et al.*, 2010; Benichoux *et al.*, 2016). Finlayson et Caspary (1991) ont également repéré chez le chinchilla une inhibition controlatérale dépendante de l'IPD en basses fréquences. L'ITD pourrait de la même manière modifier l'inhibition controlatérale chez l'homme (Koehl *et al.*, 2017).

La présente étude a néanmoins permis d'observer un effet significatif de l'ITD sur la sonie sur une plus large gamme de niveaux. Puisque les résultats obtenus pour différentes conditions de bruit externe étaient similaires, ils ont été regroupés. L'augmentation de sonie provoquée par une ITD de 772 µs a alors été révélée de façon significative sur des stimuli restitués jusqu'à un niveau de 80 phones. La taille de cet effet n'est cependant pas constante avec le niveau. Les PSE mesurés sur des stimuli de faible niveau étaient d'environ 2 dB. Pour les stimuli de niveau supérieur à 50-60 phones, les PSE étaient inférieurs à 1 dB, devenant de l'ordre de grandeur des seuils différentiels d'intensité (Fletcher, 1953).

Le fait que les stimuli comportant une ITD de 772 µs soient perçus plus forts que leurs équivalents diotiques de la même manière dans le silence et dans du bruit permet également d'évoquer un premier signe de robustesse de l'effet de l'ITD sur la sonie aux conditions d'écoute. Les conditions dans lesquelles les sons sont présentés aux auditeurs sont susceptibles de modifier la façon dont ils sont perçus. En laboratoire, les signaux sont généralement contrôlés et isolés

pour être manipulés avec précision. Dans la vie de tous les jours, les signaux sonores proviennent de sources réelles, parfois visibles, et se propagent dans un environnement bruité. L'effet de l'ITD sur la sonie a été étudié uniquement sur des sons purs auxquels une ITD fixe était appliquée lors d'une diffusion au casque dans une cabine audiométrique. L'expérience suivante a pour objectif d'étudier l'effet de l'ITD sur la sonie de sons basses fréquences dans des conditions d'écoute plus naturelle, impliquant des sources réelles identifiables situées autour des auditeurs dans un environnement non-anéchoïque.

Chapitre 4

Sonie directionnelle de bruits à bande étroite en conditions d'écoute naturelle

Cette étude a été présentée au *14eCongrès Français d'Acoustique* (CFA '18) au Havre en 2018 et a été publiée dans le *Journal of the Audio Engineering Society* (Berthomieu *et al.*, 2019a).

4.1 Introduction

Effet de la validité écologique sur la sonie

Dans les études présentées dans les chapitres 2 et 3, les sons étaient restitués aux auditeurs via un casque dans une cabine audiométrique. Cette situation d'écoute est peu écologique et s'éloigne par de nombreux aspects de conditions d'écoute naturelle. D'après les travaux d'Epstein et Florentine (Epstein et Florentine, 2009, 2012), la validité écologique d'une expérience peut avoir un effet non négligeable sur la sommation de sonie binaurale et doit être prise en compte lors d'études portant sur la sonie. Cette observation a été déduite par les auteurs à partir d'études s'intéressant au rapport de sonie binaural-à-monaural (voir section 1.3.4). Ce rapport correspond à l'augmentation de la sonie lorsque l'on passe d'une présentation monaurale, sur une seule oreille, à une présentation binaurale diotique où le même son est présenté sur les deux oreilles. Les auteurs ont évalué ce rapport pour des sons diffusés au casque ou sur haut-parleur. Lorsque les sons étaient présentés au casque, un seul des deux canaux était alimenté pour évaluer la sonie monaurale. Lorsque les sons étaient diffusés sur haut-parleur, l'écoute monaurale était effectuée en bouchant une oreille des auditeurs. Les auteurs ont reporté un rapport de sonie binaural-à-monaural plus faible lorsque la validité écologique des conditions de présentation des sons augmentait, à savoir pour les sons restitués sur haut-parleur (Epstein et Florentine, 2009) avec indices visuels (Epstein et Florentine, 2012). Bien que le rapport de sonie binaural-à-monaural ne soit pas directement lié à la sonie directionnelle, tous deux font partie du processus de sommation binaurale et ces observations témoignent de l'existence de processus de haut niveau pouvant affecter la sonie. La sonie directionnelle pourrait alors dépendre de la validité écologique des conditions de présentation des sons. De plus, Sivonen (2007) a observé une sonie directionnelle plus faible pour des sons diffusés dans des espaces réverbérés qu'en champ anéchoïque. Les effets de salle sont donc également susceptibles de diminuer les effets directionnels sur la sonie tels qu'observés en cabine audiométrique ou en salle anéchoïque.

Ainsi, les études décrites dans les chapitres 2 et 3 pourraient avoir observé des effets surestimés par rapport aux effets présents dans des conditions d'écoute naturelle. L'effet de l'ITD sur la sonie pourrait même ne plus être vérifié dans des conditions plus réalistes, par exemple pour des sons diffusés par des sources réelles situées dans une salle. De plus, les stimuli directement synthétisés et restitués au casque sont souvent internalisés (perçus « à l'intérieur de la tête ») plutôt que comme provenant de sources extérieures et distantes (Blauert, 1997a). Cette internalisation pourrait également avoir des conséquences sur les phénomènes liés à la sonie directionnelle.

Objectif de l'étude

La présente étude a été mise en place dans le but d'évaluer la sonie directionnelle de bruits basses fréquences diffusés par des haut-parleurs situés autour des auditeurs. Contrairement aux précédentes études s'intéressant à l'effet de l'ITD sur la sonie, les stimuli étaient émis par des sources réelles dans une salle non-anéchoïque. Les stimuli étaient présentés aux auditeurs de deux manières différentes : soit en les restituant sur haut-parleurs, soit en les diffusant au casque en reproduisant virtuellement leur position. Les sources sonores étaient visibles uniquement lorsque les stimuli étaient restitués par les haut-parleurs.

4.2 Matériel et méthode

La sonie directionnelle était évaluée en égalisant la sonie de bruits émis par des sources latérales sur celle de bruits émis par des sources frontales. Puisque la réponse d'une salle dépend de la position de la source sonore, la sonie directionnelle était évaluée pour différentes dispositions des sources sonores dans la salle. Les sons utilisés dans cette expérience étaient des bruits à bande étroite, moins susceptibles d'exciter un mode de salle particulier que des sons purs. Ces bruits pouvaient être restitués :

- Directement par les haut-parleurs.
- Par un casque qui en reproduisait un enregistrement binaural.
- Par un casque qui reproduisait le bruit directement synthétisé, comportant une ITD fixe. Cette restitution avait pour objectif d'évaluer l'effet de l'ITD seule sur la sonie du bruit à bande étroite afin de le comparer à l'effet précédemment observé sur des sons purs.

Visibilité des sources sonores

D'après les études de Sivonen *et al.*, la sonie directionnelle de bruits à bande étroite était identique selon que les bruits étaient directement restitués par des haut-parleurs visibles situés autour des auditeurs dans une chambre anéchoïque (Sivonen et Ellermeier, 2006) ou qu'ils étaient reproduits au casque et virtuellement placés aux mêmes positions que celles des haut-parleurs par l'intermédiaire d'HRTF individualisées (Sivonen *et al.*, 2005). Ces deux études étaient réalisées dans la même chambre anéchoïque et des haut-parleurs étaient placés dans les directions de provenance des sons même lorsque les sons étaient reproduits au casque. Dans ces conditions, les auditeurs ne parviennent pas forcément à identifier si les sons qu'ils entendent sont diffusés

par le casque ou par les haut-parleurs visibles (Langendijk et Bronkhorst, 2000). Cette confusion peut être due à des interactions audiovisuelles, où la localisation auditive des sons joués au casque est attirée par la perception visuelle du haut-parleur (voir section 1.2.5). Lorsque les sons sont restitués au casque dans le cadre d'études en laboratoire, ils ne sont généralement pas accompagnés d'une représentation visuelle de source sonore. Les études menées sur les effets de l'ITD sur la sonie (Koehl et Paquier (2015); Koehl *et al.* (2015, 2017), ainsi que les études présentées dans les chapitres 2 et 3) en sont des exemples. Puisque la présente étude a pour objectif de comparer l'effet de l'ITD dans ces conditions de laboratoire (sans représentation des sources sonores) à des conditions plus naturelles, les sons reproduits au casque n'étaient pas accompagnés d'une représentation visuelle des sources sonores. Ainsi, les auditeurs ne disposaient d'informations visuelles sur les sources sonores que lorsque les sons étaient restitués sur haut-parleurs.

4.2.1 Stimulus

Le stimulus utilisé lors de cette expérience était un bruit large d'1 ERB. La fréquence centrale de ce bruit était de 265 Hz puisque les précédentes expériences ont relevé un effet significatif de l'ITD sur la sonie entre 200 et 400 Hz (Koehl et Paquier (2015), expérience reportée dans le chapitre 2). L'ERB a été calculée d'après l'équation 1.7 de Glasberg et Moore (1990) :

$$\mathrm{ERB} = 24.7 \cdot (4.37 \cdot 10^{-3} \cdot f + 1) \tag{4.1}$$

où $f = 265\,\mathrm{Hz}$ est la fréquence centrale de l'ERB. Cette équation mène à une ERB de 53 Hz. Le bruit à bande étroite a été généré suivant la procédure décrite par Edmonds et Culling (2009). Le niveau de référence du bruit était fixé à 50 phones, niveau auquel l'effet de l'ITD est particulièrement prononcé (voir expériences des chapitres 2 et 3). Ce niveau d'isosonie correspond à un niveau de pression sonore de 58.4 dB SPL à 265 Hz au centre de la tête de l'auditeur en son absence d'après la norme ISO 226 (2003). Les stimuli duraient 1.6 s et comportaient des rampes d'attaque et de relâchement de 100 ms.

4.2.2 Procédure

Vingt auditeurs (quatre femmes et seize hommes, agés de vingt à vingt-cinq ans) n'ayant pas de problème d'audition ont pris part à l'expérience et étaient rémunérés pour leur participation. Aucun d'eux n'avait d'expérience en test d'écoute de laboratoire.

Le PSE était obtenu pour chaque stimulus à l'aide d'une procédure adaptative à choix forcé à deux intervalles et deux alternatives (2I2AFC) suivant une règle « 1-up 1-down » (Levitt, 1971), identique à la procédure utilisée lors des expériences décrites dans les chapitres 2 et 3. À chaque essai, un *son test* et un *son de référence* étaient présentés dans un ordre aléatoire, séparés par une pause de 500 ms. Le *son test* était émis par une source latérale (d'azimut −90° ou +90°) et le *son de référence* était émis par une source frontale (d'azimut 0°). Les participants devaient indiquer lequel du premier ou du second son leur semblait être le plus fort, indépendamment de

toute autre différence perçue et sans autre possibilité de réponse. La procédure était exécutée par un programme Max/MSP et les participants répondaient à l'aide de deux boutons sur une manette. Les instructions étaient indiquées oralement et quelques essais étaient effectués avant le début du test dans le but de familiariser les participants à l'interface et à la procédure. Des *sons test* frontaux (d'azimut 0°) étaient également égalisés sur les *sons de référence* frontaux dans chaque configuration spatiale des sources sonores afin de repérer d'éventuels biais systématiques liés à la procédure d'égalisation.

Les *sons test* étaient initialement présentés à un niveau aléatoirement fixé à ±10 dB du niveau de référence (58.4 dB SPL, soit un niveau d'isosonie de 50 phones), puis ce niveau était incrémenté ou décrémenté d'une certaine quantité selon les réponses des participants. Cette quantité était de ±4 dB au début de la courbe d'égalisation puis était réduite à ±1 dB après deux inversions dans la courbe d'égalisation. Après huit inversions dans la courbe d'égalisation, le PSE était dérivé depuis la moyenne du niveau du *son test* aux six dernières inversions (voir figure 2.1).

L'expérience était découpée en deux sessions séparées d'une pause de cinq minutes. Dans l'une de ces deux sessions, les sons étaient réellement joués par des haut-parleurs situés autour des auditeurs. Dans l'autre session, les sons étaient virtuellement reproduits au casque. Chaque participant effectuait les deux sessions dans un ordre aléatoire.

4.2.3 Présentation réelle

L'expérience se déroulait dans une salle d'écoute acoustiquement traitée dont le TR_{60} était de 0.3 s à 250 Hz, en accord avec les recommandations, à cette fréquence, de la norme IEC 60268-13 (1998) concernant les caractéristiques de salles utilisées dans des tests d'écoute sur haut-parleurs. La fréquence de coupure de cette salle était de 130 Hz, inférieure à tout le contenu fréquentiel des stimuli. Quatre haut-parleurs étaient disposés dans la salle (voir la figure 4.1). Ce dispositif permettait, comme le montre la figure 4.2, de placer les haut-parleurs frontaux et latéraux dans quatre configurations spatiales différentes dans la salle de test afin d'en exciter différemment la réponse. Chaque configuration spatiale était réalisée en orientant l'auditeur vers un haut-parleur particulier. Dans une configuration spatiale donnée, le haut-parleur frontal (0°) diffusant le *son de référence*, dénoté R, était positionné à 1 ou 2 m de l'auditeur. Les sources latérales (−90° et +90°) étaient situées l'une à 1 m et l'autre à 2 m, jouant les *sons test* dénotés T. Comme lors des expériences précédentes, un *son test* identique au *son de référence* était égalisé pour repérer d'éventuelles erreurs systématiques liées au protocole expérimental. Ainsi, dans chaque configuration spatiale, le haut-parleur frontal reproduisait également un *son test* à égaliser.

L'auditeur était assis sur une chaise de sorte que sa tête soit située au point d'écoute représenté par le point noir sur la figure 4.1. Sa tête était positionnée précisément à l'aide d'un appui-tête et d'une mentonnière et était maintenue fixe tout au long du test. Chaque configuration spatiale des haut-parleurs était réalisée en orientant la chaise vers l'un des quatre haut-parleurs (voir figures 4.1 et 4.2). La session était découpée en quatre sous-sessions, chacune correspondant à une configuration spatiale des haut-parleurs. L'ordre dans lequel chaque sujet passait les quatre

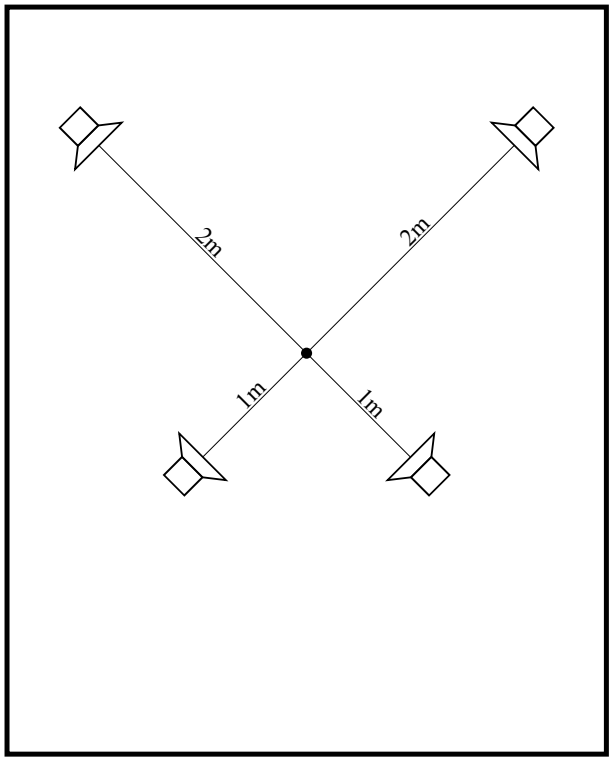


FIGURE 4.1 – Haut-parleurs placés autour du point d'écoute (représenté par le point noir) dans la salle de test, vus du dessus.

sous-sessions était déterminé aléatoirement. Chaque session durait entre cinq et dix minutes et l'auditeur pouvait se lever de la chaise entre chaque sous-session. L'appui-tête et la mentonnière étaient maintenus dans la même position tout au long de la session (comprenant les quatre sous-sessions). Ainsi, la position de la tête du participant était identique dans les quatre sous-sessions.

Les sons étaient générés depuis une carte son RME Fireface 800 et diffusés par des haut-parleurs Amadeus PMX 4 alimentés par des amplificateurs Audac DPA 154. Les quatre haut-parleurs étaient étalonnés afin de reproduire un niveau de 58.4 dB SPL (50 phones à la fréquence centrale du bruit) au point d'écoute. Chaque haut-parleur, quelle que soit sa position dans la pièce, restituait le même niveau à ce point de mesure unique.

4.2.4 Présentation virtuelle

Un enregistrement binaural des bruits diffusés par les haut-parleurs était présenté au casque dans cette session. Ainsi, les signaux acoustiques étaient similaires qu'ils soient présentés au casque ou par les haut-parleurs. Les variations interindividuelles de taille et de forme de la tête engendrent de faibles différences de pression et de temps d'arrivée aux oreilles aux fréquences étudiées dans cette expérience (Feddersen *et al.*, 1957). Ces différences sont inférieures aux plus petites différences perceptibles (Bernstein *et al.*, 1998) et les variations spectrales interindividuelles sont également faibles à ces fréquences. Les enregistrements ont donc été réalisés via une tête artificielle (Neumann KU100) permettant un enregistrement précis des sons en entrée des

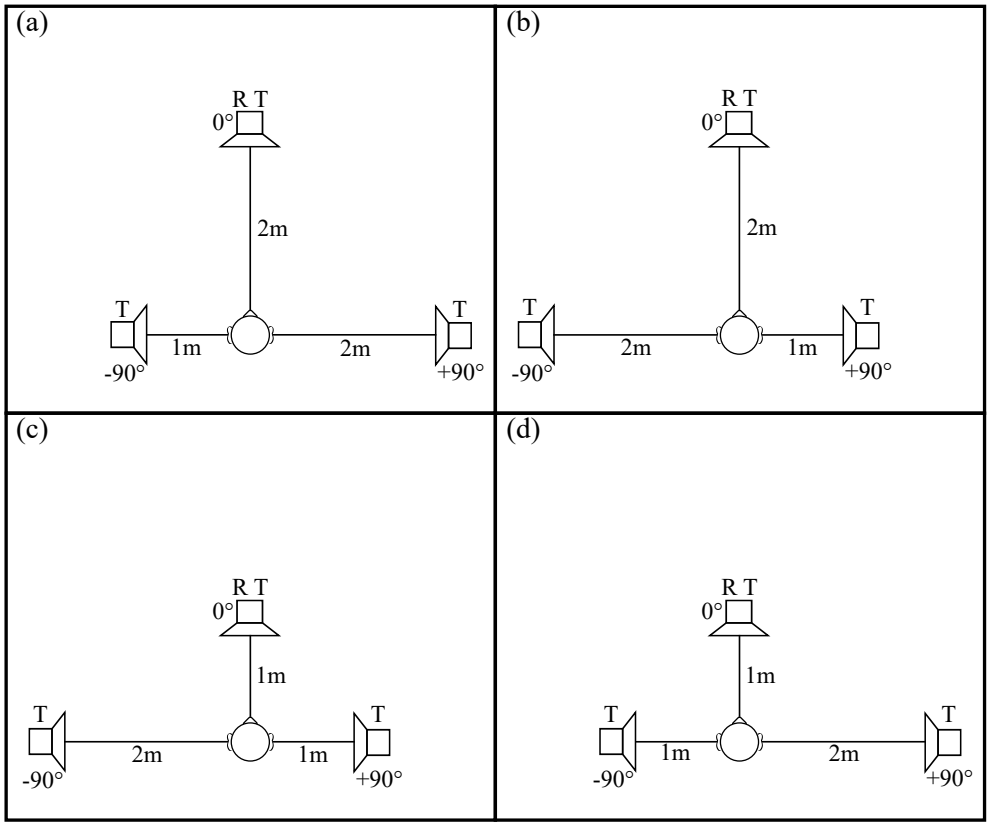


FIGURE 4.2 – Position des haut-parleurs dans chacune des quatre configurations spatiales. Les sources dénotées R jouaient les *sons de référence* et les sources dénotées T jouaient les *sons test*.

conduits auditifs externes (bloqués). La tête artificielle était précisément positionnée à la place de la tête des participants lors de la présentation réelle (voir figure 4.3) et était étalonnée à l'aide d'un pistonphone (Brüel & Kjær Type 4231) à 1 kHz. Douze stimuli ont été enregistrés de cette manière, correspondant aux sons délivrés par les haut-parleurs situés aux trois azimuts ($-90°$, $0°$ et $+90°$) dans chacune des quatre configurations spatiales (voir figure 4.2).

Les stimuli étaient restitués par un casque (Sennheiser HD650, circumaural ouvert) alimenté par la même carte son que lors de la présentation réelle (RME Fireface 800). Le niveau de sortie du système a été étalonné en plaçant le casque sur la tête artificielle étalonnée. La session se déroulait dans la même salle d'écoute que la session de présentation réelle. Les auditeurs étaient assis sur la même chaise, mais les haut-parleurs n'étaient pas visibles. Les participants avaient pour consigne de placer le casque confortablement sur leur tête et de ne pas en modifier la position tout au long du test. Les participants réalisaient douze égalisations de sonie, correspondant aux trois *sons test* (d'azimut $-90°$, $0°$ et $+90°$) enregistrés dans chacune des quatre configurations spatiales.

Les études précédentes (présentées dans les chapitres 2 et 3) avaient observé l'effet de l'ITD sur la sonie de sons purs. Afin de vérifier si cet effet est similaire pour le bruit à bande étroite utilisé lors de cette expérience, deux égalisations de sonie supplémentaires étaient réalisées lors

FIGURE 4.3 – Position de la tête artificielle, maintenue par l'appui-tête et la mentonnière durant l'enregistrement des stimuli.

de cette session. Le *son de référence* de cette égalisation était le bruit à bande étroite directement synthétisé et diffusé au casque de façon diotique (sans passer par l'enregistrement binaural via la tête artificielle), au même niveau de référence que les autres stimuli (50 phones). Les deux *sons test* à égaliser sur ce *son de référence* étaient le même bruit à bande étroite directement synthétisé, comportant pour l'un une ITD nulle (le son était donc identique à la référence) et pour l'autre une ITD de 772 µs. Cette valeur de 772 µs est identique à celle utilisée dans les expériences précédentes et correspond à l'ITD provoquée par une source positionnée à un azimut de 90° d'après le modèle basses fréquences de Kuhn (1977).

Ces quatorze égalisations étaient réalisées dans une même session qui durait approximativement trente minutes. L'ordre d'apparition des stimuli était défini aléatoirement pour chaque participant.

Source	**SS**	**DF**	**MS**	**F**	**Sig.p**
CP	35.099	1	35.099	17.946	**< .001**
CS	182.546	3	60.849	59.740	**< .001**
A	.023	1	.023	.017	.899
CP × CS	29.282	3	9.761	11.671	**< .001**
CP × A	.183	1	.183	.234	.634
CS × A	265.726	3	88.575	76.301	**< .001**
CP × CS × A	10.358	3	3.453	4.170	**.01**

Table 4.1 – Résultats de l'ANOVA conduite sur les PSE. CP désigne les Conditions de Présentation, CS la Configuration Spatiale et A l'Azimut de la source.

4.3 Résultats & Discussion

4.3.1 Résultats

Comme dans les sections 2.3 et 3.3, un PSE négatif signifie qu'à niveau égal, le *son test* était perçu plus fort que le *son de référence*. Les PSE seront donc représentés en inversant l'axe des ordonnées.

Dans un premier temps, des t-tests avec correction de Bonferroni ont été effectués sur les neuf PSE obtenus pour les *sons test* frontaux (émis par les quatre sources frontales dans les conditions de présentation réelle et virtuelle, plus le bruit diotique directement généré au casque). Ces tests avaient pour but de repérer d'éventuels biais liés à la procédure et n'ont révélé aucun écart significatif par rapport à la valeur nulle pour les neuf PSE.

Une ANOVA à mesures répétées a été réalisée sur les seize PSE obtenus dans les présentations réelle et virtuelle. Cette analyse comportait trois facteurs :

- Conditions de présentation (deux niveaux) : réelle, virtuelle.
- Configuration spatiale (quatre niveaux) : a, b, c, d (voir figure 4.2).
- Azimut (deux niveaux) : $-90°$, $+90°$.

Les résidus du modèle linéaire étaient normalement distribués. Les résultats de l'ANOVA sont présentés dans le tableau 4.1. Puisque cette étude se concentre sur les effets des conditions de présentation, les effets significatifs portant sur ce facteur seront évoqués en premier. Certains effets et interactions impliquant la configuration spatiale et l'azimut de la source sonore ont été causés par l'effet de salle et seront détaillés après l'analyse des résultats principaux.

Résultats principaux

Effet simple des Conditions de présentation : Les PSE étaient significativement plus importants (en valeur absolue) **dans les conditions de présentation virtuelle que dans les conditions de présentation réelle** ($p < .001$). Cet effet est représenté figure 4.4 (a). Des t-tests indiquent que **ces PSE étaient significativement différents de zéro dans les conditions de présentation virtuelle** ($t(159) = 13.677$, $p < .001$) **et réelle** ($t(159) = 11.433$, $p < .001$).

La figure 4.4 (b) indique le PSE moyen obtenu pour le bruit directement synthétisé et reproduit au casque, présenté avec une ITD de 772 µs. Un t-test a révélé que ce PSE était significativement différent de zéro ($t(19) = 5.296$, $p < .001$) et avait une valeur moyenne d'environ -1.5 dB, en accord avec les observations faites sur des sons purs dans les précédentes études.

Des t-tests pour échantillons indépendants ont révélé que ni le PSE obtenu dans les conditions de présentation virtuelle ($t(29.949) = 1.501$ (corrigé suite à une violation de l'hypothèse d'homogénéité des variances), $p = .144$), ni celui obtenu dans les conditions de présentation réelle ($t(178) = -.531$, $p = .596$) n'étaient significativement différents du PSE obtenu pour le stimulus comportant une ITD de 772 µs.

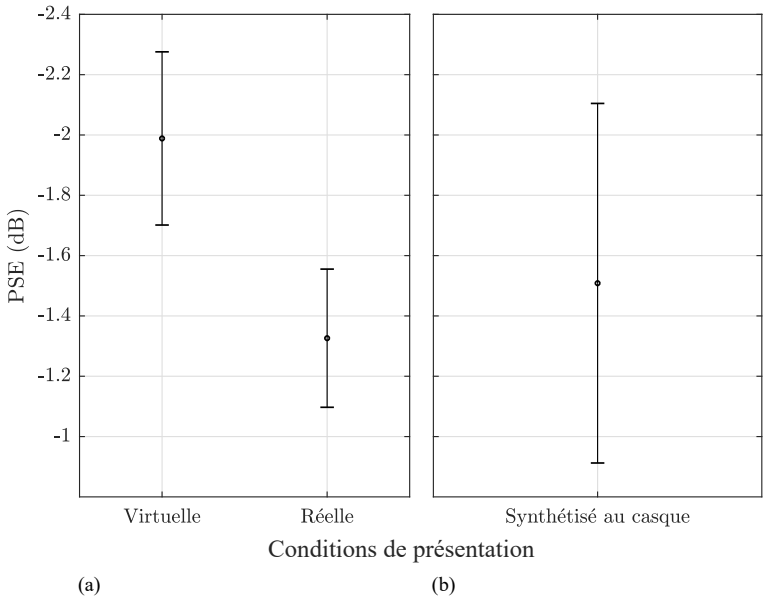


FIGURE 4.4 – PSE en fonction des conditions de présentation des sons (a) et pour les stimuli directement synthétisés et générés au casque (b), avec intervalles de confiance à 95%. Les axes des ordonnées sont inversés.

L'ITD moyenne des bruits latéraux diffusés dans les conditions de présentation réelle et virtuelle a été évaluée pour la comparer à l'ITD de 772 µs prédite par le modèle de Kuhn (1977). Pour ce faire, l'ITD moyenne a été mesurée entre les deux canaux des enregistrements binauraux pour chacun des huit *sons test* latéraux (deux azimuts dans quatre configurations spatiales). Ces huit mesures, reportées dans le tableau 4.2, mènent à une ITD moyenne légèrement supérieure à 772 µs ($\mu = 859$ µs, $\sigma = 19$ µs). Koehl *et al.* (2017) ont mesuré l'effet d'ITD supérieures à 772 µs sur la sonie. D'après leurs résultats, les augmentations de sonie dues à des ITD de 772 µs et 859 µs seraient du même ordre de grandeur.

Enfin, les stimuli ont été utilisés en entrée du modèle de sonie binaurale de Moore *et al.* (2016), qui prend en considération un certain nombre de phénomènes binauraux (tels que l'inhibition controlatérale) mais qui ne prédit pas d'augmentation de sonie liée à l'ITD. Ce modèle prédit que la sonie des sons latéraux de cette expérience est supérieure à celle des sons frontaux d'environ

Configuration spatiale	Azimut	ITD moyenne
a	−90°	865 µs
	90°	−837 µs
b	−90°	838 µs
	90°	−864 µs
c	−90°	837 µs
	90°	−865 µs
d	−90°	886 µs
	90°	−879 µs

Table 4.2 – ITD moyenne mesurée sur les enregistrements binauraux pour chaque source latérale dans chacune des quatre configurations spatiales.

0.5 phone, correspondant à un PSE plus faible que ceux révélés dans cette expérience (en valeur absolue).

Interaction Conditions de présentation × Configuration spatiale : L'effet des conditions de présentation sur le PSE n'était pas le même dans chacune des configurations spatiales des sources sonores ($p < .001$). Cette interaction est représentée figure 4.5. Des tests post-hoc LSD de Fisher avec correction de Bonferroni ont révélé un effet des conditions de présentation uniquement dans les configurations spatiales *c* ($p < .001$) et *d* ($p = .016$). Dans ces deux configurations, le PSE était plus important (en valeur absolue) dans les conditions de présentation virtuelle que dans les conditions de présentation réelle. La figure 4.2 représentant l'agencement des sources sonores dans chacune des configurations spatiales montre que les configurations *c* et *d* avaient en commun le fait que la source frontale était placée à 1 m du point d'écoute tandis qu'elle était à 2 m dans les configurations *a* et *b*. La distance de la source frontale est la seule différence permettant de distinguer ces deux paires de configurations spatiales. Le fait que le PSE dépende ou non des conditions de présentation était alors probablement liée à la distance de la source frontale.

Interaction Conditions de présentation × Configuration spatiale × Azimut : Cette interaction significative ($p = .01$) est représentée figure 4.6. Des tests post-hoc LSD de Fisher avec correction de Bonferroni ont révélé une différence significative entre les PSE obtenus dans les deux conditions de présentation dans la configuration spatiale *c* lorsque le *son test* était situé à −90° ($< .001$) et à +90° ($p = .007$). Dans cette configuration et pour ces deux azimuts, le PSE était plus important (en valeur absolue) dans la présentation virtuelle que dans la présentation réelle. Ainsi, l'augmentation de sonie dans la présentation virtuelle a été observée pour les deux sources latérales dans la configuration *c* tandis qu'elle n'a été observée qu'en combinant les deux sources latérales dans la configuration *d*. Dans cette représentation, la distance d'un haut-parleur latéral est renseignée par la combinaison de la configuration spatiale et de l'azimut du *son test*. Ainsi, dans les configurations spatiales *b* et *c*, le haut-parleur situé à un azimut de −90° était à une distance de 2 m et celui situé à un azimut de +90° était à une distance de 1 m du point d'écoute, ce qui était l'inverse dans les configurations spatiales *a* et *b* (voir figure 4.2).

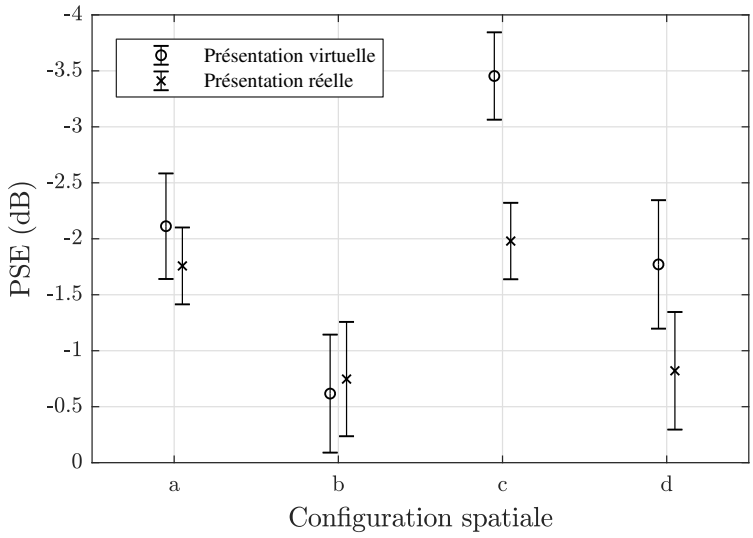


FIGURE 4.5 – PSE en fonction de la configuration spatiale des sources sonores pour les deux conditions de présentation, avec intervalles de confiance à 95%. L'axe des ordonnées est inversé.

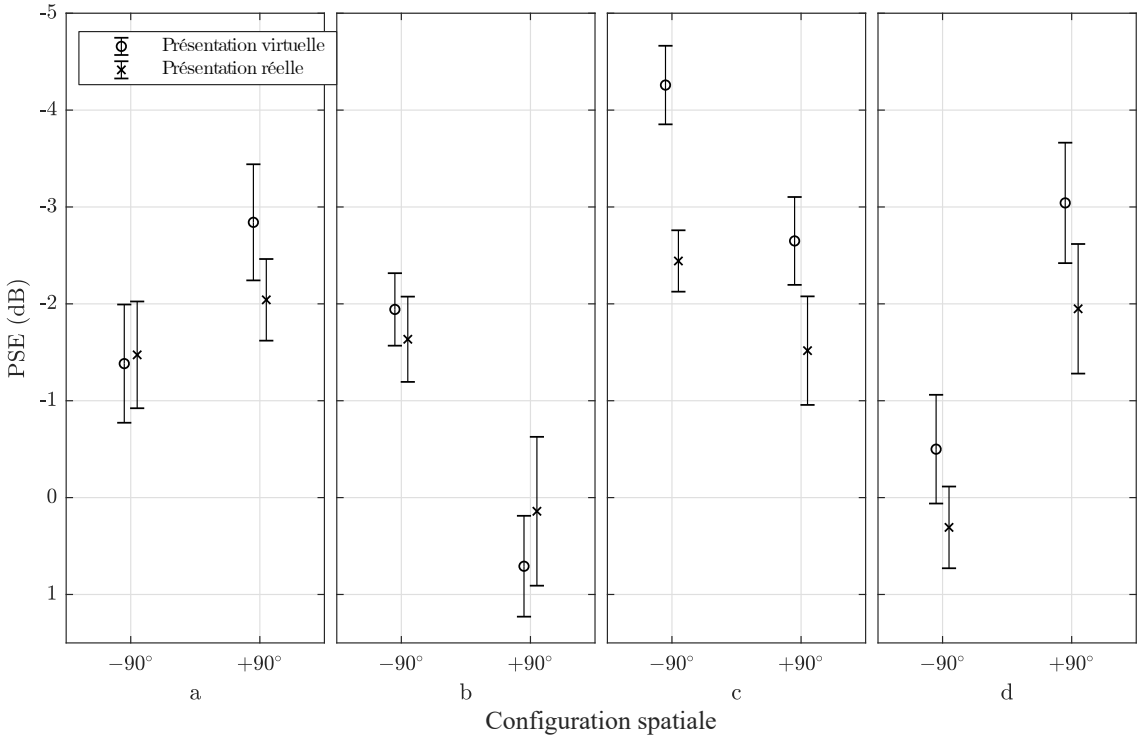


FIGURE 4.6 – PSE en fonction de l'azimut du *son test* dans chacune des quatre configurations spatiales pour chacune des deux conditions de présentation, avec intervalles de confiance à 95%. Les axes des ordonnées sont inversés.

Autres résultats

Effet simple de la Configuration spatiale : Cet effet simple indique que les PSE étaient différents selon la configuration spatiale des sources sonores ($p < .001$) et est représenté figure 4.7

(a). Puisque la salle d'écoute n'était pas anéchoïque, cet effet est susceptible d'avoir pour cause des variations de pression sonore aux oreilles des auditeurs et de la tête artificielle bien que les quatre haut-parleurs aient été étalonnés pour délivrer le même niveau sonore au centre de la tête des auditeurs et de la tête artificielle. Le niveau de pression sonore du *son de référence* a été mesuré dans chacune des quatre configurations spatiales aux oreilles de la tête artificielle. La figure 4.7 (b) représente la moyenne des mesures effectuée sur les deux oreilles dans chacune de ces configurations spatiales. Ces mesures montrent un motif globalement similaire à celui de l'effet simple de la configuration spatiale sur les PSE (figure 4.7 (a)). Les variations de niveau sonore aux oreilles semblent alors expliquer la dépendance de les PSE vis-à-vis de la configuration spatiale.

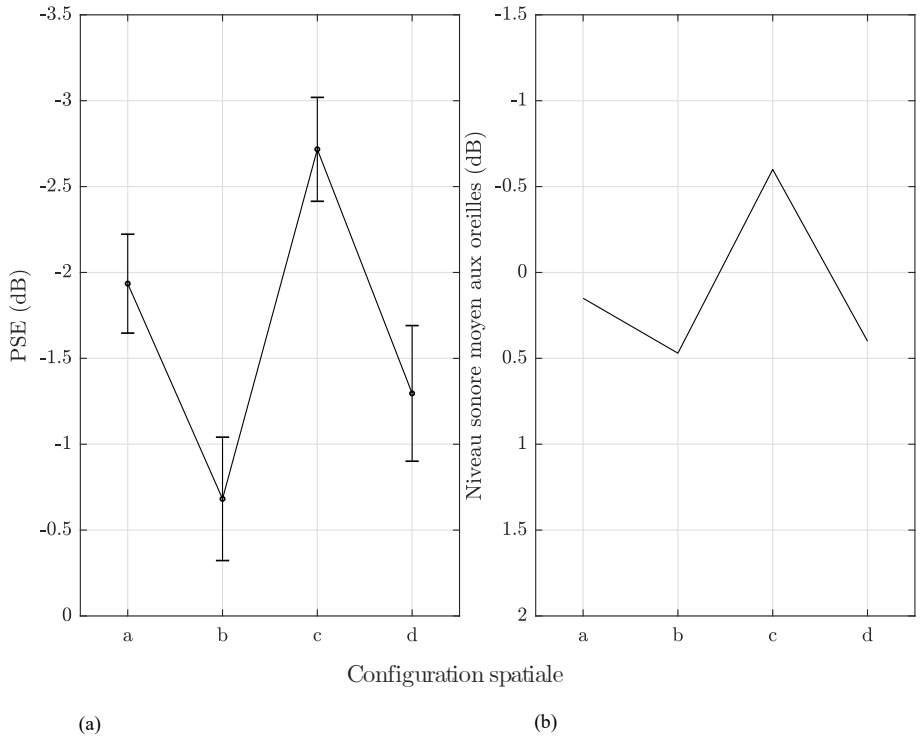


FIGURE 4.7 – PSE en fonction de la configuration spatiale avec intervalles de confiance à 95% (a) et niveau de pression sonore moyen entre les deux oreilles de la tête artificielle pour le *son de référence* de chacune des configurations spatiales (b). La valeur 0 dB dans cette figure correspond au niveau délivré par les sources de référence au centre de la tête (50 phones). Les axes des ordonnées sont inversés.

Interaction Configuration spatiale × Azimut : Cette interaction significative ($p < .001$) a également pour origine l'effet de salle, puisque les différentes sources ne délivraient pas toutes le même niveau aux oreilles des auditeurs et de la tête artificielle. Cependant, les différences entre les niveaux sonores aux oreilles délivrés par les quatre sources étaient présentes dans les conditions de présentation réelle et virtuelle. Ces différences ne sont donc pas susceptibles d'expliquer les effets observés des conditions de présentation sur les PSE.

4.3.2 Discussion

Les résultats de cette expérience montrent que :

- La sonie provoquée par des bruits à bande étroite basses fréquences était plus forte lorsqu'ils provenaient du côté des auditeurs que lorsqu'ils étaient frontaux, dans les conditions de présentation réelle (dans laquelle les sons étaient présentés par des haut-parleurs) et virtuelle (dans laquelle un enregistrement binaural des stimuli était présenté au casque).
- Cette augmentation était du même ordre de grandeur, dans chaque mode de présentation, à celle observée lorsque les sons étaient latéralisés par le seul ajout d'une ITD de 772 µs.
- Cette augmentation de sonie était plus faible dans la présentation réelle que dans la présentation virtuelle lorsque la source frontale était située à une distance de 1 m des auditeurs (dans les configurations *c* et *d*, voir figure 4.2).
- L'augmentation de sonie était la même dans les deux conditions de présentation lorsque la source frontale était située à une distance de 2 m des auditeurs (dans les configurations *a* et *b*).

Effet des Conditions de présentation

La sonie directionnelle observée dans les conditions de présentation réelle était globalement plus faible que celle observée dans les conditions de présentation virtuelle. Cette observation est en accord avec celles d'Epstein et Florentine (2009, 2012), qui ont mesuré un rapport de sonie binaural-à-monaural plus faible en augmentant la validité écologique de leurs expériences en diffusant leurs stimuli sur haut-parleurs plutôt qu'au casque et en ajoutant un support visuel aux auditeurs.

D'après les résultats de la présente étude, la distance de la source sonore frontale (jouant les *sons de référence*) a affecté différemment la sonie directionnelle dans les deux conditions de présentation des sons. Dans les conditions de présentation réelle, la tête des auditeurs était fixée par un appui-tête et une mentonnière. Ainsi, seule la source frontale était visible dans chaque configuration spatiale. Puisque la sonie peut être affectée par les informations auditives (Zahorik et Wightman, 2001) et visuelles (Mershon *et al.*, 1981) de distance, il est possible que ces informations soient à l'origine des différences de PSE observées dans les deux conditions de présentation selon la distance de la source sonore frontale.

Indices auditifs

Puisque le niveau de pression sonore était fixé en un même point d'écoute pour les quatre haut-parleurs, il ne variait pas selon que les sources sonores étaient situées à 1 m ou 2 m des auditeurs et ne peut donc pas expliquer les différences de PSE. D'autres caractéristiques acoustiques du son peuvent être affectées par la distance de leur source (*i.e.*, le rapport d'énergie champ direct à champ réverbéré, la distribution de l'énergie spectrale, voir section 1.2.3). Ces caractéristiques étaient cependant similaires pour les sons présentés aux oreilles des auditeurs dans les conditions de présentation réelle et aux oreilles de la tête artificielle lors de l'enregistrement binaural. Quelques variations interindividuelles ont toutefois pu résulter des différences entre les

HRTF des auditeurs (lesquelles filtraient les sons dans la présentation réelle) et les HRTF de la tête artificielle (lesquelles filtraient les sons dans la présentation virtuelle). Cependant, les HRTF ne dépendent de la distance de la source que dans l'espace proximal (< 1 m voire encore plus près aux fréquences étudiées d'après Brungart et Rabinowitz (1999) et Otani *et al.* (2009)). Ainsi, le fait que les stimuli aient été filtrés par des HRTF différentes selon les conditions de présentation n'est pas susceptible d'expliquer les différences observées entre ces deux modes pour les sons restitués à 1 m et 2 m.

Indices visuels

La dépendance des PSE vis-à-vis de la distance réelle des sources pourrait provenir de leur distance perçue. La distance perçue d'une source peut être affectée par des indices visuels de distance (voir section 1.2.3).

Les indices visuels de distance n'étaient pas les mêmes dans les deux conditions de présentation. Dans la présentation virtuelle, ni les sources sonores frontales ni les sources sonores latérales n'étaient présentes. Dans la présentation réelle, les auditeurs voyaient la configuration des haut-parleurs en entrant dans la salle d'écoute et en se positionnant sur la chaise. Puisque leur tête était maintenue dans une position fixe par l'appui-tête et la mentonnière (voir figure 4.3), seules les sources frontales étaient visibles tout au long de l'expérience. Ainsi, les participants n'avaient pas le même accès à l'information visuelle de distance selon les conditions de présentation et selon l'azimut des sources sonores.

Lorsque les sources sonores frontales étaient visibles et proches des auditeurs (donc dans les conditions de présentation réelle et dans les configurations spatiales *c* et *d*, voir figure 4.2), les PSE étaient significativement plus faibles (en valeur absolue) que lorsque les sources n'étaient pas visibles (dans les conditions de présentation virtuelle). Le fait que l'amplitude des PSE était plus faible signifie que la sonie des *sons de référence* et celle des *sons test* étaient plus proches. Puisque les PSE étaient négatifs (*i.e.,* les *sons test* étaient perçus plus forts que les *sons de référence*), cela peut être dû soit à une augmentation de la sonie du *son de référence*, soit à une diminution de la sonie du *son test*. Dans chacune des configurations spatiales des sources sonores, les *sons test* étaient diffusés à une distance de 1 m et 2 m (selon l'azimut de leur source). Puisque l'effet de l'ITD sur la sonie est symétrique (Koehl et Paquier, 2015), un potentiel effet de la distance de la source latérale sur les PSE aurait été compensé dans chaque configuration spatiale.

Les différences de PSE observées entre les deux conditions de présentation pourraient alors être dues à des différences de sonie des *sons de référence* restitués par les sources frontales. À un même niveau sonore aux oreilles, les sons restitués par une source proche seraient perçus comme étant plus forts si leur source est visible, ce qui ne serait pas le cas pour des sources plus lointaines. Un effet similaire a été reporté pour des sources visibles en salle anéchoïque par Mershon *et al.* (1981), où un même son était perçu plus fort lorsqu'un haut-parleur muet était présenté à une distance de 2.25 m des auditeurs que lorsqu'il était situé à 3.75 m de ces derniers, alors que le

niveau de pression sonore était fixé au point d'écoute et que les sons étaient diffusés par un même haut-parleur caché dans les deux conditions. Ce raisonnement basé sur les indices visuels de distance pourrait également expliquer le fait qu'aucune différence de sonie directionnelle n'ait été observée par Sivonen *et al.* (Sivonen *et al.*, 2005; Sivonen et Ellermeier, 2006) selon que les sons étaient présentés au casque ou sur haut-parleurs. En effet, les haut-parleurs étaient placés aux mêmes positions dans les deux conditions de présentation des sons de sorte que les auditeurs puissent visuellement identifier les sources sonores même lorsque les sons étaient restitués au casque.

Autres facteurs

Dans la présente expérience, les auditeurs étaient conscients que les sons qu'ils entendaient étaient diffusés soit par des haut-parleurs (dans les conditions de présentation réelle), soit par un casque (dans les conditions de présentation virtuelle). Le simple fait d'avoir conscience de la nature des sources sonores pourrait déclencher des phénomènes de haut niveau susceptibles d'avoir un effet sur la façon dont est traitée la sonie des sons perçus (Epstein et Florentine, 2012). La sonie de deux sons strictement identiques pourrait alors varier selon les conditions dans lesquelles ces sons étaient présentés aux auditeurs et, dans le cas présent, mener à des différences de PSE.

Les auditeurs auraient notamment pu identifier, dans les conditions de présentation réelle, que les sources sonores disposées autour d'eux jouaient le même son quelles que soient leurs positions. Ceci aurait pu mener à une certaine forme de constance de sonie, selon laquelle les jugements de sonie étaient moins dépendants des stimuli aux oreilles que lorsque les sources sonores n'étaient pas identifiées. Ainsi, la sonie des sons frontaux et latéraux aurait été moins différente que dans les conditions de présentation virtuelle, aboutissant à des PSE globalement plus faibles. Ce phénomène a notamment été évoqué par Sivonen et Ellermeier (2006) ainsi que par Meunier *et al.* (2016) pour expliquer certaines différences interindividuelles observées sur la sonie directionnelle.

La perception de distance pourrait également avoir interagi avec d'autres facteurs de haut niveau. Les sons sont susceptibles d'être internalisés (*i.e.*, perçus comme provenant de l'intérieur de la tête) lorsqu'ils sont reproduits au casque, surtout lorsqu'ils sont situés à l'avant (azimut de 0°) et dans des proportions moindres lorsqu'ils sont situés sur les côtés (azimut de ±90°) de l'auditeur (Hendrickx *et al.*, 2017a,b). Les sons reproduits sur des sources sonores réelles sont d'ordinaire externalisés quelle que soit leur position relative à l'auditeur (Hartmann et Wittenberg, 1996). Ainsi, les *sons de référence* (localisés à 0° d'azimut) auraient pu être moins bien externalisés que les *sons test* (localisés à ±90° d'azimut) dans les conditions de présentation virtuelle, tandis que les deux types de sons pourraient avoir été externalisés de la même manière dans les conditions de présentation réelle. Puisque l'externalisation est un pré-requis pour percevoir la distance d'une source sonore (Kolarik *et al.*, 2016), la distance perçue n'était probablement pas identique dans les deux conditions de présentation des sons, ce qui aurait pu mener à des différences de sonie traduites par des différences de PSE dans les deux conditions.

4.3.3 Conclusion

L'effet de l'ITD sur la sonie a été observé sur des bruits basses fréquences restitués par des haut-parleurs situés à différents azimuts, impliquant des indices de localisation binauraux naturels. Les deux précédents chapitres avaient étudié cet effet sur des sons purs restitués au casque, latéralisés par le seul biais d'une ITD.

Les PSE obtenus entre les sources frontales et latérales dépendaient des conditions dans lesquelles les sons étaient présentés. Plus précisément, l'augmentation de sonie provoquée par la latéralisation d'une source sonore était moins importante lorsque les sons étaient présentés par des sources réelles visibles que lorsqu'un enregistrement binaural était présenté aux auditeurs via un casque. Cette observation n'a été faite que lorsque la source frontale, utilisée en tant que référence, était située à la distance étudiée la plus proche (1 m) des auditeurs. L'une des hypothèses fondées sur ces résultats est que lorsque les sources étaient situées à 1 m, les sons étaient perçus comme étant plus forts lorsque la source était visible, ce qui n'était pas le cas pour les sources situées à 2 m. Les différences de PSE pourraient également être expliquées par d'autres paramètres ayant varié entre les deux conditions de présentation des sons. Par exemple, le seul fait d'écouter des sons restitués au casque plutôt que sur enceinte aurait pu modifier la façon dont les sons sont perçus. La sonie aurait pu être affectée différemment dans les deux conditions de présentation des sons par des processus de haut niveau tels que le degré d'externalisation des sources sonores.

Conclusion de la première partie

La première partie de cette thèse comporte trois expériences consacrées aux effets de la position et de l'ITD sur la sonie de sons basses fréquences. Cette partie s'inscrit directement à la suite des travaux entamés par Koehl *et al.* (Koehl et Paquier, 2015; Koehl *et al.*, 2015, 2017), qui ont relevé une augmentation de sonie lorsque des sons purs basses fréquences de faible niveau comportaient une ITD non nulle.

La première étude de cette partie a observé l'effet de l'ITD sur la sonie de sons purs de niveaux et fréquences différents de ceux utilisés dans les expériences passées. Les résultats ont montré un effet de l'ITD globalement plus prononcé à 200 Hz qu'à 100 Hz. Pour ces deux fréquences, l'ITD affectait la sonie de manière plus importante pour les sons restitués à faible niveau que pour les sons restitués à fort niveau. À 200 Hz, l'augmentation de sonie était significative pour des sons purs restitués jusqu'à un niveau de 60 phones. Cette observation est en accord avec l'hypothèse, formulée par Koehl et Paquier (2015), selon laquelle l'augmentation de sonie avec l'ITD pourrait être liée au bruit interne inhérent au système auditif.

Dans la seconde étude de cette partie, l'effet de l'ITD sur la sonie était étudié sur des sons purs restitués en présence d'un bruit externe diffusé à faible niveau. Les résultats n'ont pas montré de dépendance de l'effet de l'ITD sur la sonie vis-à-vis du bruit externe ajouté. La comparaison des résultats des deux premières études de cette partie ne montre aucune influence significative liée à l'ajout d'un bruit externe diotique, ne permettant pas de tirer de conclusion sur le rôle du bruit interne dans la sonie de sons basses fréquences. En revanche, les résultats cumulés de ces deux expériences montrent une augmentation de sonie liée à l'ITD jusqu'à des niveaux supérieurs à ceux précédemment observés, allant jusqu'à 80 phones, le plus fort niveau testé. Toutefois, la taille de cet effet était plus importante à faibles niveaux (de 40 phones à 50 phones) qu'aux niveaux supérieurs (de 60 phones à 80 phones). Cette expérience a permis de confirmer l'augmentation de sonie due à l'ITD lorsque les sons étaient présentés dans du bruit.

Dans la troisième étude de cette partie, des bruits basses fréquences provenant de différentes directions ont été présentés à des auditeurs soit via des haut-parleurs frontaux et latéraux, soit via un casque restituant une prise de son binaurale de ces bruits. L'augmentation de sonie pour les sons latéraux était similaire à celle observée lors des précédentes études, dans lesquelles l'ITD était synthétisée sur des sons purs restitués au casque. De plus, la sonie directionnelle était plus prononcée lorsque les sons étaient restitués au casque que lorsqu'ils l'étaient sur haut-parleurs, et ce uniquement lorsque les sources frontales étaient proches des auditeurs. Cette observation a permis de supposer l'existence d'interactions entre la distance perçue d'une source sonore et la sonie des sons qu'elle émet.

Deuxième partie

Sonie et distance

Chapitre 5

Sonie *aux oreilles* de bruit en fonction de la distance

Cette étude a été présentée au *23rd International Congress on Acoustics* à Aachen en 2019 (Berthomieu *et al.*, 2019b).

5.1 Introduction

Les résultats de l'expérience décrite dans le chapitre 4 permettent d'émettre l'hypothèse selon laquelle la sonie de sons émis par une source serait dépendante de la distance perçue de cette source. Plus précisément, les PSE étaient plus faibles lorsque les sons étaient restitués par des haut-parleurs visibles qu'au casque lorsque les sources frontales étaient positionnées à 1 m. Cette observation n'a pas été faite lorsque les sources frontales étaient à 2 m du point d'écoute, pour une même pression sonore aux oreilles.

Dans notre environnement quotidien, nous combinons l'information reçue par le biais de nos différents sens dans le but d'appréhender le monde dans lequel nous évoluons. La recherche auditive s'intéresse particulièrement aux interactions audiovisuelles puisque la vision et l'audition interagissent de façon particulièrement prononcée. Si les informations visuelles et auditives sont conflictuelles, le cerveau peut interpréter différemment chacune de ces informations pour résoudre le conflit (Kohlrausch et van de Par, 2005). L'effet ventriloque (Thurlow et Jack, 1973; Hendrickx *et al.*, 2015) permet par exemple de percevoir un stimulus audiovisuel cohérent malgré l'existence d'une disparité spatiale entre les stimuli visuel et sonore. En localisation auditive (Weerts et Thurlow, 1971; Warren, 1979; Bertelson et Radeau, 1981), la localisation d'un stimulus sonore peut être modifiée par l'ajout d'un stimulus visuel situé à une position différente (voir section 1.2.5).

La sonie peut dépendre de la position de la source sonore (voir section 1.4, *e.g.* Sivonen et Ellermeier (2006) pour la sonie directionnelle et Zahorik et Wightman (2001) pour la constance de sonie avec la distance). Puisque l'information visuelle peut modifier la localisation d'une source sonore, il est envisageable que cette modification de localisation ait un effet sur la sonie. La présente étude s'intéresse à ces relations et se focalise sur les interactions audiovisuelles dans des conditions où une source sonore est située à différentes distances d'un auditeur. Lorsqu'un

son est émis par une source sonore, il se propage dans l'environnement avant d'atteindre les oreilles de l'auditeur et son intensité diminue au fur et à mesure qu'il s'éloigne de sa source (voir section 1.2.3). Il pourrait alors être supposé que la sonie, qui dépend largement du niveau sonore aux oreilles de l'auditeur, décroisse de la même manière au fur et à mesure que l'auditeur s'éloigne de la source sonore. Cependant, si l'attention des auditeurs est orientée vers la sonie *à la source*, la sonie peut être indépendante de la distance (Mohrmann, 1939; Zahorik et Wightman, 2001). Les résultats de ces études témoignent d'une habilité des auditeurs à estimer la sonie *à la source* (correspondant à une estimation de puissance) de façon constante malgré les variations de niveau sonore à leurs oreilles (liées à l'éloignement ou au rapprochement des sources sonores).

Lors de l'expérience présentée dans le chapitre 4, l'attention des auditeurs n'était pas portée sur la sonie *à la source*. Pourtant, les auditeurs ont semblé, lorsque les sources étaient visibles, avoir pris en compte leur distance dans leurs jugements de sonie. La présente expérience a été mise en œuvre pour étudier les relations entre distance perçue et sonie, en particulier dans le cas où la distance de la source est perçue par le biais d'indices visuels.

Les participants avaient pour objectif d'estimer la sonie *aux oreilles* (correspondant à la force sonore perçue au niveau des oreilles des auditeurs) ainsi que la distance égocentrique de bruits délivrés depuis différentes distances du point d'écoute. Ces estimations étaient réalisées en fonction de différentes sources d'information sur la distance de la source sonore. La source pouvait être visible ou cachée, impliquant ou non des informations visuelles de distance. Les sons étaient diffusés dans des environnements anéchoïques ou réverbérants, impliquant ou non des indices auditifs absolus de distance. L'objectif de ces différentes estimations était d'observer :

- Si les informations visuelles modifiaient la sonie *aux oreilles*.
- Si les informations visuelles modifiaient la distance perçue de la source sonore.
- Si la modification de la sonie *aux oreilles* était due à la modification de la distance perçue de la source sonore.

5.2 Matériel et méthode

L'étude des relations entre sonie et distance de sources visibles requiert l'utilisation de protocoles expérimentaux relativement complexes. À titre d'exemple, l'étude de Calcagno *et al.* (2012), qui s'intéressait aux perceptions visuelle et auditive de distance, utilisait un haut-parleur monté sur un rail permettant de le déplacer à différentes distances des participants. Ces déplacements devaient être réalisés sans être perçus par les participants. Le protocole expérimental incluait alors des haut-parleurs supplémentaires, diffusant un bruit à fort niveau pendant les déplacements visant à masquer les bruits du rail. Cette méthode comporte certains désavantages, puisque la perception des stimuli sonores pourrait être affectée par la perception de ces bruits masquants diffusés à forts niveaux.

Afin de s'affranchir de ces contraintes, un environnement virtuel a été créé pour permettre aux expérimentateurs de manipuler la distance de la source, le niveau de diffusion de la source, la visibilité de la source ainsi que l'acoustique de la salle indépendamment tout en plaçant les

auditeurs dans des conditions visuelles et auditives réalistes. Durant l'expérience, les auditeurs étaient assis sur une chaise et étaient équipés d'un casque audio Sennheiser HD 650 (circumaural ouvert) ainsi que d'un casque de réalité virtuelle (HMD, pour *Head Mounted Display*) HTC Vive affichant un environnement visuel virtuel. Les sons étaient restitués au casque, mais un haut-parleur était représenté dans cet environnement et y jouait le rôle de source sonore. Dans l'environnement virtuel, l'auditeur faisait face au haut-parleur. Un grand panneau était également représenté dans l'environnement et pouvait totalement occulter le haut-parleur s'il était placé entre ce dernier et l'auditeur. L'environnement virtuel, d'un point de vue extérieur à celui du participant, est représenté figure 5.1 (a). Dans cette configuration, le panneau était décalé et ne cachait pas le haut-parleur. La figure 5.1 (b) représente l'environnement virtuel depuis le même point de vue lorsque le haut-parleur était caché par le panneau. La même scène vue du point de vue de l'auditeur est représentée figure 5.1 (c) lorsque le haut-parleur était visible, et figure 5.1 (d) lorsque le haut-parleur était caché. Dans cette dernière configuration, les grandes dimensions du panneau permettaient de totalement obstruer le champ de vision de l'auditeur.

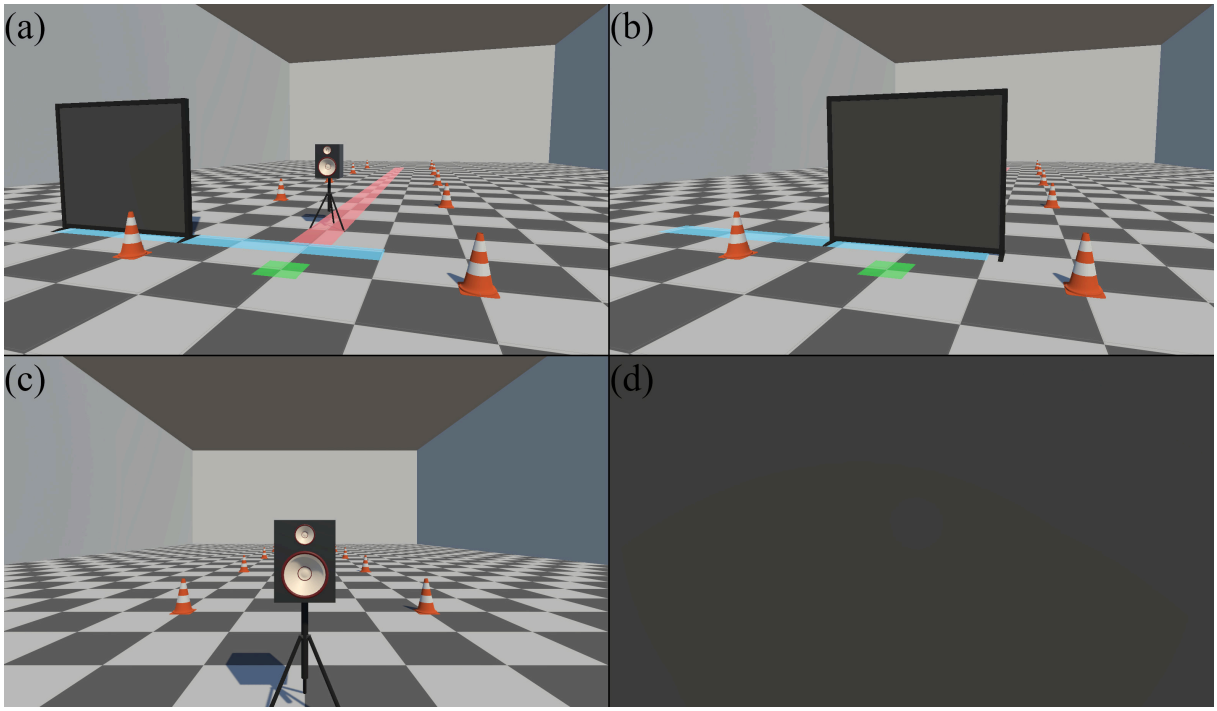


FIGURE 5.1 – Représentation de l'environnement visuel virtuel comprenant le haut-parleur, le panneau ainsi que des cônes de signalisation. Les images (a) et (b) montrent cet environnement d'un point de vue extérieur à celui des participants. Dans ces images, le carré vert représente la position de l'auditeur dans l'environnement virtuel, dont le point de vue faisait face au haut-parleur. Le chemin rouge représente la course du haut-parleur et le chemin bleu celle du panneau. Lorsque le panneau est dans la position de l'image (a), le haut-parleur est visible par les participants, dont le point de vue est représenté dans l'image (c). Lorsque le panneau est dans la position de l'image (b), le haut-parleur est totalement obstrué par le panneau, comme en témoigne le point de vue des participants représenté dans l'image (d). Le sol est texturé de carrés de surface $1\,\mathrm{m}^2$.

5.2.1 Stimuli auditifs

Trois salles différentes ont été étudiées. L'une était anéchoïque, une autre était une grande salle de sport dont le temps de réverbération TR_{60} est de 2 s (voir la photo figure 5.2 a.) et la troisième une petite salle de concert dont le TR_{60} est de 0.5 s (voir la photo figure 5.2 b.). Lors de l'expérience, les bruits larges bandes utilisés comme stimuli étaient restitués via un casque audio. Les source sonores étaient virtuellement positionnées à différentes distances des auditeurs dans ces différents environnements. Pour celà, les réponses impulsionnelles des salles ont été enregistrées en ambisonie d'ordre élevé (HOA, pour *Higher-Order Ambisonics*) d'ordre quatre. Un bruit blanc de durée 200 ms, généré sur le logiciel MATLAB, a ensuite été convolué par ces réponses impulsionnelles. Les signaux HOA ont permis de synthétiser des stimuli binauraux (restitués via un casque étalonné) avec suivi de mouvement. La procédure utilisée est détaillée en annexe A.

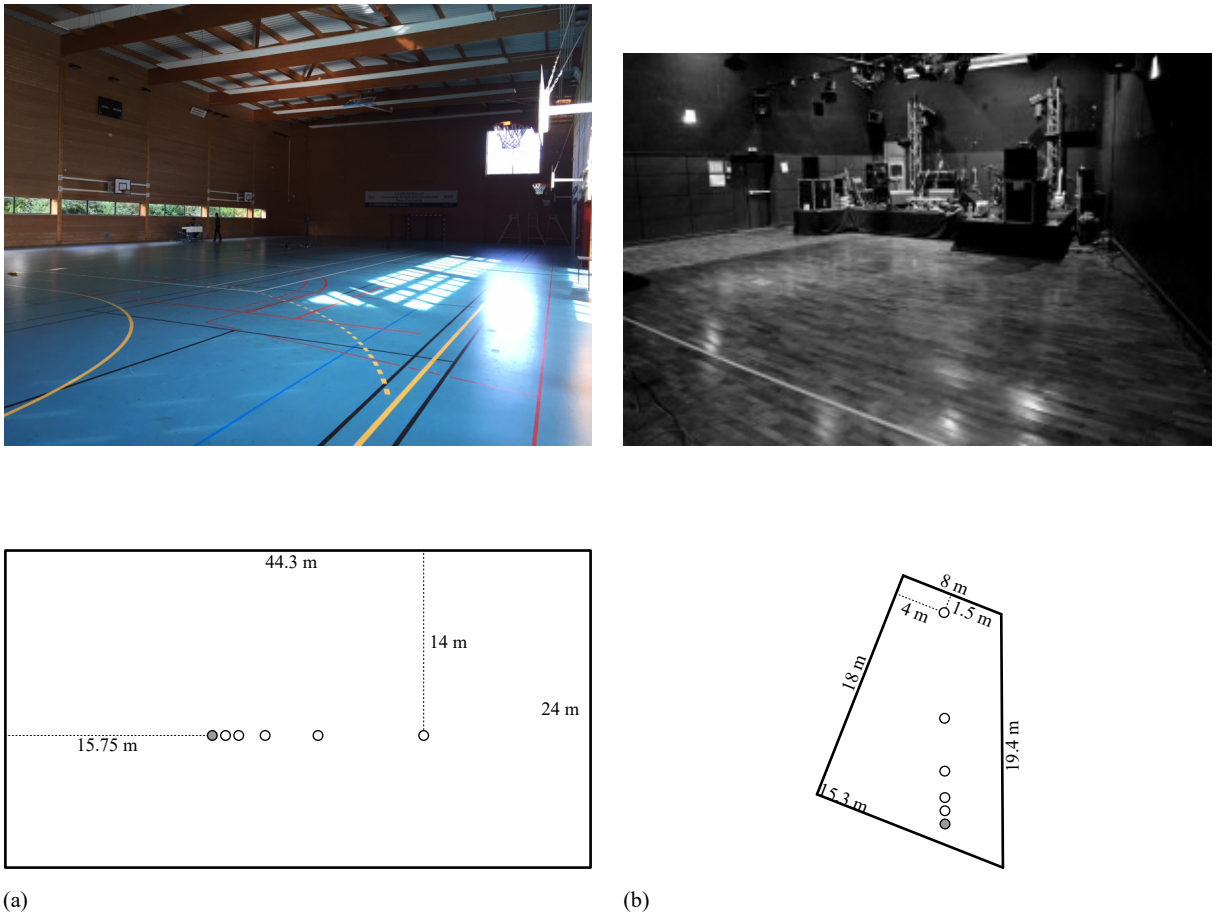


FIGURE 5.2 – Photos et schémas de la grande salle de sport (a) et de la petite salle de concert (b) dans lesquelles les enregistrements ont été réalisés. Dans les schémas, les cercles pleins représentent les positions du microphone (et donc des participants), les cercles vides représentent celles de la source sonore.

Les sources pouvaient restituer les bruits depuis cinq distances du point d'écoute : 1 m, 2 m, 4 m, 8 m et 16 m (voir les schémas de la figure 5.2). Les niveaux de champ direct et de champ

réverbéré ont été mesurés dans les deux environnements réverbérants et sont donnés en annexe B. Les bruits pouvaient être joués à trois niveaux de diffusion différents de sorte qu'un niveau de pression sonore aux oreilles donné ne correspondait pas à une distance de la source donnée. Ces niveaux étaient fixés à 0 dB, −6 dB et −12 dB par rapport à un niveau de référence. Ce niveau de référence était ajusté de sorte que le niveau de pression sonore soit de 80 dB SPL aux oreilles de l'auditeur lorsque la source était située à 1 m du point d'écoute.

45 stimuli ont été créés de cette manière, correspondant aux trois niveaux de diffusion pour une source située à cinq distances différentes dans chacun des trois environnements.

5.2.2 Stimuli visuels

Dans la revue d'études concernant la perception de distance égocentrique en environnement virtuel effectuée par Renner *et al.* (2013), les auteurs concluent avec une liste d'éléments susceptibles d'améliorer la qualité de la perception de distance en réalité virtuelle. Ils établissent que « dans le but de faciliter le plus possible la perception de distance, il est important de fournir une disparité binoculaire, d'utiliser une grande qualité de graphismes, d'ajuster soigneusement les paramètres de la caméra virtuelle, de représenter un environnement virtuel riche contenant une texture de sol régulièrement structurée, et d'améliorer la sensation de présence de l'utilisateur ». La présence décrit la sensation de se trouver dans un environnement donné bien que l'on se situe physiquement dans un autre environnement (Witmer et Singer, 1998). L'environnement visuel de cette expérience a été mis au point en respectant ces recommandations. L'utilisation d'un HMD permet de restituer aux participants la plupart des indices visuels monoculaires (l'occlusion, la taille relative, la hauteur dans le champ visuel, les différentes perspectives, la parallaxe de mouvement...) et binoculaires (la convergence, la disparité binoculaire) de distance. Les modèles 3D ont été créés sur le logiciel de modélisation 3D Blender (voir figure 5.3) et rendus en haute qualité sur les écrans de résolution 1200 × 1080 px du HMD à un taux de rafraîchissement de 90 images par seconde. Le sol était texturé de carrés de dimensions 1 m × 1 m. Des cônes de signalisation étaient situés à différentes distances du point d'observation afin d'enrichir l'environnement virtuel. Le suivi de mouvement de la tête (*head tracking*) fourni par le HMD constitue un élément permettant une bonne qualité d'immersion et améliore la sensation de présence de l'utilisateur (Renner *et al.*, 2013). De plus, ce suivi de mouvement a été lié au suivi de mouvement des stimuli sonores (voir annexe A) de sorte que la cohérence des informations transmises par les différentes modalités soit préservée, ce qui constitue un facteur important de la sensation de présence (Held et Durlach, 1992).

L'impression visuelle de la salle dans laquelle les sources sonores sont placées peut affecter l'extrernalisation des sons (Udesen *et al.*, 2015). De ce fait, trois environnements de tailles différentes ont été modélisés sur le logiciel Blender, correspondant à chacune des trois salles dans lesquelles les sons étaient virtuellement restitués. Pour les deux salles réverbérantes (la grande salle de sport et la petite salle de concert), les environnements virtuels étaient des salles ayant exactement les mêmes dimensions et formes que les salles réelles dans lesquelles les enregistrements ont été réalisés (voir les schémas des salles figure 5.2 et une capture d'écran des deux

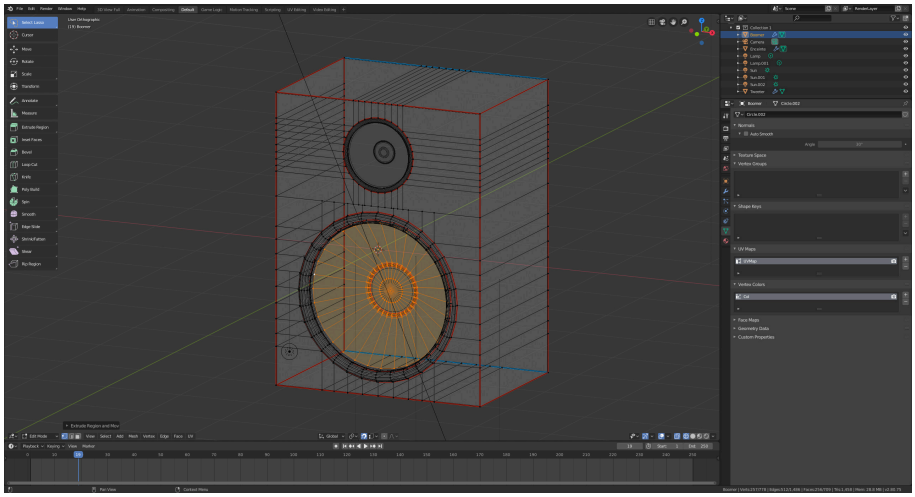
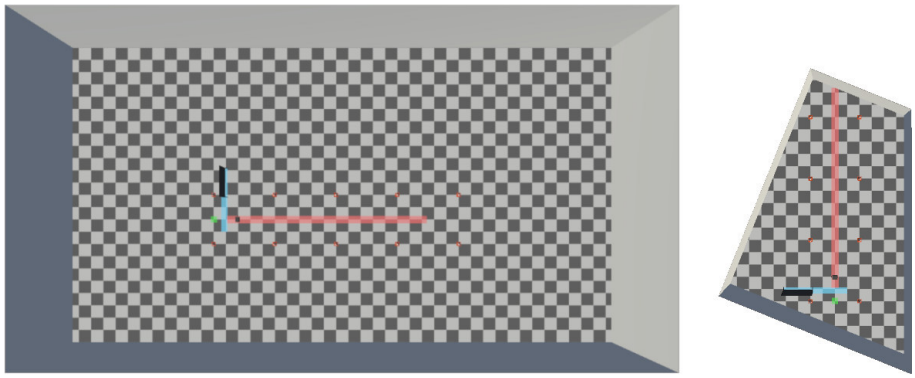


FIGURE 5.3 – Capture d'écran de l'interface de Blender pendant la création du modèle 3D du haut-parleur.

environnements virtuels vus du ciel figure 5.4). Pour la salle anéchoïque, aucun mur n'était représenté et l'environnement était composé d'un sol s'étendant à l'infini et dont la texture était régulièrement structurée.



(a) (b)

FIGURE 5.4 – Vue aérienne (prise à une hauteur de 50 m) des deux salles virtuelles représentant la grande salle de sport (a) et la petite salle de concert (b). Les indications visuelles sont les mêmes que dans la figure 5.1.

5.2.3 Procédure

L'expérience s'est déroulée dans la même cabine audiométrique que celle des expériences détaillées dans les chapitres 2 et 3. Les participants étaient assis sur une chaise et l'interface de réponse était affichée sur une tablette tactile posée devant eux (voir figure 5.5). Le HMD dispose d'une webcam permettant de filmer l'environnement réel, dont la sortie était affichée en temps réel dans l'environnement virtuel sur un panneau posé au sol devant le point d'observation. De cette façon, les participants pouvaient voir leurs mains ainsi que l'interface de réponse dans

l'environnement virtuel (voir figure 5.6). L'expérience était découpée en deux sessions, chacune correspondant à une tâche différente qui était soit d'estimer la distance égocentrique de la source sonore, soit d'estimer la sonie *aux oreilles* des sons diffusés par cette source.

FIGURE 5.5 – Photographie d'un participant durant l'expérience. Le participant porte le HMD et le casque audio et tape sa réponse sur la tablette tactile. La figure 5.6 présente ce que le participant voyait à ce moment précis.

Chaque session était ensuite découpée en trois sous-sessions correspondant à chacune des salles. Dans chacune de ces sous-sessions, chaque combinaison de distance, de niveau de diffusion et de visibilité de la source sonore était présentée à chaque participant quatre fois dans un ordre aléatoire. Cette procédure a donné lieu à cent-vingt estimations (cinq distances, trois niveaux de diffusion, deux conditions de visibilité et quatre répétitions) par participant dans chacune des trois sous-sessions (anéchoïque, salle de sport et salle de concert) composant chaque session (estimation de sonie *aux oreilles* ou de distance). Le tableau 5.1 donne un résumé de chaque session, facteur et niveau de facteur présents dans l'expérience.

La sonie *aux oreilles* était reportée à l'aide d'un protocole d'estimation de grandeur sans référence (Marks et Florentine, 2011). Plus précisément, les participants assignaient à chaque bruit un nombre, sans référence et sans borne, dont l'amplitude correspondait selon eux à la sonie *aux oreilles* évoquée. Les instructions données par l'expérimentateur précisaient que leur attention devait être portée sur le niveau perçu *aux oreilles*, en leur indiquant « plus vous entendez le son fort, plus le nombre que vous assignez doit être élevé ». Meunier *et al.* (2018) ont utilisé une procédure d'estimation d'amplitude pour évaluer la sonie directionnelle provoquée par des sources en mouvement autour des auditeurs. La sensibilité directionnelle en sonie ainsi évaluée

FIGURE 5.6 – Interface de réponse vue par le participant dans l'environnement virtuel. La sortie de la webcam présente sur le HMD était affichée sur un panneau virtuel, posé au sol devant les participants. La figure 5.5 montre le participant portant le HMD et tapant sa réponse sur l'interface à ce moment précis.

était similaire à celle mesurée par une procédure adaptative 2I2AFC (Meunier *et al.*, 2016). Ponsot *et al.* (2015) ont évalué la sonie de sons dont l'intensité croissait ou décroissait en utilisant ces deux procédures différentes et ont également obtenu des résultats similaires avec les deux procédures.

La distance égocentrique apparente était reportée en mètres.

Les participants tapaient leur estimation sur l'interface de réponse, qui s'affichait en temps réel précédée du mot « Sensation : » dans la session d'estimation de sonie *aux oreilles* et « Distance : » dans la session d'estimation de distance (voir figure 5.6). Les participants pouvaient réécouter le son de chaque essai autant de fois que nécessaire avant d'effectuer leur estimation. L'expérimentateur indiquait cependant aux participants que les sons ne devaient être réécoutés que si nécessaire et que les estimations devaient rester spontanées.

Tâche (session)	Estimation de sonie · Estimation de distance
Facteur	**Niveaux**
Salle (sous-session)	Anechoïque · Salle de concert · Salle de sport
Distance	1 m · 2 m · 4 m · 8 m · 16 m
Niveau de diffusion	0 dB · −6 dB · −12 dB
Haut-parleur	Visible · Caché
Répétition	1 · 2 · 3 · 4

TABLE 5.1 – Résumé des conditions expérimentales évaluées par chaque participant.

Après la réponse du participant à l'essai $n-1$, le panneau cachait le haut-parleur pendant que ce dernier se déplaçait à la position qu'il occuperait dans l'essai n. Ensuite, le panneau révélait ou non le haut-parleur (selon que ce dernier était visible ou caché dans l'essai n), le son était joué puis le participant pouvait taper sa réponse. Si le haut-parleur était caché dans l'essai n, le panneau restait dans sa position après la réponse du participant et durant le mouvement du haut-parleur vers sa position pour l'essai $n+1$ afin de ne pas révéler la position cachée du haut-parleur lors de l'essai n. Puisque le participant devait baisser la tête afin de répondre sur l'interface (voir figures 5.5 et 5.6), le programme ne jouait pas le son de l'essai suivant si le participant n'avait pas relevé la tête. Si tel était le cas, le programme attendait que l'angle de la tête par rapport à une position horizontale soit de $\pm 15°$ pour que le son soit joué. Le programme était exécuté en temps réel sur Max/MSP, qui communiquait par le protocole Open Sound Control (OSC) avec le moteur de jeu Unity dans lequel les objets virtuels étaient représentés. Un enregistrement vidéo de l'expérience est disponible à l'adresse `https://youtu.be/V1ifR558VO4`.

Vingt participants (cinq femmes et quinze hommes, âgés de vingt à vingt-cinq ans) ne reportant pas de problème d'audition ont participé à l'expérience et étaient rémunérés pour leur participation. Les participants portant des lunettes de vue pouvaient les garder lors du test, sans que cela ne pose de problème vis-à-vis du port du HMD. Les instructions étaient affichées telles que proposées par Marks et Florentine (2011) sur un panneau dans l'environnement virtuel (voir figure 5.7). Ces instructions étaient également lues par l'expérimentateur au début de chaque session (estimation de sonie *aux oreilles* ou de distance égocentrique), qui s'assurait que les participants avaient compris quelle grandeur évaluer.

Les participants réalisaient les deux sessions sur deux semaines différentes. Chaque sous-session démarrait par un pré-test d'une dizaine d'essais dont les réponses n'étaient pas gardées afin de se familiariser avec l'interface, les stimuli et l'environnement virtuel. L'écart inter-pupille de chaque participant était mesuré au début du test et était configuré sur le HMD afin d'éviter toute sensation d'inconfort durant l'expérience. La moitié des participants débutait par la session d'estimation de distance et l'autre moitié par la session d'estimation de sonie *aux oreilles*. Chaque session durait approximativement une heure (environ un quart d'heure par sous-session, plus les temps nécessaires aux instructions, aux changements de sous-session et aux pré-tests).

Après avoir réalisé la première session, les participants devaient répondre à un questionnaire comportant sept propositions différentes afin d'évaluer la qualité de la sensation de présence dans l'environnement virtuel. Ce questionnaire était inspiré de celui de Rébillat *et al.* (2012) et a été adapté afin de correspondre à l'environnement virtuel de la présente expérience (voir tableau 5.2). Les participants répondaient sur une échelle à sept points allant de « pas du tout d'accord » à « tout à fait d'accord » (Likert, 1932). Chaque réponse correspondait à un score compris entre -3 et 3, où -3 signifiait une faible sensation de présence et 3 une forte sensation de présence. Les scores ont été moyennés sur les sept propositions pour chaque participant et une moyenne générale a été calculée sur les moyennes de tous les participants. Le score global ainsi obtenu était positif ($\mu = 0.93$, $\sigma = 0.63$), témoignant d'une sensation de présence satisfaisante dans l'environnement virtuel.

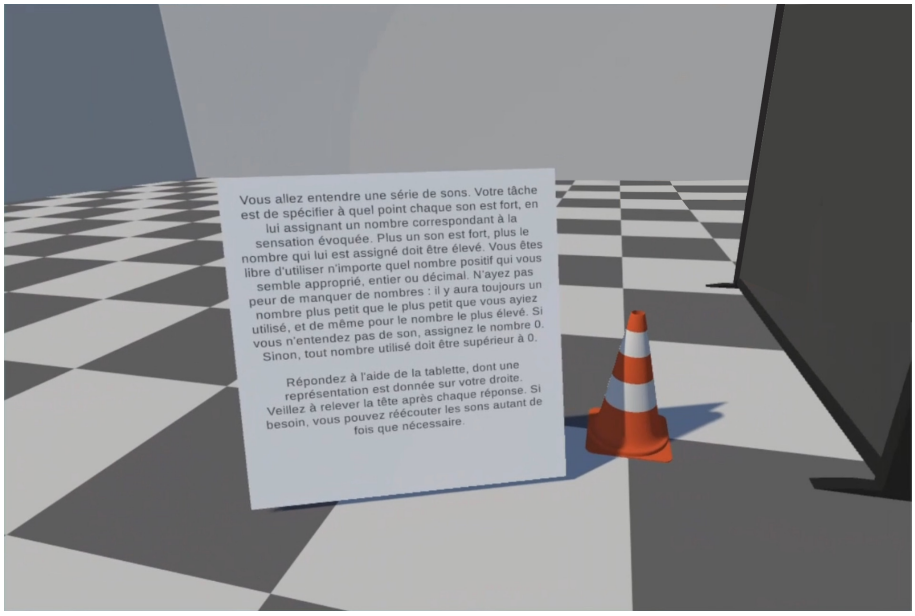


FIGURE 5.7 – Panneau d'instruction, situé à l'arrière gauche du participant dans l'environnement virtuel.

Proposition 1	J'avais l'impression de percevoir des objets réels.
Proposition 2	J'avais l'impression de faire face à un écran plutôt qu'à l'espace qui y était représenté.
Proposition 3	L'environnement virtuel m'a paru réel et je ne me souvenais plus être dans la cabine d'expérimentation.
Proposition 4	Je me souviens de l'environnement virtuel comme d'un endroit où je suis allé plutôt que comme d'une image générée par ordinateur.
Proposition 5	J'ai eu l'impression de pouvoir toucher les objets virtuels.
Proposition 6	Je me sentais présent dans le monde virtuel.
Proposition 7	Je me sentais entouré par le monde virtuel.

TABLE 5.2 – Questionnaire proposé aux participants à la fin de la première session afin d'évaluer la qualité de la sensation de présence dans l'environnement virtuel.

5.3 Résultats & Discussion

Deux ANOVA à mesures répétées ont été effectuées sur les résultats afin d'évaluer les effets et interactions mettant en jeux les différents facteurs sur les estimations de sonie *aux oreilles* et de distance égocentrique. Quatre facteurs ont été inclus dans chacune des deux analyses : la salle (trois niveaux), la distance de la source (cinq niveaux), le niveau de diffusion de la source (trois niveaux) ainsi que la visibilité de la source (deux niveaux).

5.3.1 Estimations de sonie *aux oreilles*

Les estimations de sonie *aux oreilles* ont été obtenues en utilisant une échelle libre. Ces résultats ont alors été normalisés pour que les réponses de tous les participants soient ramenées à

un référentiel commun avant d'être analysées statistiquement. Cette normalisation a été réalisée d'après la méthode décrite par Altmann *et al.* (2012) :

1. Chaque estimation est convertie en son logarithme.
2. Pour chaque condition expérimentale et pour chaque sujet, les quatre valeurs obtenues dans l'étape 1 (correspondant aux quatre répétitions d'une même condition expérimentale) sont moyennées.
3. Les valeurs calculées dans l'étape 2 sont moyennées pour chaque participant. Pour un participant donné, cette étape donne la moyenne de toutes ses estimations, toutes conditions expérimentales confondues.
4. Les vingt valeurs (une par participant) déterminées lors de l'étape 3 sont moyennées. Cette étape donne la moyenne générale des réponses de tous les participants confondus à toutes les conditions expérimentales confondues.
5. La valeur obtenue lors de l'étape 4 est retranchée aux valeurs déterminées lors de l'étape 3 pour chacun des participants. Les résultats de cette étape sont les écarts entre les moyennes individuelles et la moyenne générale de tous les participants.
6. Pour chaque participant, la valeur obtenue lors de l'étape 5 est retranchée aux valeurs obtenues lors de l'étape 2. Les estimations de chaque participant pour chacune des conditions expérimentales sont désormais normalisées par rapport à la moyenne générale.
7. Le logarithme inverse de chaque valeur obtenue dans l'étape 6 est calculé pour revenir à l'ordre de grandeur des estimations brutes. L'analyse statistique a été réalisée en amont de cette étape, qui n'a pour but que de tracer les estimations dans leur espace d'origine.

Une ANOVA à mesures répétées a été effectuée sur ces logarithmes normalisés des estimations de sonie *aux oreilles*. Les résidus du modèle linéaire étaient normalement distribués. Les résultats de cette analyse de variance sont disponibles dans le tableau 5.3.

Absence d'effets impliquant la visibilité de la source sonore

Le résultat principal de cette étude est que l'analyse de variance **n'a pas relevé d'effet significatif de la visibilité de la source** sur les estimations de sonie *aux oreilles*. De plus, **aucune interaction** ne mettant en jeu la visibilité de la source (notamment avec la distance) n'a été significative d'après cette analyse. Les estimations de sonie *aux oreilles* étaient alors similaires que le haut-parleur soit visible ou caché.

Autres résultats

Effet simple de la Distance : Les estimations de sonie *aux oreilles* étaient d'autant plus élevées que la source était proche des auditeurs ($p < .001$).

Effet simple du Niveau : Plus les sons étaient joués forts, plus la sonie *aux oreilles* était élevée ($p < .001$).

	Estimations de sonie *aux oreilles*				
Source	**SS**	**DF**	**MS**	**F**	**Sig.p**
S	1.783	2	.892	9.879	**< .001**
D	36.353	1.125*	32.316	186.970	**< .001**
N	24.763	1.042*	23.759	217.213	**< .001**
V	.017	1	.017	2.125	.161
S × D	2.059	2.883*	.714	37.202	**< .001**
S × N	.055	2.451*	.022	1.492	.233
D × N	.269	4.077*	.066	5.960	**< .001**
S × V	.004	2	.002	.528	.594
D × V	.014	2.363*	.006	.712	.518
N × V	.009	2	.004	1.230	.304
S × D × N	.051	6.806*	.007	1.005	.430
S × D × V	.010	4.612*	.002	.408	.828
S × N × V	.024	4	.006	1.888	.121
D × N × V	.025	4.379*	.006	.838	.514
S × D × N × V	.058	6.429*	.009	1.219	.299

* Le degré de liberté a été ajusté par une correction de Greenhouse-Geisser suite à une violation de l'hypothèse de sphéricité des résultats.

Table 5.3 – Résultats de l'ANOVA conduite sur les logarithmes normalisés des estimations de sonie *aux oreilles*. S désigne la salle, D la distance de la source, N le niveau de diffusion de la source et V la visibilité de la source.

Effet simple de la Salle : Les estimations de sonie *aux oreilles* étaient significativement plus élevées dans les environnements réverbérants que dans l'environnement anéchoïque ($p < .001$). Le niveau sonore aux oreilles des auditeurs était globalement plus fort dans les environnements réverbérants du fait de l'énergie supplémentaire apportée par le champ réverbéré, puisque les niveaux d'émission de la source sonore ont été étalonnés en égalisant le niveau sonore à 1 m dans les trois environnements (voir annexe B).

Interaction Salle × Distance : L'évolution des estimations de sonie *aux oreilles* avec la distance était différente dans les trois environnements ($p < .001$). Cette interaction est représentée figure 5.8 (a). Des tests post-hoc LSD de Fisher avec correction de Bonferroni ont été menés sur cette interaction. Lorsque la source était à 8 m, les estimations étaient plus faibles dans l'environnement anéchoïque que dans la salle de concert ($p < .001$). Lorsque la source était à 16 m, les estimations étaient plus faibles dans l'environnement anéchoïque que dans les deux salles réverbérantes ($p < .001$ pour les deux comparaisons).

Une régression des résultats sur une fonction puissance de forme $L = k \cdot r^b$ a été effectuée pour chaque courbe correspondant à l'évolution des estimations de sonie en fonction de la distance du haut-parleur dans chacune des trois salles. Dans cette équation, L désigne la quantité de sonie perçue, r la distance réelle de la source sonore et k est une constante. La valeur de l'exposant b de cette fonction de sonie permet d'évaluer la pente de la fonction de sonie, tel que l'expliquent Zahorik et Wightman (2001). Une constance de sonie parfaite serait alors traduite par un exposant $b = 0$. Les valeurs de b qu'obtiennent Zahorik et Wightman (2001) et pour lesquelles les auteurs considèrent que la sonie est constante sont approximativement comprises entre 0 et 0.1 (en valeur absolue). Dans l'une de leurs conditions expérimentales, les auteurs ne

Salle	**Exposant** b	**MSE**
Anéchoïque	-0.44	0.1228
Salle de concert	-0.28	0.1394
Salle de sport	-0.28	0.141

TABLE 5.4 – Exposants b des fonctions puissance $L = k \cdot r^b$ obtenues par régression sur les résultats d'estimation de sonie montrant l'évolution de la sonie avec la distance de la source sonore dans chacune des trois salles. Un exposant b proche de zéro témoignerait d'une constance de sonie avec la distance de la source. Les valeurs sont affichées avec la MSE entre les valeurs estimées lors de la régression et les valeurs réelles.

Niveau	**Exposant** b	**MSE**
-12 dB	-0.37	0.0388
-6 dB	-0.33	0.0167
0 dB	-0.3	0.0825

TABLE 5.5 – Exposants b des fonctions puissance $L = k \cdot r^b$ obtenues par régression sur les résultats d'estimation de sonie montrant l'évolution de la sonie avec la distance de la source sonore à chacun des trois niveaux de diffusion des bruits. Plus la valeur de l'exposant b est proche de zéro, moins les estimations étaient dépendantes de la distance.

relèvent pas de constance de sonie. L'exposant relevé dans ces conditions avait une valeur de -0.35. Les exposants b calculés pour chaque salle sur les estimations de sonie *aux oreilles* en fonction de la distance sont donnés dans le tableau 5.4. Ce tableau contient également l'erreur quadratique moyenne (MSE, pour *Mean Square Error*) entre les moyennes des estimations et les approximations faite lors de la régression. Ces coefficients témoignent d'une absence de constance de sonie avec la distance de la source sonore dans chacun des environnements. La décroissance plus douce de la sonie avec la distance dans les environnements réverbérants que dans l'environnement anéchoïque est susceptible de n'être due qu'au niveau total du stimulus aux oreilles des auditeurs. En effet, l'énergie du champ réverbéré, absent dans l'environnement anéchoïque, décroît plus faiblement avec la distance que celle du champ direct (voir annexe B). Aux distances lointaines, le niveau de pression sonore aux oreilles était alors plus élevé dans les environnements réverbérants que dans l'environnement anéchoïque.

Interaction Distance × Niveau : Les estimations de sonie *aux oreilles* n'évoluaient pas de la même manière avec la distance de la source sonore selon le niveau auquel elle émettait les sons ($p < .001$). Cette interaction est représentée figure 5.8 (b), qui ne présente cependant pas de différence marquée au niveau de l'allure des trois courbes. Une régression sur une fonction puissance de forme $L = k \cdot r^b$ a été effectuée pour ces trois courbes. Les coefficients b ainsi obtenus, présentés dans le tableau 5.5, sont du même ordre de grandeur pour les trois niveaux et témoignent également d'une absence de constance de sonie *aux oreilles* avec la distance.

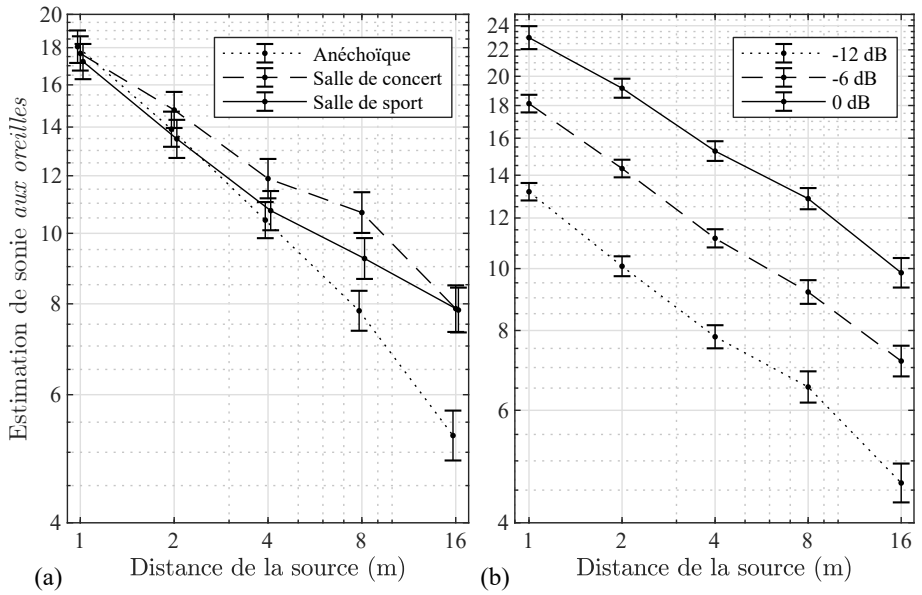


FIGURE 5.8 – Moyennes géométriques des estimations de sonie *aux oreilles* en fonction de la distance de la source sonore dans chacune des salles (a) et pour chacun des trois niveaux de diffusion des bruits (b), avec intervalles de confiance à 95%. Les moyennes et les limites des intervalles de confiance ont été calculées sur les logarithmes des estimations puis converties vers les estimations d'origine.

Bilan sur la sonie *aux oreilles*

Les résultats de cette étude indiquent que les estimations de sonie étaient similaires que le haut-parleur soit visible ou caché. De plus, les résultats indiquent une absence de constance de sonie lors d'estimations de sonie *aux oreilles* de bruits diffusés à plusieurs distances du point d'écoute. Cette absence de constance de sonie a été observée aussi bien lorsque le haut-parleur était caché que lorsqu'il était visible, en environnement anéchoïque comme en environnement réverbérant (voir figure 5.8 (a)). En définitive, les estimations de sonie *aux oreilles* suivaient le niveau sonore de pression aux oreilles des auditeurs, qui variait avec la distance de la source.

5.3.2 Estimations de distance

Les estimations de distance avaient été récoltées dans le but d'expliquer de potentiels effets de l'information visuelle sur les estimations de sonie, lesquels se sont révélés non-significatifs. Puisque ces résultats sont secondaires, ils seront analysés de façon moins exhaustive.

Puisque les estimations de distance sont globalement normalement distribuées sur une échelle logarithmique (Zahorik, 2005), chaque estimation de distance individuelle a été convertie en son logarithme avant d'être soumise à l'analyse de variance. La moyenne arithmétique des logarithmes des quatre répétitions de chaque essai a ensuite été calculée pour chaque essai et chaque participant. Ces valeurs n'ont pas été normalisées, puisque les participants utilisaient tous une unité commune dans leurs estimations : le mètre. Une ANOVA à mesures répétées a été effectuée

sur les résultats ainsi convertis (tableau 5.6). Les résidus du modèle linéaire étaient normalement distribués.

		Estimations de distance			
Source	**SS**	**DF**	**MS**	**F**	**Sig.p**
S	9.165	1.132*	8.097	14.601	**.001**
D	163.111	1.136*	143.526	244.122	**< .001**
N	16.564	1.183*	14.002	235.614	**< .001**
V	.170	1	.170	3.543	.075
S × D	.693	2.423*	.286	3.948	**.020**
S × N	.807	2.070*	.390	11.130	**< .001**
D × N	1.799	2.562*	.702	21.464	**< .001**
S × V	.294	1.374*	.214	7.659	**.006**
D × V	1.531	1.610*	.951	13.736	**< .001**
N × V	1.476	1.075*	1.373	14.996	**.001**
S × D × N	.196	4.847*	.040	1.525	.191
S × D × V	.062	8	.008	.910	.510
S × N × V	.005	2.595*	.002	.170	.893
D × N × V	.098	3.446*	.028	1.631	.184
S × D × N × V	.091	7.430*	.012	.848	.555

* Le degré de liberté a été ajusté par une correction de Greenhouse-Geisser suite à une violation de l'hypothèse de sphéricité des résultats.

Table 5.6 – Résultats de l'ANOVA conduite sur les estimations de distance égocentrique. S désigne la salle, D la distance de la source, N le niveau de diffusion de la source et V la visibilité de la source.

L'analyse n'a pas révélé d'effet simple de la visibilité de la source sonore sur les estimations de distance. En revanche, ce facteur a eu des interactions significatives avec les trois autres facteurs (niveau de diffusion de la source, distance de la source et salle).

Interaction Niveau × Visibilité : La figure 5.9 (a) montre que les estimations de distance dépendaient moins du niveau de diffusion de la source (et en conséquence du niveau de pression sonore aux oreilles) lorsque cette dernière était visible que lorsqu'elle était cachée. Ainsi, l'indice relatif de distance qu'est le niveau sonore affectait moins la distance perçue lorsque les indices visuels de distance étaient disponibles.

Interaction Distance × Visibilité : L'évolution des estimations de distance de la source avec la distance réelle était différente selon que la source était visible ou non. Cette interaction est représentée figure 5.9 (b), qui indique que les estimations de distance étaient plus proches de la distance réelle (ligne pointillée) lorsque la source était visible (ligne continue) que lorsqu'elle était cachée (ligne discontinue). Ce résultat est en accord avec la littérature, qui indique que les estimations de distance sont souvent surestimées pour des sources proches et sous-estimées pour des sources lointaines en perception de distance auditive (Zahorik, 2005), ce qui n'est pas le cas en perception de distance visuelle (Kolarik *et al.*, 2016).

Interaction Salle × Visibilité : Les estimations de distance dépendaient différemment des indices visuels dans les différentes salles. Des tests post-hoc LSD de Fisher avec correction de

Bonferroni ont révélé un effet significatif de la visibilité de la source sur les estimations de distance dans la salle anéchoïque ($p = .042$) uniquement, comme le montre la figure 5.9 (c). Cette figure représente les estimations moyennes de distance dans chacune des trois salles lorsque le haut-parleur était visible (ligne continue) ou caché (ligne discontinue). Les indices visuels de distance ont ainsi affecté la distance perçue uniquement lorsqu'ils étaient les seuls indices absolus disponibles.

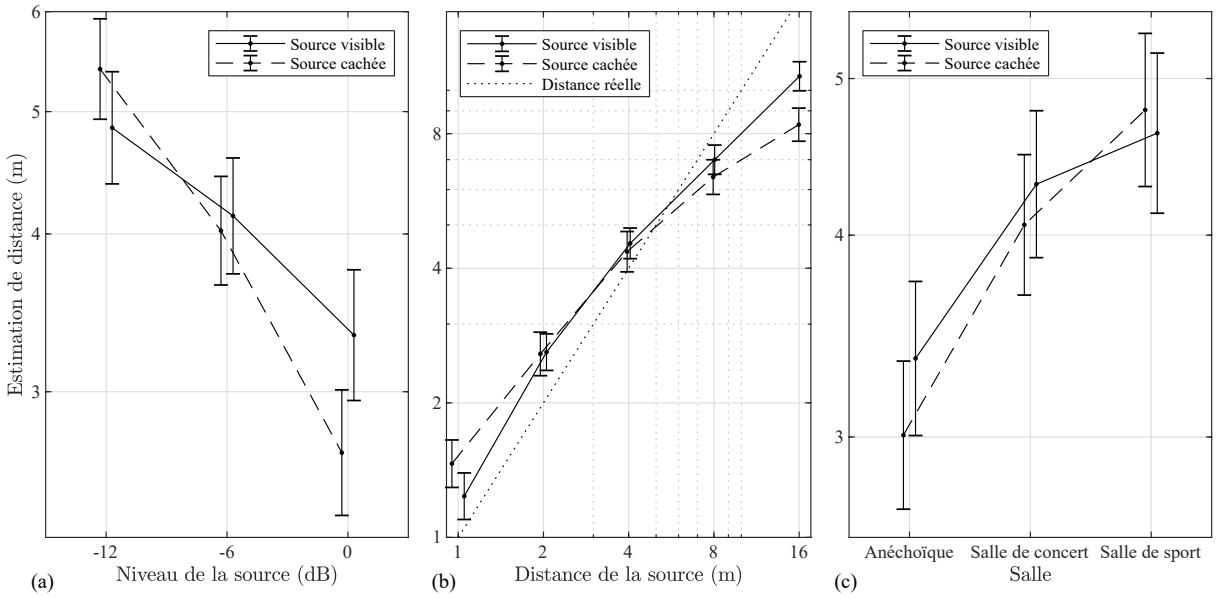


FIGURE 5.9 – Moyennes géométriques des estimations de distance en fonction du niveau (a), de la distance (b) et de la salle (c) pour les deux conditions de visibilité de la source, avec intervalles de confiance à 95% calculées comme expliqué dans la figure 5.8.

5.3.3 Conclusion

L'objectif de cette étude était d'observer si la sonie *aux oreilles* dépendait de la distance perçue de la source sonore. Les résultats de cette expérience montrent que les estimations de sonie *aux oreilles* étaient identiques que la source sonore soit visible ou cachée dans toutes les conditions expérimentales testées. Pourtant, la distance égocentrique perçue dépendait significativement de la visibilité de la source sonore. La sonie *aux oreilles* du bruit blanc diffusé par cette source était donc basée sur le niveau sonore aux oreilles des auditeurs, indépendamment de la distance perçue de la source sonore. Ces résultats sont en accord avec les études précédentes n'ayant pas relevé de constance de sonie avec la distance de sources de sons non-familiers (Stevens et Guirao, 1962; Petersen, 1990; Altmann *et al.*, 2012). La prochaine étude de cette thèse s'intéressera au rôle joué par les indices visuels de distance dans les estimations de sonie *à la source*, similaires à des estimations de puissance de la source sonore, dans un environnement réverbérant et un environnement anéchoïque.

Chapitre 6

Sonie *à la source* de bruit en fonction de la distance

6.1 Introduction

L'expérience décrite dans le chapitre 5 montre que les estimations de sonie *aux oreilles* de bruits blancs diffusés depuis différentes distances du point d'écoute étaient identiques que leur source sonore soit visible ou cachée. Ces estimations étaient indépendantes de la perception de la distance et étaient pleinement définies par les niveaux sonores aux oreilles des auditeurs.

Mohrmann (1939) ainsi que Zahorik et Wightman (2001) ont étudié les relations entre sonie et distance en orientant l'attention des auditeurs vers la sonie *à la source*, correspondant à la puissance sonore. Les auteurs reportent alors un phénomène de constance de sonie avec la distance, où la perception de sonie *à la source* reste constante malgré une décroissance du niveau sonore aux oreilles des auditeurs.

Dans l'étude de Zahorik et Wightman (2001), les sources sonores étaient virtuellement positionnées à différentes distances des auditeurs dans un environnement réverbérant et les sons étaient restitués au casque. Les sources n'étaient pas visibles et le champ réverbéré fournissait les informations de distance (par le biais du rapport d'énergie champ direct à champ réverbéré (DRR)) et de puissance (par le biais de l'énergie du champ réverbéré) de la source sonore (voir section 1.4.2). Dans ces conditions, la sonie *à la source* pourrait être évaluée de deux façons différentes. Il serait possible soit d'utiliser directement l'information de puissance délivrée par le champ réverbéré, soit d'utiliser la distance perçue ainsi que le niveau sonore perçu aux oreilles. En effet, si un auditeur entend un son faible mais qu'il perçoit sa source comme étant lointaine, il pourrait estimer que la source est puissante et que la faiblesse du son entendu est due à la distance. Afin d'observer laquelle de ces deux méthodes était utilisée par les auditeurs, Zahorik et Wightman (2001) ont également récolté des estimations de distance de la source. Les distances ainsi reportées (par estimation en mètres) étaient différentes des distances réelles, favorisant l'hypothèse selon laquelle les auditeurs se basaient directement sur l'énergie du champ réverbéré. Cependant, l'écart entre les distances estimées et les distances réelles pourrait être expliqué par le protocole utilisé, qui donne des estimations peu précises (voir section 1.2.3).

En définitive, le rôle joué par la distance perçue dans l'estimation de sonie *à la source* n'est pas pleinement défini. De plus, les participants de l'étude de Zahorik et Wightman (2001) disposaient en permanence d'indices directs de la puissance de la source puisque tous les sons étaient présentés dans un environnement réverbérant. Dans un environnement anéchoïque, cet indice n'est pas disponible et les auditeurs ne disposent pas d'indice auditif absolu de puissance ou de distance de la source sonore.

Si la source est visible dans un environnement anéchoïque, les auditeurs disposent alors d'indices absolus de distance. La présente étude observera la capacité des auditeurs à estimer la sonie *à la source* en environnement anéchoïque lorsque les auditeurs disposent d'informations visuelles de distance. Le rôle joué par la perception visuelle de distance dans la sonie *à la source* sera également observé dans un environnement réverbérant, dans lequel des indices de distance et de puissance de la source sont disponibles simultanément.

6.2 Matériel et méthode

Le protocole expérimental de cette étude était similaire à celui de l'expérience précédente (expliqué dans la section 5.2) sur de nombreux points. L'expérience se déroulait dans la même cabine d'expérimentation et les participants disposaient du même environnement virtuel, restitué par le même HMD et le même casque audio. L'interface de réponse était également identique.

6.2.1 Stimuli auditifs

Les stimuli auditifs étaient les mêmes que ceux utilisés lors de l'expérience du chapitre 5. Les haut-parleurs virtuels diffusaient les bruits aux trois mêmes niveaux de diffusion (0 dB, −6 dB et −12 dB par rapport à un niveau de pression sonore de 80 dB SPL mesuré à 1 m de la source) et étaient situés aux cinq mêmes distances (1 m, 2 m, 4 m, 8 m, et 16 m). Seuls l'environnement anéchoïque et l'environnement le plus réverbérant de l'expérience précédente (la salle de sport) ont été étudiés dans cette expérience.

30 stimuli sonores ont ainsi été étudiés, correspondant à un bruit blanc diffusé à trois niveaux différents par des sources situées à cinq distances différentes dans chacun des deux environnements. Les sons étaient restitués en binaural avec suivi de mouvement selon la technique détaillée en annexe A. Le matériel utilisé pour restituer les sons était le même que dans la section 5.2.

6.2.2 Stimuli visuels

Les stimuli visuels étaient également identiques à ceux utilisés lors de la précédente expérience. Les modèles 3D étaient restitués en temps réel sur l'écran du HMD via le moteur de jeu vidéo Unity.

6.2.3 Procédure

La procédure était similaire à la procédure de l'expérience précédente, détaillée dans la section 5.2. L'expérience se déroulait au sein d'une seule et même session dans laquelle la tâche des participants était d'estimer la sonie *à la source* des bruits qu'ils entendaient. Les instructions données aux participants étaient d'estimer la *puissance* de la source sonore, ce terme ayant été préféré pour des raisons de clarté et de compréhensibilité de la tâche par les participants. L'expérimentateur indiquait notamment aux auditeurs « plus la source joue le son fort, plus le nombre que vous assignez doit être élevé ».

La session de test était découpée en deux sous-sessions correspondant à chacun des deux environnements étudiés. Dans chacune de ces deux sous-sessions, cent-vingt estimations de sonie *à la source* étaient réalisées (cinq distances, trois niveaux de diffusion, deux conditions de visibilité de la source sonore et quatre répétitions de chaque combinaison) par chaque participant. Le tableau 6.1 regroupe les différents facteurs de l'expérience ainsi que leurs différents niveaux. Les estimations de sonie *à la source* étaient réalisées selon le même protocole d'estimation de grandeur sans référence (Marks et Florentine, 2011) que celui utilisé dans la section 5.2 pour les estimations de sonie *aux oreilles*. Les participants tapaient leur estimation sur la tablette tactile, qui s'affichait en temps réel à la suite du mot « Puissance : ». Les participants pouvaient réécouter les sons autant de fois que nécessaire avant de valider leur réponse. Le reste de la procédure était identique à celle de l'expérience du chapitre 5.

Tâche (session)	Estimation de sonie *à la source*
Facteur	**Niveaux**
Salle (sous-session)	Anechoïque · Salle de sport
Distance	1 m · 2 m · 4 m · 8 m · 16 m
Niveau de diffusion	0 dB · −6 dB · −12 dB
Haut-parleur	Visible · Caché
Répétition	1 · 2 · 3 · 4

Table 6.1 – Résumé des conditions expérimentales évaluées par chaque participant.

Dix-sept participants (trois femmes et quatorze hommes âgés de vingt à cinquante-deux ans) ne reportant pas de problème d'audition ont participé à l'expérience. Aucun d'eux n'avait d'expérience en tests psychoacoustiques. Les instructions étaient écrites sur un panneau dans l'environnement virtuel et étaient également lues par l'expérimentateur au début de la session, qui s'assurait que les participants avaient compris quelle grandeur évaluer. Les participants portant des lunettes pouvaient les garder durant l'expérience sans que cela ne pose de problème vis-à-vis du HMD. L'écart inter-pupille de chaque participant était mesuré au début de la session et était configuré sur le HMD. Un pré-test d'une dizaine d'essais était réalisé par les participants au début de chaque sous-session afin de se familiariser avec l'environnement et la procédure. L'ordre dans lequel les participants réalisaient les deux sous-sessions était aléatoire. Une pause était proposée entre les deux sous-sessions durant laquelle les participants pouvaient retirer le

HMD et sortir de la cabine d'expérimentation. L'expérience durait approximativement trois quarts d'heure pour chaque participant (environ un quart d'heure par session, plus les temps nécessaires aux instructions, au changement de session et aux pré-tests).

6.3 Résultats & Discussion

Les résultats ont été traités de la même manière que les estimations de sonie *aux oreilles* effectuées lors de l'expérience du chapitre 5 (Altmann *et al.*, 2012).

Une ANOVA à mesures répétées a été conduite sur les estimations de sonie *à la source*. Quatre facteurs ont été inclus dans cette analyse : la salle (deux niveaux), la distance de la source (cinq niveaux), le niveau de diffusion de la source (trois niveaux) ainsi que la visibilité de la source (deux niveaux). Les résidus du modèle linéaire étaient normalement distribués. Les résultats de cette ANOVA sont donnés tableau 6.2.

Variable dépendante		**Sonie *à la source***			
Source	**SS**	**DF**	**MS**	**F**	**Sig.p**
S	.007	1	.007	.037	.851
D	3.240	1.091*	2.971	5.938	**.024**
N	11.049	1.094*	10.103	71.034	**< .001**
V	.003	1	.003	.167	.688
S × D	2.340	4	.585	60.862	**< .001**
S × N	.119	2	.059	4.533	**.018**
D × N	.036	3.204*	.011	.589	.636
S × V	.217	1	.217	10.666	**.005**
D × V	1.634	1.647*	.992	21.641	**< .001**
N × V	.001	2	.000	.118	.889
S × D × N	.040	3.342*	.012	.892	.460
S × D × V	.369	2.422*	.152	14.645	**< .001**
S × N × V	.004	2	.002	.325	.725
D × N × V	.126	3.773*	.034	3.878	**.008**
S × D × N × V	.025	8	.003	.865	.548

* Le degré de liberté a été ajusté par une correction de Greenhouse-Geisser suite à une violation de l'hypothèse de sphéricité des résultats.

Table 6.2 – Résultats de l'ANOVA conduite sur les estimations de sonie *à la source*. S désigne la salle, D la distance de la source, N le niveau de diffusion de la source et V la visibilité de la source.

Principaux résultats

Effet simple de la Distance : L'effet significatif de la distance témoigne d'une absence de constance de sonie sur la globalité des résultats ($p = .024$). Cependant, certaines interactions faisant intervenir la distance de la source sont significatives (e.g., « Salle × Distance »). L'évolution de la sonie *à la source* avec la distance a de ce fait été étudiée plus en détail au niveau de ces interactions.

Interaction Salle × Distance : Les estimations de sonie *à la source* n'évoluaient pas de la même manière avec la distance dans les deux environnements ($p < .001$). La figure 6.1 (a) représente cette interaction. Il apparaît nettement que les estimations de sonie *à la source* dépendaient moins de la distance de la source sonore dans la salle de sport (l'environnement réverbérant, en trait continu) que dans l'environnement anéchoïque (en trait discontinu). Cette observation est conforme à l'hypothèse de Zahorik et Wightman (2001), puisque la sonie *à la source* est moins dépendante de la distance (et donc du niveau sonore aux oreilles) lorsque les informations (de puissance et de distance) fournies par le champ réverbéré sont disponibles.

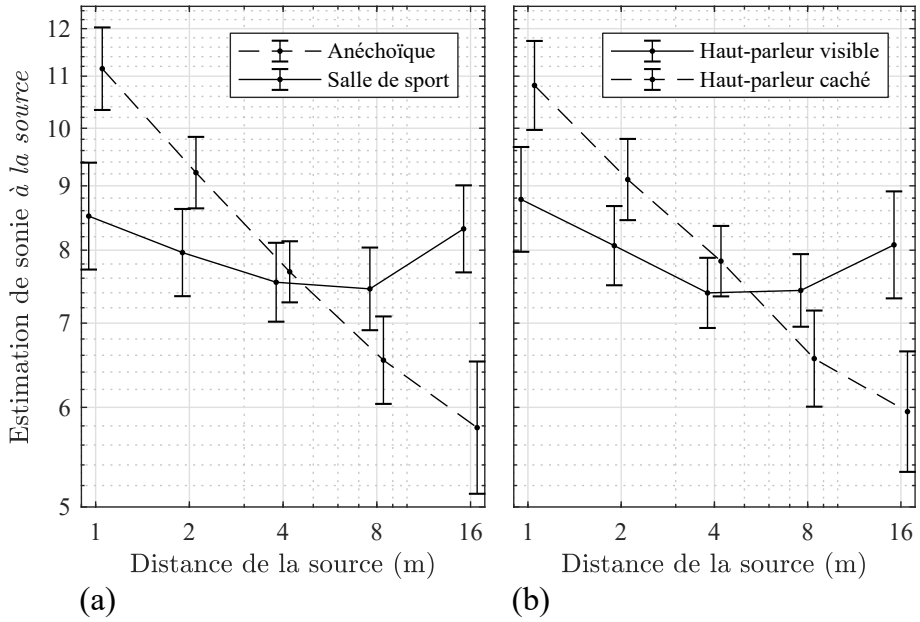


FIGURE 6.1 – Moyennes géométriques des estimations de sonie *à la source* en fonction de la distance de la source sonore effectuées dans chacun des deux environnements (a) et pour les deux conditions de visibilité (b), avec intervalles de confiance à 95%. Les moyennes et les limites des intervalles de confiance ont été calculées comme expliqué figure 5.8.

Interaction Distance × Visibilité : La figure 6.1 (b) représente cette interaction significative ($p < .001$), qui indique que la sonie *à la source* dépendait moins de la distance de la source lorsque le haut-parleur était visible (trait continu) que lorsqu'il était caché (trait discontinu). La comparaison des figures 6.1 (a) et 6.1 (b) montre que **les rôles joués par le champ réverbéré et les indices visuels sur la sonie *à la source* étaient similaires**.

Interaction Salle × Distance × Visibilité : La figure 6.2 représente cette interaction significative ($p < .001$), qui contient les résultats majeurs de cette expérience. Les estimations de sonie *à la source* réalisées dans la salle anéchoïque (a) et dans la salle de sport (b) y sont représentées en fonction de la distance lorsque le haut-parleur était visible (trait continu) ou caché (trait discontinu). Les courbes obtenues avec indices visuels et/ou champ réverbéré semblent moins pentues (donc moins dépendantes de la distance) que la courbe obtenue en environnement

Environnement	Visibilité	Exposant b	MSE
Anéchoïque	Visible	-0.11	0.1661
	Caché	-0.37	0.0291
Réverbérant	Visible	0.03	0.2286
	Caché	-0.07	0.1114

TABLE 6.3 – Exposants b des fonctions puissance $L = k \cdot r^b$ obtenues par régression sur les résultats d'estimation de sonie *à la source* montrant l'évolution de la sonie avec la distance de la source sonore dans chacun des environnements lorsque le haut-parleur était soit visible, soit caché. Un exposant $b = 0$ témoigne d'une indépendance de la sonie *à la source* vis-à-vis de la distance de la source.

anéchoïque sans indices visuels. Chacune de ces courbes a été soumise à une régression sur une fonction puissance de forme $L = k \cdot r^b$ (voir section 5.3) afin d'en étudier l'exposant b. Les valeurs des exposants b ainsi calculées sont présentées dans le tableau 6.3.

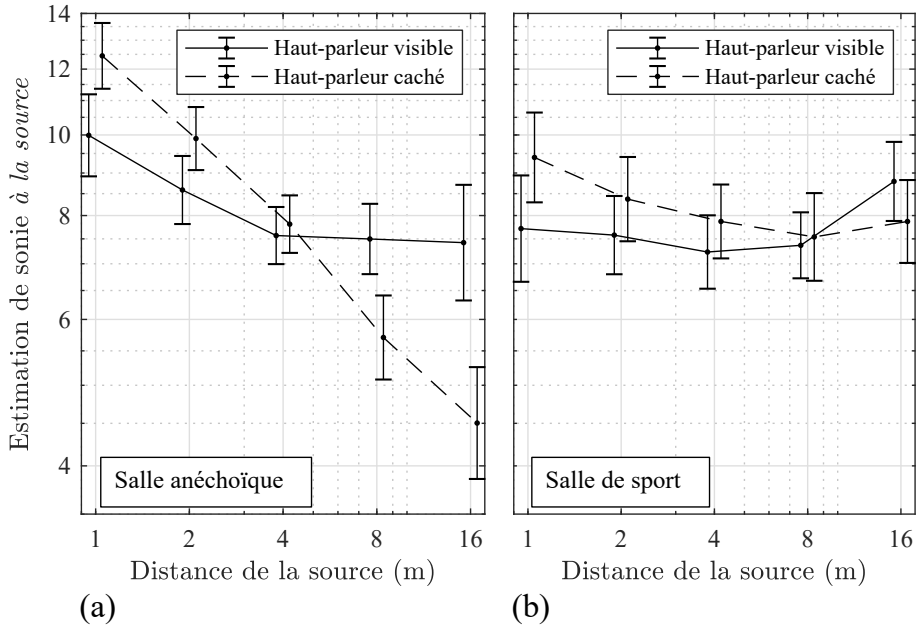


FIGURE 6.2 – Moyennes géométriques des estimations de puissance en fonction de la distance de la source sonore effectuées dans chacun des deux environnements pour chacune des deux conditions de visibilité du haut-parleur, avec intervalles de confiance à 95%. Les moyennes et les limites des intervalles de confiance ont été calculées comme expliqué figure 5.8.

La régression effectuée sur la courbe d'estimation de sonie *à la source* en fonction de la distance de la source sonore obtenue sans indices visuels en salle anéchoïque a un exposant $b = -0.37$. Les estimations étaient alors semblables aux estimation de sonie *aux oreilles* (obtenues dans la section 5.3) puisque le seul indice disponible était le niveau de pression aux oreilles, dépendant de la distance.

Lorsque le haut-parleur était visible dans l'environnement anéchoïque, les estimations de sonie *à la source* devenaient moins dépendantes de la distance de la source. La régression estime un coefficient $b = -0.11$, plus faible (en valeur absolue) que ceux observés jusqu'à présent.

Les coefficients obtenus après régression dans l'environnement réverbérant sont $b = -0.07$ sans indices visuels et $b = 0.03$ avec indices visuels, tous deux proches de zéro.

Des tests post-hoc LSD de Fisher avec correction de Bonferroni ont été effectués sur ces résultats afin de valider ou non la présence de constance de sonie dans chacune de ces conditions expérimentales. Les estimations de sonie *à la source* obtenues à chacune des cinq distances ont été comparées deux-à-deux dans chaque environnement et pour chaque condition de visibilité du haut-parleur. D'après ces tests, les estimations de sonie *à la source* étaient constantes avec la distance **dans l'environnement réverbérant que la source soit visible ou cachée** et **dans l'environnement anéchoïque lorsque la source était visible**. Dans ces trois conditions, aucune différence significative n'a été relevée entre les estimations réalisées à des distances différentes. En environnement anéchoïque lorsque le haut-parleur était caché, chaque comparaison deux-à-deux des résultats obtenus aux cinq distances reportait une différence significative ($p = 0.019$ pour la différence la moins significative, 4 m vs. 8 m).

Autres résultats

Effet simple du Niveau : Plus les sons étaient joués forts, plus la sonie *à la source* était forte ($p < .001$). Des tests post-hoc LSD de Fisher avec correction de Bonferroni montrent que les estimations de sonie *à la source* étaient significativement différentes pour les trois niveaux de diffusion ($p < .001$ pour chacune des trois comparaison deux à deux).

Interaction Salle × Niveau : La figure 6.3 (a) représente cette interaction significative ($p = .018$). Les estimations semblent varier sur une dynamique légèrement plus grande dans l'environnement réverbérant que dans l'environnement anéchoïque. Des tests post-hoc LSD de Fisher avec correction de Bonferroni n'indiquent cependant pas de différence significative entre les estimations obtenues dans les deux environnements.

Interaction Salle × Visibilité : La figure 6.3 (b) représente cette interaction significative ($p = .005$). La sonie *à la source* semble plus forte lorsque la source était cachée que lorsqu'elle était visible dans l'environnement réverbérant tandis que l'inverse semble avoir eu lieu dans l'environnement anéchoïque. Des tests post-hoc LSD de Fisher avec correction de Bonferroni n'indiquent cependant pas de différence significative entre les estimations obtenues pour les deux conditions de visibilité.

Interaction Distance × Niveau × Visibilité : La figure 6.4 représente cette interaction significative ($p = .008$). Chacune des trois cases représente les estimations de sonie *à la source* effectuées lorsque les sons étaient diffusés aux niveaux −12 dB (a), −6 dB (b) et 0 dB (c). Dans chacune de ces cases, les estimations de sonie *à la source* apparaissent en fonction de la distance de la source sonore, lorsque cette dernière était soit visible (trait continu) soit cachée (trait discontinu). Cette interaction n'apporte cependant pas d'informations supplémentaires vis-à-vis de cette étude comparée à l'interaction « Distance × Visibilité ».

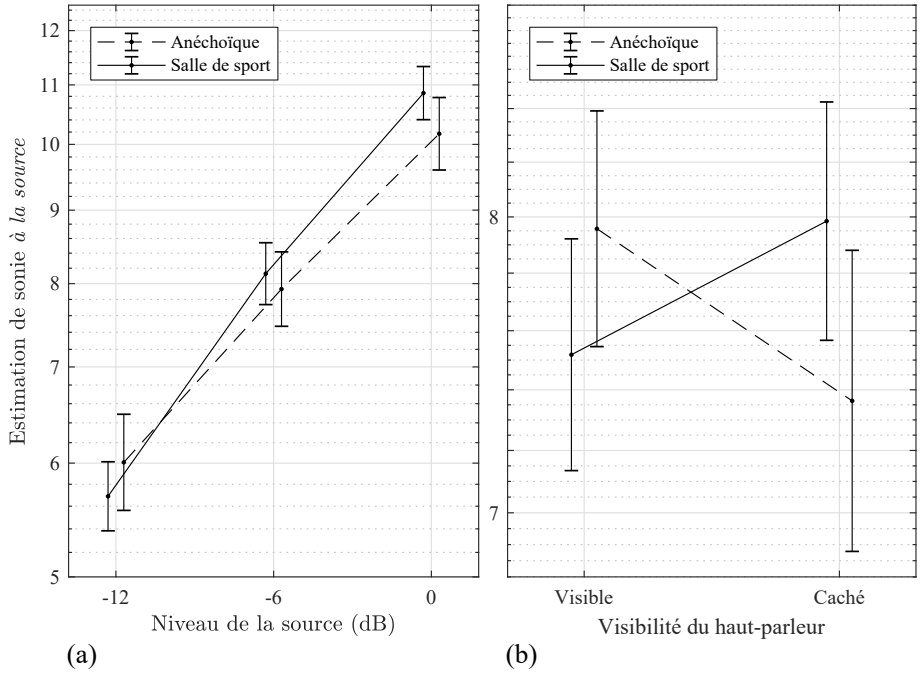


FIGURE 6.3 – Moyennes géométriques des estimations de sonie *à la source* en fonction du niveau de la source (a) et de sa visibilité (b) dans chacun des deux environnements, avec intervalles de confiance à 95%. Les moyennes et les limites des intervalles de confiance ont été calculées comme expliqué figure 5.8.

Bilan des résultats

En résumé, les résultats de cette expérience attestent que :

- Le rôle joué par les indices visuels (fournissant une information de distance) était identique à celui joué par le champ réverbéré (fournissant des informations de distance et de puissance) dans les estimations de sonie *à la source*.
- Les estimations de sonie *à la source* réalisées en environnement anéchoïque étaient constantes en fonction de la distance de la source lorsque cette dernière était visible.
- Les estimations de sonie *à la source* réalisées dans l'environnement réverbérant étaient constantes avec la distance de la source que cette dernière soit visible ou cachée.

Conformément aux résultats de Zahorik et Wightman (2001), la présente étude observe des estimations de sonie *à la source* indépendantes de la distance dans un environnement réverbérant. D'après les auteurs, les estimations de sonie *à la source* seraient basées sur l'énergie du champ réverbérant, laquelle fournit une information directe de puissance de la source sonore. Les résultats de la présente étude ne remettent pas en question cette hypothèse mais révèlent également une constance de sonie *à la source* en absence de champ réverbérant, lorsque la source sonore était visible par les auditeurs. Cette observation prouve la capacité des auditeurs à mettre en commun le niveau sonore et la distance visuellement perçue pour réaliser des estimations de sonie *à la source*. Typiquement, si les auditeurs entendaient un son faible mais qu'ils voyaient la source lointaine, ils en déduisaient que la source était puissante. Pourtant, les estimations de

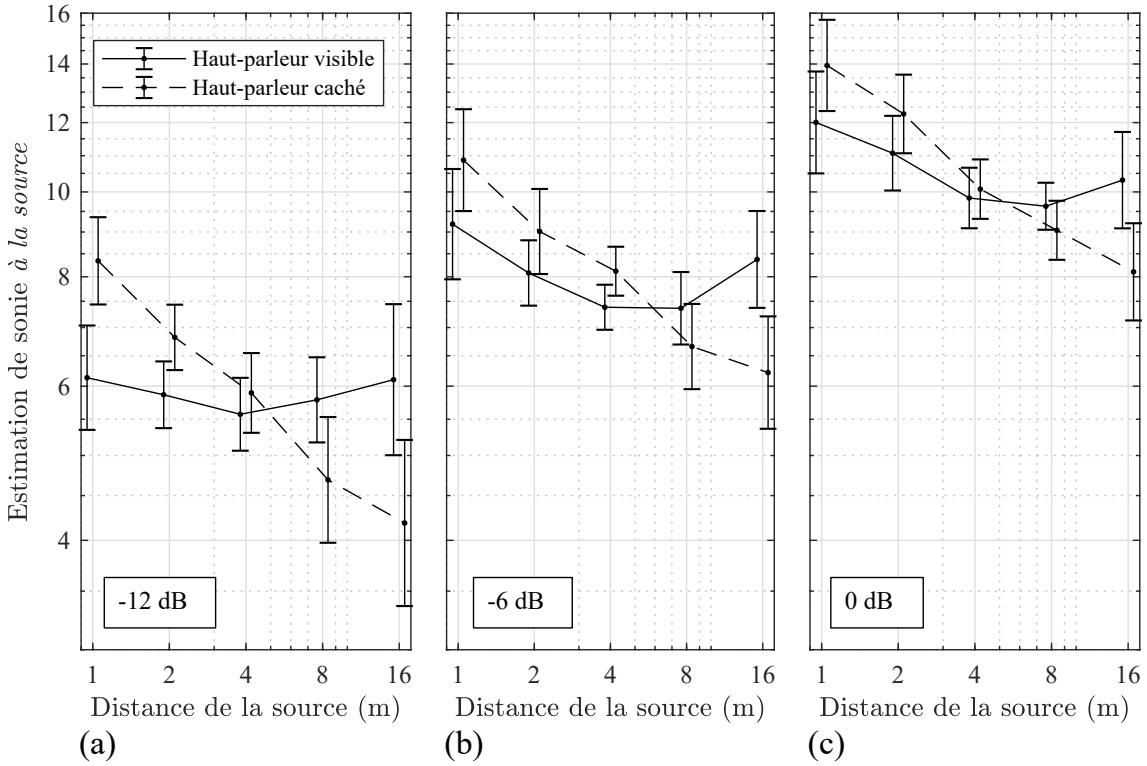


FIGURE 6.4 – Moyennes géométriques des estimations de puissance en fonction de la distance de la source sonore effectuées pour chaque niveau de diffusion et pour chacune des deux conditions de visibilité du haut-parleur, avec intervalles de confiance à 95%. Les moyennes et les limites des intervalles de confiance ont été calculées comme expliqué figure 5.8.

distance réalisées dans les mêmes conditions expérimentales lors de l'étude présentée dans le chapitre 5 n'étaient pas exactes. Ces erreurs pourraient être dues au protocole utilisé (l'estimation directe en mètres), donnant une estimation incorrecte des distances perçues (qui auraient été plus proches des distances réelles).

L'étude suivante, qui clôture cette thèse, s'intéresse à la sonie *aux oreilles* et à la sonie *à la source* de parole. Ce signal sonore, contrairement au bruit blanc, a la particularité de véhiculer une information de puissance sonore de sa source (via l'effort vocal) ayant potentiellement des effets sur les relations entre sonie et distance.

Chapitre 7

Sonie *aux oreilles* et sonie *à la source* de signaux de parole en fonction de la distance

7.1 Introduction

Les deux études précédentes ont observé l'évolution de la sonie *aux oreilles* et de la sonie *à la source* de bruit blanc en fonction de la distance séparant la source des auditeurs. Lorsque les bruits étaient diffusés dans un environnement réverbérant, les indices de distance et de puissance de la source fournis par le champ réverbéré permettaient aux auditeurs d'estimer la sonie *à la source* de façon indépendante de la distance. Ce même constat était réalisé lorsque les sources sonores étaient visibles, même en l'absence de champ réverbéré. En revanche, les estimations de sonie *aux oreilles* dépendaient pleinement de la distance de la source sonore (et donc du niveau de pression sonore aux oreilles des auditeurs) dans l'ensemble des conditions de présentation des sons (en environnement anéchoïque ou réverbérant, lorsque la source était visible ou cachée).

Le lien entre niveau de pression sonore et sonie semble dépendre du type de stimulus. Dans l'étude de Mohrmann (1939), des sons non-familiers (des sons purs et des bruits) et des sons familiers (de la parole et de la musique) étaient diffusés depuis différentes distances du point d'écoute. Les participants devaient ajuster le niveau d'une source pour que la sonie des sons qu'elle diffusait corresponde à la sonie de sons diffusés depuis une distance différente. Deux ajustements différents étaient réalisés, l'un ayant pour objectif d'égaliser la sonie *à la source* et l'autre d'égaliser la sonie *aux oreilles*. Lorsque les sons étaient familiers, les auteurs ont relevé de meilleures performances sur les ajustements de sonie *à la source* (après ajustement, les puissances des deux sources étaient relativement proches) que sur les ajustements de sonie *aux oreilles* (après ajustement, les niveaux au point d'écoute étaient relativement éloignés). Ce constat n'a pas été réalisé lorsque les sources diffusaient des sons non-familiers. Par ailleurs, les observations faites par Pollack (1952) et Warren (1973) témoignent d'un traitement différent du niveau sonore aux oreilles dans la sonie pour les sons familiers ou non-familiers. D'après ces études, la sonie de signaux de parole était moins dépendante du niveau sonore aux oreilles que la sonie de sons non-familiers tels que des bruits ou des sons purs. Les auteurs ne dirigeaient pas explicitement l'attention des participants vers la sonie *aux oreilles* ou la sonie *à la source*.

La parole, contrairement aux bruits ou aux sons purs, contient une information intrinsèque de l'effort vocal du locuteur, renseignant sur la puissance sonore (voir section 1.2.3). Si l'on prend un exemple extrême, la source sonore d'un chuchotement sera perçue comme étant moins puissante que celle d'un hurlement, et ce indépendamment du niveau sonore aux oreilles de l'auditeur. L'évolution de la sonie *à la source* de parole avec la distance est alors susceptible de différer de celle observée pour du bruit.

D'après ces différentes observations, les estimations de sonie *à la source* et de sonie *aux oreilles* pourraient hypothétiquement varier différemment avec la distance de la source selon que celle-ci diffuse du bruit ou de la parole. La présente étude a pour objectif d'étudier les sonies *à la source* et *aux oreilles* de voix parlée à différentes distances d'auditeurs afin de vérifier cette hypothèse. Comme dans les expériences des chapitres 5 et 6, la source sonore (en l'occurrence un locuteur) pourra être soit visible, soit cachée afin d'observer le rôle des indices visuels de distance dans les estimations de sonie.

7.2 Matériel et méthode

Le protocole expérimental de cette expérience ressemble de près à celui des études présentées dans les chapitres 5 et 6. Les participants se trouvaient dans la même cabine audiométrique que lors de ces études et l'environnement virtuel était similaire et reproduit sur les mêmes appareils. En revanche, les stimuli sonores et visuels n'étaient pas les mêmes. Dans cette étude, un locuteur prononçait des mots à différents efforts vocaux, contrairement aux deux études précédentes dans lesquelles un haut-parleur diffusait des bruits blancs.

7.2.1 Stimuli auditifs

Deux acoustiques différentes ont été étudiées, l'une étant anéchoïque et l'autre réverbérante (la salle de sport).

Les stimuli auditifs étaient des mots prononcés par un locuteur à différentes distances du point d'écoute. La distance séparant le locuteur du point d'écoute a été créée virtuellement. Les mots ont dans un premier temps été enregistrés dans une cabine d'enregistrement (quasi-anéchoïque) puis ont été placés à différentes distances grâce au procédé expliqué en annexe A. Lors de l'enregistrement, le locuteur était assis sur une chaise et devait prononcer des mots à trois niveaux sonores, produisant 68, 74 et 80 dB SPL à 1 m, donnant lieu à trois efforts vocaux différents. Ainsi, les niveaux d'émission des mots étaient les mêmes que les niveaux d'émission des bruits lors des expériences des chapitres 5 et 6. Les mots étaient enregistrés en proximité par un microphone (DPA 4006A) placé à 15 cm de la bouche du locuteur. Un filtre anti-pop était placé entre le locuteur et ce microphone afin d'éliminer les bruits de bouche. Un second microphone (DPA 4006A également) placé à 1 m de la bouche du locuteur mesurait le niveau sonore produit par sa voix. D'après ces mesures, un écran affichait en temps réel des indications permettant au locuteur d'ajuster son effort vocal pour produire les niveaux requis. Un schéma de ce dispositif d'enregistrement est représenté figure 7.1. La figure 7.2 représente une photographie

du locuteur pendant la séance d'enregistrement. L'acquisition des sons était réalisée via une carte son RME Fireface 800. Les sons étaient encodés sur 24 bits avec une fréquence d'échantillonnage de 44.1 kHz et enregistrés sur le logiciel de DAW (*Digital Audio Workstation*) Reaper.

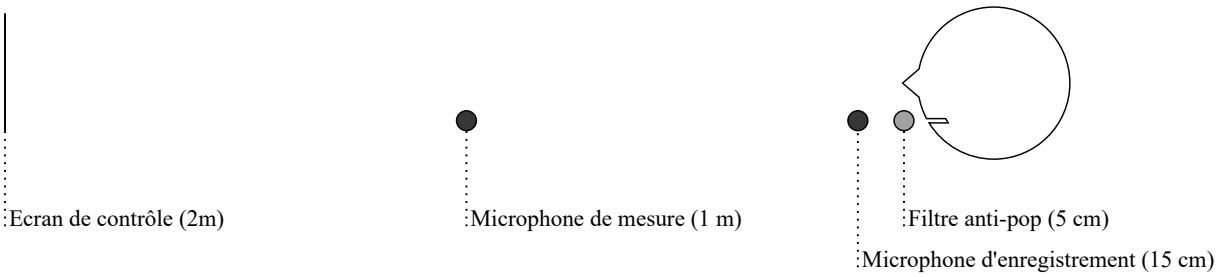


FIGURE 7.1 – Schéma du dispositif mis en place pour la prise de son des signaux de parole. Le locuteur, à droite, prononce les mots aux niveaux indiqués (68 dB SPL, 74 dB SPL et 80 dB SPL), contrôlés par le microphone de mesure placé à 1 m de sa bouche. L'écran de contrôle lui indiquait si le mot était prononcé pas assez fort, trop fort ou au bon niveau. La voix était enregistrée par un microphone placé à 15 cm.

FIGURE 7.2 – Photographie du locuteur pendant l'enregistrement des stimuli sonores.

Les mots enregistrés étaient des mots dissyllabiques, proches des *spondees* utilisés par Epstein et Florentine (2009) dans leur étude portant sur la sommation de sonie binaurale de voix parlée. Trois mots différents ont été sélectionnés afin de multiplier les stimuli pour éviter que les participants ne réalisent leurs estimations par identification : si un seul mot était prononcé, les participants auraient pu identifier que les sons étaient prononcés à trois efforts vocaux distincts, ce qui aurait pu résulter en une discrétisation des réponses. Les trois mots sélectionnés étaient « réchaud », « normand » et « caveau », extraits des listes dissyllabiques de Fournier (1951) notamment utilisées en audiométrie vocale. La sortie du microphone de mesure, placé à 1 m de la

bouche du locuteur, était analysée par un programme Max/MSP qui contrôlait le niveau sonore des mots prononcés. Ce programme indiquait en temps réel au locuteur, grâce à des signaux lumineux, s'il produisait un niveau trop faible, correct ou trop fort. Une marge de tolérance de ±1 dB était définie autour de chacun des niveaux à produire. Par exemple, si le locuteur devait prononcer un mot en produisant un niveau sonore de 74 dB SPL à 1 m, le programme acceptait des niveaux mesurés allant de 73 dB SPL à 75 dB SPL. Le programme contrôlait également la dynamique des mots prononcés.

Plusieurs répétitions d'un même mot à un même effort vocal était enregistrées. Après la session d'enregistrement, un seul enregistrement était sélectionné par l'expérimentateur pour chaque combinaison de mot et d'effort vocal. Le choix était porté sur une évaluation subjective, faite par l'expérimentateur, de la clarté de prononciation du mot. La forme d'onde des neuf mots ainsi sélectionnés est représentée figure 7.3, où chaque ligne correspond à un mot différent et chaque colonne à un effort vocal différent.

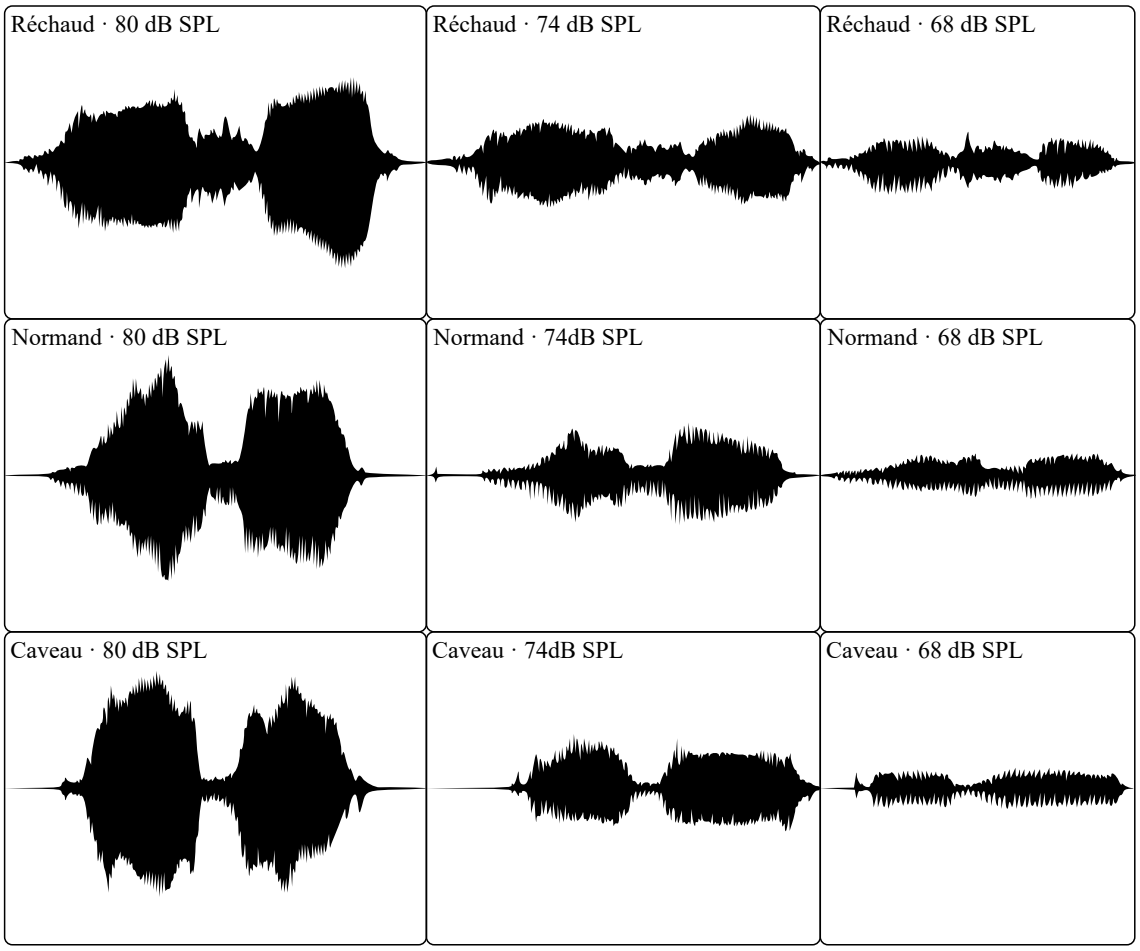


FIGURE 7.3 – Formes d'onde de chacun des neuf stimuli sonores, correspondant aux trois mots dissyllabiques (réchaud, normand et caveau) exprimés à trois efforts vocaux différents (produisant 80 dB SPL, 74 dB SPL et 68 dB SPL à un mètre de la bouche de locuteur).

Puisque le programme autorisait une certaine marge de tolérance lors de la mesure du niveau sonore, les niveaux des enregistrements des trois mots ont été égalisés pour chaque effort vocal donné. Enfin, les stimuli auditifs binauraux ont été synthétisés à partir de ces enregistrements par la procédure décrite en annexe A et restitués sur un casque étalonné.

7.2.2 Stimuli visuels

Lors des expériences précédentes, détaillées dans les chapitres 5 et 6, les sons étudiés étaient des bruits large bande diffusés par des haut-parleurs virtuels. Un haut-parleur avait alors été modélisé en 3D pour être placé dans l'environnement virtuel (voir figures 5.1 et 5.3). Dans la présente étude, les sons étudiés étaient des mots prononcés par un locuteur. Les expressions du visage et du corps sont différentes selon l'effort vocal produit, constituant une information visuelle de puissance de la source sonore. D'après les résultats de Rosenblum et Fowler (1991), cette information visuelle pourrait avoir une influence sur la sonie de parole. Une représentation visuelle du locuteur a de ce fait été implémentée dans l'environnement virtuel. Pour ce faire, le locuteur a été filmé sur un fond bleu pendant qu'il prononçait à nouveau les mots enregistrés précédemment. Une photographie du plateau de tournage est disponible figure 7.4. Chaque mot enregistré était joué au locuteur, qui avait pour objectif de reproduire le même effort vocal lors de l'enregistrement vidéo. Les scènes ont été filmées avec une caméra de cinéma Blackmagic URSA Mini Pro.

FIGURE 7.4 – Photographie du plateau sur lequel les stimuli visuels ont été tournés.

Un certain nombre d'enregistrements vidéo étaient réalisés pour chaque mot à chaque effort vocal. Après la session de tournage, une seule de ces vidéos était sélectionnée par l'expérimentateur pour chaque combinaison de mot et d'effort vocal. La vidéo sélectionnée était celle qui correspondait (subjectivement) le mieux au mot enregistré lorsque les enregistrements audio et

vidéo étaient joués simultanément (en *playback*). Chaque enregistrement vidéo a ensuite été détouré (pour effacer le fond bleu) sur le logiciel DaVinci Resolve. Enfin, les enregistrements vidéo étaient incrustés dans l'environnement virtuel. La vidéo du locuteur était affichée sur un plan positionné à la distance souhaitée du point d'écoute (1 m, 2 m, 4 m, 8 m ou 16 m). La figure 7.5 montre les différentes étapes de ce procédé, qui comporte la vidéo originale (a), la vidéo détourée (b) et la vidéo incrustée (c). La vidéo correspondant au stimulus sonore à présenter aux participants était alors jouée en temps réel, synchronisée avec l'enregistrement audio.

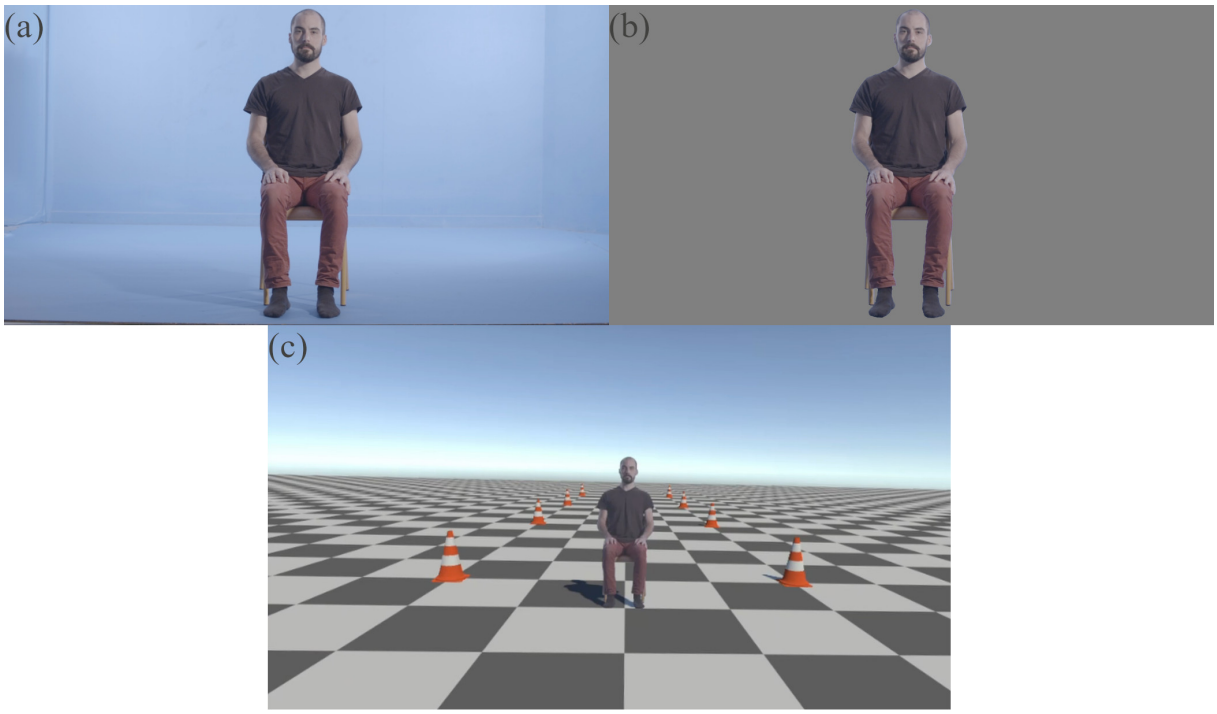


FIGURE 7.5 – Différents stades de l'incrustation des vidéos dans l'environnement virtuel. Les vidéos originales (a) sont détourées pour supprimer le fond bleu (b) puis incrustées dans l'environnement en les projetant sur un plan à la distance souhaitée du point d'observation (c).

7.2.3 Procédure

La procédure est similaire à celles des expériences précédentes, décrites dans les sections 5.2 et 6.2. L'expérience était partagée en deux sessions, l'une étudiant la sonie *aux oreilles* et l'autre la sonie *à la source* des mots prononcés par le locuteur. Dans la session d'estimation de sonie *aux oreilles*, les instructions données par l'expérimentateur indiquaient aux auditeurs « plus vous entendez le son fort, plus le nombre que vous assignez doit être élevé ». Dans la session d'estimation de sonie *à la source*, les instructions données par l'expérimentateur indiquaient aux auditeurs « plus le locuteur a parlé fort, plus le nombre que vous assignez doit être élevé ».

Chaque session était découpée en deux sous-sessions, correspondant à chacun des deux environnements (l'environnement anéchoïque et la salle de sport). Chaque combinaison de variables

expérimentales était présentée deux fois aux participants. Cent quatre-vingts estimations de sonie étaient alors réalisées dans chacune des quatre sous-sessions par chaque participant. Le tableau 7.1 présente chaque facteur de l'expérience associé à ses différents niveaux. Les estimations de sonie *aux oreilles* et de sonie *à la source* étaient toutes deux effectuées par un protocole d'estimation de grandeur sans référence (Marks et Florentine, 2011) identique à celui utilisé lors des deux expériences précédentes. Les participants répondaient à l'aide de la tablette tactile et leurs estimations étaient précédées de la mention « Sensation : » lors de la session d'estimation de sonie *aux oreilles* et de la mention « Effort vocal : » lors de la session d'estimation de sonie *à la source*. Les participants pouvaient rejouer les stimuli à leur guise. Le reste de la procédure était identique à celles des expériences détaillées dans les chapitres 5 et 6.

Tâche (session)	Sonie *aux oreilles* · Sonie *à la source*
Facteur	**Niveaux**
Salle (sous-session)	Anechoïque · Salle de sport
Distance	1 m · 2 m · 4 m · 8 m · 16 m
Niveau de diffusion	0 dB · −6 dB · −12 dB
Locuteur	Visible · Caché
Mot	Réchaud · Normand · Caveau
Répétition	1 · 2

Table 7.1 – Résumé des conditions expérimentales évaluées par chaque participant.

Dix-sept participants (six femmes et onze hommes âgés de vingt à vingt-six ans) ne reportant pas de problème d'audition ont pris part à l'expérience. Les participants portant des lunettes pouvaient les garder durant l'expérience sans que cela ne pose de problème vis-à-vis du HMD. L'écart inter-pupille des participants était mesuré au début de la première session et le HMD était configuré en conséquence. Aucun des participants n'avait d'expérience en tests psychoacoustiques. Les instructions étaient inscrites sur le même panneau que présenté figure 5.7 et étaient également lues par l'expérimentateur au début de chaque session, qui s'assurait que les participants avaient compris quelle grandeur évaluer. Les deux sessions étaient réparties sur deux semaines différentes pour chaque participant. Un pré-test d'une dizaine d'essais était réalisé au début de chacune des quatre sous-sessions afin de familiariser les participants avec les stimuli et la procédure. L'ordre dans lequel les participants réalisaient chaque session et chaque sous-session était aléatoire. Les participants avaient la possibilité de retirer le HMD et de sortir de la salle d'expérimentation entre chaque sous-session. La durée totale de l'expérience était d'environ deux heures par participant (quatre sessions d'environ vingt minutes, plus les temps consacrés à la mise en place du matériel de restitution, aux instructions, aux pré-tests et aux pauses). Un enregistrement vidéo de l'expérience est disponible à l'adresse `https://youtu.be/F3xAx_j0YZw`.

7.3 Résultats & Discussion

Les estimations de sonie ont été traitées suivant la procédure détaillée dans la section 5.2 (Altmann *et al.*, 2012). Une ANOVA à mesures répétées a été conduite sur les logarithmes normalisés des résultats de chacune des deux sessions. Cinq facteurs ont été inclus dans chacune des deux analyses : la salle (deux niveaux), la distance de la source (cinq niveaux), le niveau d'émission de la voix (trois niveaux), la visibilité du locuteur (deux niveaux) et le mot prononcé (trois niveaux). Il est probable qu'au vu du grand nombre de variables indépendantes relativement au nombre de participants, l'analyse puisse révéler des interactions significatives peu pertinentes (Kirk, 1995). Ainsi, certaines interactions présentées dans la section **Autres résultats** sont statistiquement significatives mais ne montrent pas de variations flagrantes de la variable dépendante représentée en fonction des différents facteurs.

7.3.1 Estimations de sonie *aux oreilles*

Les résultats de l'ANOVA réalisée sur les estimations de sonie *aux oreilles* sont présentés tableau 7.2. Les résidus du modèle linéaire étaient normalement distribués.

Résultats principaux

Effet simple de la Visibilité : Les estimations de sonie *aux oreilles* étaient globalement plus élevées lorsque le locuteur était caché que lorsqu'il était visible ($p = .003$), comme le montre la figure 7.6.

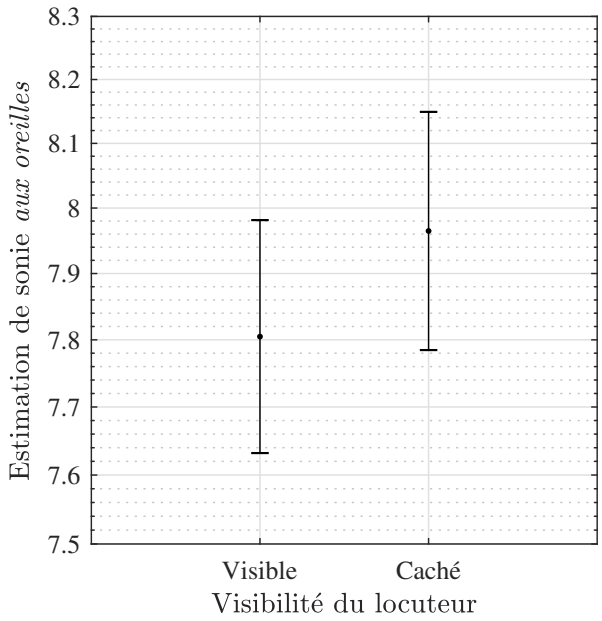


FIGURE 7.6 – Moyennes géométriques des estimations de sonie *aux oreilles* lorsque le locuteur était soit visible, soit caché. Les moyennes géométriques sont représentées avec intervalles de confiance à 95%. Les moyennes et les limites des intervalles de confiance ont été calculées comme expliqué figure 5.8.

Variable dépendante		**Sonie *aux oreilles***			
Source	**SS**	**DF**	**MS**	**F**	**Sig.p**
S	9.873	1	9.873	$1.867*10^8$	**< .001**
D	44.402	1.096*	40.502	69.48	**< .001**
M	.214	2	.107	17.132	**< .001**
N	21.653	1.09*	19.872	86.924	**< .001**
V	0.59	1	.059	12.613	**.003**
S × D	4.75	1.493*	3.181	39.47	**< .001**
S × M	.063	2	.031	5.831	**.007**
S × N	.255	2	.128	9.275	**.001**
S × V	.004	1	.004	.338	.569
D × M	.101	4.444*	.023	2.013	.095
D × N	.31	3.603*	.086	3.071	**.027**
D × V	.032	4	.008	1.016	.406
M × N	.036	4	.009	1.173	.331
M × V	.025	2	.012	3.224	.053
N × V	.001	2	.001	.136	.873
S × D × M	.064	3.522*	.018	1.105	.36
S × D × N	.084	3.953*	.021	1.422	.237
S × D × V	.059	4	.015	1.467	.223
S × M × N	.047	2.542*	.018	2.186	.114
S × M × V	.019	2	.009	2.155	.132
S × N × V	.006	1.366*	.005	.528	.53
D × M × N	.124	6.611*	.019	1.205	.308
D × M × V	.059	3.934*	.015	1.228	.308
D × N × V	.055	3.795*	.014	.727	.57
M × N × V	.012	4	.003	.52	.721
S × D × M × N	.058	5.775*	.01	.677	.663
S × D × M × V	.031	3.884*	.008	.558	.689
S × D × N × V	.037	3.966*	.009	.787	.537
S × M × N × V	.005	4	.001	.266	.899
D × M × N × V	.056	4.916*	.011	.453	.807
S × D × M × N × V	.133	6.106*	.022	1.265	.28

* Le degré de liberté a été ajusté par une correction de Greenhouse-Geisser suite à une violation de l'hypothèse de sphéricité des résultats.

TABLE 7.2 – Résultats de l'ANOVA conduite sur les logarithmes des estimations de sonie *aux oreilles*. S désigne la salle, D la distance de la source, N le niveau de diffusion de la source (assimilable à l'effort vocal du locuteur), M le mot prononcé par le locuteur et V la visibilité du locuteur.

Effet simple de la Distance : Les estimations de sonie *aux oreilles* étaient plus élevées lorsque le locuteur était proche des auditeurs ($p < .001$). Ce résultat témoigne d'une absence de constance de sonie sur les résultat globaux.

Interaction Salle × Distance : Les estimations de sonie *aux oreilles* n'évoluaient pas de la même manière avec la distance du locuteur dans les deux environnements ($p < .001$). Cette interaction, représentée figure 7.7, montre une décroissance plus faible de la sonie *aux oreilles* avec la distance dans l'environnement réverbérant que dans l'environnement anéchoïque. Ces deux courbes ont été soumises à une régression sur une fonction puissance de type $L = k \cdot r^b$ (voir section 5.3), donnant des exposants $b = -0.37$ (MSE = 0.0364) dans l'environnement anéchoïque et $b = -0.2$ (MSE = 0.1266) dans l'environnement réverbérant. Ces deux exposants sont relativement éloignés de zéro. Des tests post-hoc LSD de Fisher avec correction de Bonferroni

ont été effectués pour comparer deux-à-deux les estimations obtenues à chacune des cinq distances dans ces deux salles. Chacune de ces comparaisons a révélé une différence significative (la valeur de p la plus élevée est $p = 0.002$, obtenue lors de la comparaison des estimations aux distances 8 m et 16 m dans la salle de sport). Ainsi, même si l'évolution de la sonie *aux oreilles* avec la distance n'était pas la même dans les deux environnements, aucune des deux ne présente de constance de sonie. Les différences observées entre les deux environnements sont susceptibles d'être dues uniquement au niveau sonore aux oreilles des auditeurs. Ce niveau était plus élevé dans la salle de sport que dans l'environnement anéchoïque, surtout aux distances lointaines, puisque l'énergie du champ réverbéré décroît peu avec la distance (voir annexe B).

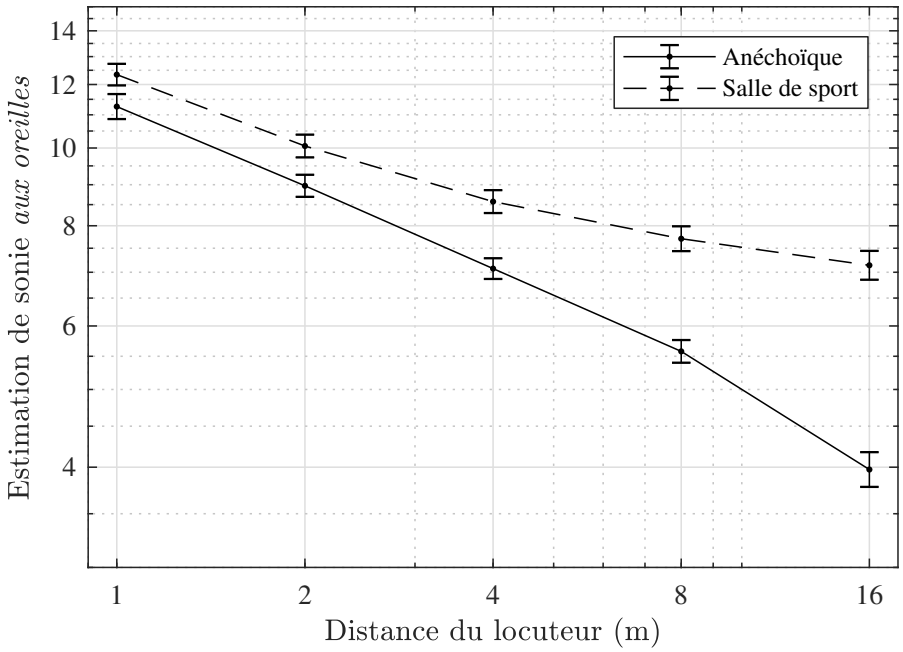


FIGURE 7.7 – Moyennes géométriques des estimations de sonie *aux oreilles* en fonction de la distance du locuteur dans chacun des deux environnements. Les moyennes géométriques sont représentées avec intervalles de confiance à 95%. Les moyennes et les limites des intervalles de confiance ont été calculées comme expliqué figure 5.8.

Autres résultats

Effet simple de la Salle : Les estimations de sonie *aux oreilles* étaient globalement plus élevées pour les sons présentés dans l'environnement réverbérant que pour ceux présentés dans l'environnement anéchoïque ($p < .001$). Cet effet a pour cause le niveau sonore aux oreilles des auditeurs, qui était globalement plus élevé dans l'environnement réverbérant à cause du champ réverbéré (voir annexe B).

Effet simple du Niveau : Cet effet indique que la sonie *aux oreilles* était plus élevée lorsque les mots étaient prononcés plus forts ($p < .001$).

Effet simple du Mot : Les différents mots prononcés par le locuteur ont mené à des estimations de sonie *aux oreilles* différentes ($p < .001$).

Interaction Salle × Mot : Cette interaction est représentée figure 7.8 (a). La dépendance des estimations de sonie *aux oreilles* envers le mot prononcé était plus marquée dans la salle de sport que dans l'environnement anéchoïque. Ces différences pourraient venir de la salle de sport, qui aurait répondu de façon plus importante à la sollicitation provoquée par les mots « réchaud » et « caveau » que par celle provoquée par le mot « normand ». La réponse de la salle pourrait également être responsable de l'effet simple du mot.

Interaction Salle × Niveau : Cette interaction significative ($p = .001$) est représentée figure 7.8 (b), mais ne montre pas d'évolution très différente des estimations de sonie *aux oreilles* en fonction du niveau de prononciation des mots dans les deux salles.

Interaction Distance × Niveau : Cette interaction significative ($p = .027$) est représentée figure 7.8 (c). Cette figure montre l'absence de constance de sonie pour les trois niveaux de prononciation des mots. Elle ne montre cependant pas de grande différence entre les évolutions des estimations de sonie *aux oreilles* avec la distance du locuteur pour les trois niveaux de prononciation des mots.

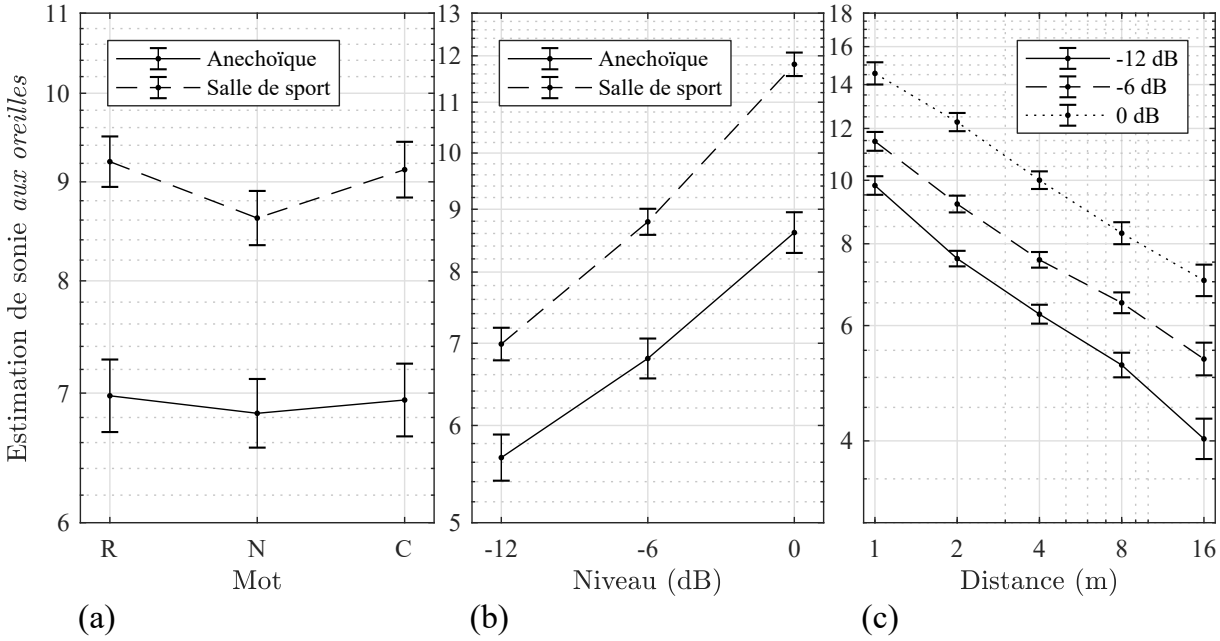


FIGURE 7.8 – Moyennes géométriques des estimations de sonie *aux oreilles* en fonction du mot prononcé (R pour « réchaud », N pour « normand » et C pour « caveau ») dans chacune des salles (a), du niveau de prononciation des mots dans chacune des salles (b) et de la distance du locuteur à chacun des niveaux de prononciation des mots (c). Les moyennes géométriques sont représentées avec intervalles de confiance à 95%. Les moyennes et les limites des intervalles de confiance ont été calculées comme expliqué figure 5.8.

Bilan sur la sonie *aux oreilles*

Les résultats de cette étude indiquent que les estimations de sonie *aux oreilles* de signaux de parole étaient plus élevées lorsque le locuteur était caché que lorsqu'il était visible. Cet effet n'a cependant pas interagi avec la distance du locuteur, ne montrant pas de dépendance de la sonie *aux oreilles* avec la distance perçue.

Dans ces deux conditions, la sonie *aux oreilles* décroissait avec la distance de la source sonore. Cette décroissance était du même ordre de grandeur que celle observée pour des bruits blancs (voir section 5.3). Les estimations de sonie *aux oreilles* étaient alors dirigées par le niveau aux oreilles des auditeurs, lequel variait avec la distance de la source. Ce résultat s'éloigne des observations faites par Pollack (1952) et Warren (1973), selon lesquelles la sonie de parole était moins dépendante du niveau aux oreilles que la sonie de sons non-familiers tels que les bruits blancs utilisés dans le chapitre 5. Les protocoles utilisés et les instructions données aux participants de ces deux études étaient cependant différents de ceux de la présente expérience par plusieurs points :

- Les participants de l'étude de Pollack (1952) avaient pour objectif d'ajuster le niveau d'un signal pour qu'il atteigne la sonie (ou la moitié de la sonie, ou le double de la sonie) d'un signal de référence. Ceux de l'étude de Warren (1973) devaient estimer la sonie d'un signal en lui attribuant un score relatif à celui d'un signal identique diffusé à un niveau supérieur, dont le score était arbitrairement fixé à 100. Ces deux méthodes donnaient des estimations relatives de sonie, contrairement aux estimations absolues obtenues dans la présente étude.
- L'attention des auditeurs n'était pas explicitement portée vers la sonie du stimulus *aux oreilles*. Les participants de l'étude de Warren (1973) devaient comparer la sonie de deux sons et la question posée était « *what number would you use to describe the loudness of the fainter sound ?* », soit « quel nombre utiliseriez-vous pour décrire la sonie du son le plus faible ? ».

Si l'on pose la question à un auditeur de savoir à quel point un son « lui paraît fort », sans précision supplémentaire, plusieurs interprétations sont possibles. Pollack (1952) établit un rapprochement entre ses observations sur la sonie de voix parlée et les observations établies sur la « constance de taille » en perception visuelle (voir section 1.4.2). Prenons l'exemple d'un observateur à qui l'on demande si une personne « lui paraît grande », sans précision supplémentaire. Il est possible que cet observateur ne prenne pas en compte uniquement la taille formée par cette personne sur sa rétine (qui serait assimilable à la sonie *aux oreilles* pour un son), mais également la taille « réelle » qu'il perçoit de cette personne (assimilable à la sonie *à la source* pour un son). Si l'on demande à ce même observateur d'évaluer la taille que forme l'image de cette personne dans son champ de vision (assimilable à une estimation de sonie *aux oreilles* pour un son), son estimation pourrait être indépendante de la perception de taille réelle. Ainsi, le fait d'orienter l'attention de l'auditeur (ou de l'observateur) sur le stimulus proximal (la sonie *aux oreilles* ou la taille de l'image sur la rétine) ou sur le stimulus distal (la sonie *à la source* ou la taille réelle

de l'objet) pourrait mener à des estimations différentes de celles obtenues en laissant l'auditeur (ou l'observateur) interpréter librement la consigne.

De plus, Pollack (1952) a pris l'exemple d'un locuteur se déplaçant dans une pièce tout en parlant pour expliquer ses résultats (voir section 1.3.6). Le niveau de pression du signal de parole aux oreilles d'un auditeur donné variera en accord avec les déplacements du locuteur. Pourtant, d'après l'auteur, cet auditeur ne percevra pas de fluctuation considérable sur la voix du locuteur. Les signaux de parole sont traités de manière particulière par le cerveau humain (Moore, 2012d). Par exemple, si un signal de parole est présenté avec des distorsions sévères (ajout de bruit, modification du spectre, écrêtage, etc.), le cerveau est capable de le recomposer et d'y reconnaître des syllabes, mots ou phrases. Les faibles dépendances de la sonie observée par Pollack (1952) et Warren (1973) envers le niveau sonore de signal de parole pourraient être issues de phénomènes similaires. Les auditeurs comparaient deux signaux de paroles (des extraits de phrase) dont l'unique différence était le niveau sonore. Le cerveau pourrait identifier les signaux comme étant identiques et « compenser » les variations de niveaux perçues, entraînant une variation de sonie relativement faible.

7.3.2 Estimations de sonie *à la source*

Les résultats de l'ANOVA réalisée sur les estimations de sonie *à la source* sont présentés dans le tableau 7.3. Les résidus du modèle linéaire étaient normalement distribués.

Résultats principaux

La visibilité de la source sonore n'a pas eu d'effet simple significatif sur les estimations de sonie *à la source* de voix parlée, mais a interagi significativement avec certains autres facteurs expérimentaux.

Effet simple de la Distance : Cet effet significatif ($p < .001$) indique que la sonie *à la source* était plus forte lorsque le locuteur était proche des auditeurs. L'évolution des estimations avec la distance sera analysée plus en détail sur les interactions significatives faisant intervenir ce facteur.

Interaction Distance × Visibilité : Les estimations de sonie *à la source* évoluaient différemment avec la distance du locuteur selon que ce dernier était visible ou caché ($p < .001$).

Interaction Distance × Niveau × Visibilité : Les estimations de sonie *à la source* en fonction de la distance du locuteur étaient plus ou moins affectées par le fait que ce dernier soit visible ou caché selon le niveau auquel il prononçait les mots ($p = .017$). La figure 7.9 représente cette interaction.

Des tests post-hoc LSD de Fisher avec correction de Bonferroni ont été effectués sur cette interaction et indiquent que les estimations n'étaient strictement constantes avec la distance que dans le cas où le locuteur était visible et qu'il prononçait les mots à faible niveau ($-12\,\mathrm{dB}$). À

Variable dépendante		**Sonie *à la source***			
Source	**SS**	**DF**	**MS**	**F**	**Sig.p**
S	.279	1	.279	$3.542 \cdot 10^8$	**< .001**
D	2.344	1.689*	1.388	29.126	**< .001**
M	1.131	1.414*	.800	17.106	**< .001**
N	45.140	1.053*	42.861	104.642	**< .001**
V	.011	1.000	.011	1.051	.321
S × D	.043	1.811	.024	.909	.405
S × M	.007	2	.003	.503	.610
S × N	.167	1.309*	.128	3.669	.059
S × V	.000	1	.000	.030	.864
D × M	.056	4.595*	.012	2.034	.089
D × N	.041	2.873*	.014	.695	.554
D × V	.103	4	.026	6.535	**< .001**
M × N	3.003	1.591*	1.887	31.807	**< .001**
M × V	.047	2	.024	4.630	**.017**
N × V	.038	1.433*	.027	3.099	.079
S × D × M	.051	8	.006	1.245	.278
S × D × N	.027	3.578*	.008	.761	.542
S × D × V	.026	4	.007	2.110	.090
S × M × N	.020	2.153*	.009	.489	.631
S × M × V	.008	1.442*	.005	.576	.516
S × N × V	.006	2	.003	.663	.522
D × M × N	.062	5.623*	.011	.851	.528
D × M × V	.055	3.258*	.017	1.329	.274
D × N × V	.100	4.704*	.021	3.015	**.017**
M × N × V	.012	2.346*	.005	.431	.685
S × D × M × N	.062	5.886*	.010	.789	.579
S × D × M × V	.040	3.438	.012	.991	.412
S × D × N × V	.020	3.314*	.006	.411	.765
S × M × N × V	.060	2.726*	.022	4.206	**.013**
D × M × N × V	.071	4.898*	.014	.730	.601
S × D × M × N × V	.102	5.839*	.018	1.367	.237

* Le degré de liberté a été ajusté par une correction de Greenhouse-Geisser suite à une violation de l'hypothèse de sphéricité des résultats.

TABLE 7.3 – Résultats de l'ANOVA conduite sur les logarithmes des estimations de sonie *à la source*. S désigne la salle, D la distance de la source, N le niveau de diffusion de la source (assimilable à l'effort vocal du locuteur), M le mot prononcé par le locuteur et V la visibilité du locuteur.

ce niveau de prononciation des mots seulement, les auditeurs ont estimé différemment la sonie *à la source* lorsque le locuteur était visible et lorsqu'il était caché.

Chacune des courbes représentées dans cette figure a été soumise à une régression sur une fonction puissance de forme $L = k \cdot r^b$ (voir section 5.3). Le tableau 7.4 affiche les résultats de ces régressions, qui indiquent **une dépendance relativement faible des estimations de sonie *à la source* envers la distance du locuteur dans chacune des conditions**.

Autres résultats

Effets simples : La sonie *à la source* était plus forte quand les mots étaient prononcés plus fort (effet simple du Niveau, $p < .001$). Les estimations étaient aussi globalement plus élevées

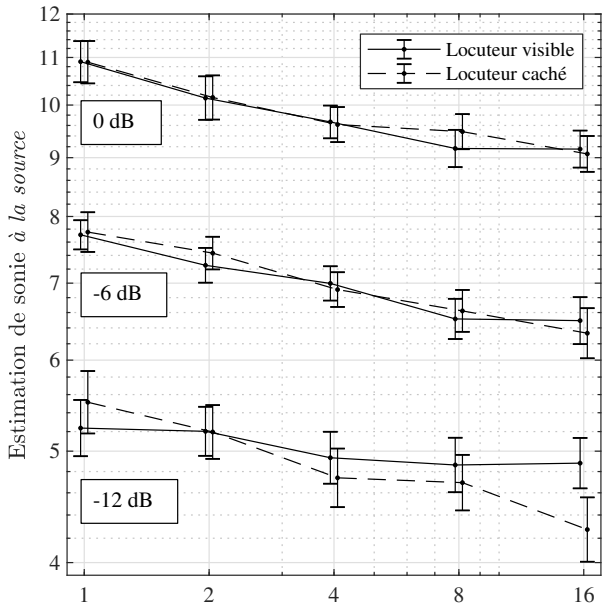


FIGURE 7.9 – Moyennes géométriques des estimations de sonie *à la source* en fonction de la distance du locuteur pour chaque niveau de prononciation des mots et pour chaque condition de visibilité du locuteur, avec intervalles de confiance à 95%. Les moyennes et les limites des intervalles de confiance ont été calculées comme expliqué figure 5.8.

dans la salle de sport que dans l'environnement anéchoïque (effet simple de la Salle, $p < .001$) et dépendaient également du mot prononcé (effet simple du Mot, $p < .001$).

Interaction Mot × Niveau : D'après cette interaction, les estimations de sonie *à la source* ne suivaient pas le niveau de prononciation des trois mots de la même manière ($p < .001$). Ces différences sont vraisemblablement dues à la prononciation propre à chaque mot et n'apportent pas d'éléments supplémentaires sur les autres résultats de cette étude.

Interaction Mot × Visibilité : Le fait de voir le locuteur n'a pas eu le même effet sur les estimations de sonie *à la source* selon le mot qu'il prononçait ($p = .017$). Les expressions faciales et corporelles lors de la prononciation des mots étaient propres à chaque mot et certaines ont vraisemblablement eu un effet plus prononcé que d'autres sur la sonie *à la source*.

Interaction Salle × Mot × Niveau × Visibilité : Cette interaction significative ($p = .013$) regroupe les facteurs impliqués dans les deux dernières interactions présentées ainsi que le facteur « Salle ». Des tests post-hoc LSD de Fisher avec correction de Bonferroni ont été menés sur cette interaction. Ces tests comparaient, dans chaque salle, pour chaque niveau de prononciation des mots et pour chaque mot les estimations de sonie *à la source* réalisées lorsque le locuteur était soit visible, soit caché. Aucune de ces comparaisons n'a donné de différence significative.

Niveau	Visibilité	Exposant b	MSE
−12 dB	Visible	−0.03	0.0041
	Caché	−0.09	0.0063
−6 dB	Visible	−0.07	0.0091
	Caché	−0.08	0.0025
0 dB	Visible	−0.07	0.0227
	Caché	−0.06	0.0192

TABLE 7.4 – Exposants b des fonctions puissance $L = k \cdot r^b$ obtenues par régression sur les résultats d'estimation de sonie montrant l'évolution de la sonie *à la source* avec la distance de la source sonore à chacun des trois niveaux de prononciation des mots lorsque le locuteur était soit visible, soit caché. Plus la valeur de l'exposant b est proche de zéro, moins les estimations étaient dépendantes de la distance. Les valeurs sont affichées avec la MSE entre les valeurs estimées lors de la régression et les valeurs réelles.

Bilan sur la sonie *à la source*

Les estimations de sonie *à la source* n'étaient pas dans leur ensemble strictement constantes avec la distance séparant le locuteur des auditeurs. La décroissance de ces estimations avec la distance était cependant relativement faible dans chacune des conditions expérimentales. Les exposants révélés dans chaque condition expérimentale par les régressions sur fonction puissance (voir section 5.3) sont relativement faibles (Zahorik et Wightman, 2001) et les estimations pourraient être considérées constantes d'après la littérature.

L'analyse de variance n'a pas relevé d'interaction significative faisant intervenir les facteurs « Distance » et « Salle ». En d'autres termes, les estimations de sonie *à la source* dépendaient de la même manière de la distance lorsque l'environnement comportait un champ réverbéré ou non. L'information de puissance sonore fournie par le timbre des mots prononcés a de ce fait joué un rôle comparable à celle fournie par l'énergie du champ réverbéré de la salle de sport.

Le fait de voir le locuteur a tout de même permis d'atteindre une réelle constance des estimations de sonie *à la source* avec la distance lorsque les mots étaient prononcés à faible niveau.

7.3.3 Conclusion

Les résultats de cette étude montrent que les estimations de sonie *aux oreilles* dépendaient plus de la distance que les estimations de sonie *à la source* pour des signaux de parole.

La sonie *aux oreilles* dépendait de la distance dans des proportions similaires à celles observées pour des bruits non-familiers. Les études de Pollack (1952) et Warren (1973) avaient pourtant observé des fonctions de sonie différentes pour ces deux types de stimuli. D'après ces études, la sonie de parole était moins dépendante du niveau sonore aux oreilles que celle de bruits ou de sons purs. Ces observations pourraient être expliquées par des différences dans les instructions données aux participants et dans les procédures utilisées.

Lorsque les instructions étaient d'estimer la sonie *à la source*, soit l'effort vocal perçu du locuteur, les estimations dépendaient peu de la distance et, par conséquent, du niveau aux oreilles des auditeurs. Cette observation a été faite aussi bien dans l'environnement réverbérant (lequel fournit des indices auditifs absolus de distance et de puissance de la source) que dans l'environnement anéchoïque, même lorsque le locuteur n'était pas visible. Dans ce dernier cas, seules les informations transmises par le timbre des signaux de paroles permettait aux auditeurs d'estimer la sonie *à la source*.

Conclusion de la deuxième partie

Les trois expériences de cette partie ont étudié les relations entre sonie et distance de sons non-familiers (des bruits blancs) et de sons familiers (des signaux de parole). Les instructions données aux participants distinguaient clairement la sonie *aux oreilles*, correspondant à la sensation de niveau sonore « entendu » évoquée par le signal, de la sonie *à la source*, correspondant à la puissance perçue de la source sonore. Ces deux grandeurs étaient évaluées lorsque la source sonore était placée à différentes distances des auditeurs et que différents indices de distance et de puissance (visuels et/ou auditifs) étaient disponibles. La distance perçue de la source sonore était également évaluée pour les sources diffusant du bruit blanc. Les résultats ont montré que :

- La sonie *aux oreilles* de bruits était dirigée par le niveau de pression aux oreilles des auditeurs et donc par la distance de la source sonore. La sonie *aux oreilles* ne dépendait pas de la visibilité de la source et n'était pas liée à la distance perçue (qui dépendait pour sa part de la visibilité de la source).
- La sonie *à la source* de bruits était indépendante de la distance de la source (et, par conséquent, du niveau de pression sonore aux oreilles) lorsque des indices de distance ou de puissance étaient disponibles. Ces indices pouvaient être auditifs (le rapport d'énergie champ direct à champ réverbéré comme indice de distance, l'énergie du champ réverbéré comme indice de puissance) ou visuels (fournis par le casque de réalité virtuelle). Ces différents indices ont joué un rôle similaire dans les estimations de sonie *à la source*.
- La sonie *aux oreilles* de signaux de parole dépendait du niveau de pression sonore aux oreilles (et donc de la distance de la source) dans des proportions similaires à celles observées sur du bruit blanc.
- La sonie *à la source* de signaux de parole était peu dépendante de la distance de la source, même lorsque les sons étaient restitués dans un environnement anéchoïque en l'absence d'indices visuels de distance. Dans ces conditions, les auditeurs basaient leurs estimations sur l'effort vocal du locuteur, perçu par le biais du timbre de sa voix.

Ces résultats exposent l'importance des instructions données aux auditeurs dans les expériences étudiant la sonie. Lorsque l'attention des auditeurs était activement dirigée vers la sonie *aux oreilles*, les estimations étaient clairement dépendantes du niveau de pression sonore aux oreilles quel que soit le type de stimulus. Lorsque l'attention était portée sur la sonie *à la source*, les estimations étaient nettement moins dépendantes du niveau de pression sonore aux oreilles dès lors que des indices de distance et/ou de puissance sonore (quels qu'ils soient) étaient disponibles.

Conclusion générale

Cette thèse a mis en évidence plusieurs relations entre la position d'une source sonore et la sonie des sons qu'elle émet. Ces relations étaient cependant radicalement différentes selon le type de sonie qu'évaluaient les auditeurs.

La sonie désigne le niveau sonore perçu d'un son. Ainsi, lorsque l'on souhaite déterminer la sonie d'un son, il est commun de demander à un auditeur : « Est-ce que ce son est fort ? ». Cependant, cette question peut se rapporter à l'évaluation de plusieurs grandeurs différentes. La force du son peut notamment être différente selon qu'elle est estimée au niveau de sa source ou au niveau de l'auditeur. Prenons l'exemple de la sirène d'une ambulance lointaine. L'auditeur pourrait considérer que le son de cette sirène **n'est pas fort** puisque l'ambulance est lointaine et qu'il entend le son à un faible niveau. Il pourrait également considérer que le son de cette sirène **est fort** puisque, par familiarité avec ce stimulus, il sait qu'une sirène d'ambulance restitue des sons de fort niveau.

Une partie des études de cette thèse a observé des relations entre position de la source et sonie *aux oreilles*. La sonie *aux oreilles* était évaluée dans les chapitres 2 et 3 puisque la nature des stimuli (des sons purs restitués au casque) ne laissait vraisemblablement pas d'autre choix aux auditeurs dans l'interprétation de la tâche. Dans le chapitre 5 et dans une partie du chapitre 7, la tâche était d'effectuer des estimations de sonie *aux oreilles* et l'attention des auditeurs était explicitement dirigée vers le niveau sonore « entendu ». La sonie *aux oreilles* observée dans ces études s'est révélée directement définie par les propriétés physiques des stimuli, qui dépendaient de l'azimut (ITD) et de la distance (niveau sonore aux oreilles) :

- Dans les études reportées dans les chapitres 2 et 3, les sons purs basses fréquences étaient perçus plus forts lorsqu'ils comportaient une ITD.
- Dans les études reportées dans les chapitres 5 et 7, le niveau perçu était entièrement défini par le niveau sonore aux oreilles des auditeurs, qui dépendait de la distance de la source. Cette relation était identique pour les sons non-familiers (des bruits) et les sons familiers (des signaux de parole). Les sons étaient donc perçus d'autant moins forts que la distance entre leur source et l'auditeur augmentait.

Une autre partie des études de cette thèse a observé des relations entre position de la source et sonie *à la source*. Dans le chapitre 6 et dans une partie du chapitre 7, la tâche était d'effectuer des estimations de sonie *à la source* et l'attention des auditeurs était dirigée explicitement vers la puissance de la source sonore. Les auditeurs prenaient alors en compte la position et la nature qu'ils percevaient de la source dans leurs estimations, qui n'étaient plus uniquement définies par les propriétés physiques des stimuli :

- Dans l'étude reportée dans le chapitre 6, la sonie *à la source* d'un bruit était indépendante du niveau sonore aux oreilles des auditeurs (et donc de la distance de la source) dès lors que des indices absolus de distance (le rapport d'énergie champ direct à champ réverbéré, les indices visuels) ou de puissance (l'énergie du champ réverbéré) étaient disponibles.
- L'étude reportée dans le chapitre 7 a réalisé le même constat pour des signaux de parole. Lorsque ces signaux étaient restitués en conditions anéchoïques sans indices visuels, les indices de puissance fournis par le timbre suffisaient à rendre la sonie *à la source* indépendante du niveau sonore aux oreilles.

Lorsque les instructions ne sont pas explicites, les auditeurs sont alors libres d'interpréter la question qui leur est posée (e.g. « Le son de la sirène est-il fort ? »). Ainsi, lors de l'étude reportée dans le chapitre 4, des bruits étaient restitués par paire depuis plusieurs positions et la question posée aux auditeurs était : « Lequel des deux sons est le plus fort ? ». Les résultats de cette étude montrent une sonie directionnelle moins marquée lorsque les sons étaient restitués par des haut-parleurs visibles que lorsqu'ils l'étaient via un casque sans indices visuels sur les sources. Cette observation pourrait être due à une interprétation différente de la question posée dans les deux conditions de présentation des sons. Lorsque les bruits étaient restitués au casque sans source visible, la sonie se rapportait vraisemblablement à la sonie *aux oreilles*, et l'ITD impliquait une augmentation de sonie relativement importante. Lorsque les bruits étaient restitués par des haut-parleurs visibles, le niveau sonore évalué par les auditeurs n'était possiblement pas directement le niveau sonore *aux oreilles*, mais considérait en partie le niveau sonore *à la source*, constant quelle que soit la position de la source sonore.

Ces résultats mettent en valeur le rôle joué par les instructions données aux auditeurs (et de l'interprétation qu'ils en font) dans l'évaluation de la sonie. Ce rôle est d'autant plus important dans les études s'intéressant aux relations entre position et sonie. La restitution d'un même son depuis différentes positions (direction et distance) résultera en des stimuli différents reçus par l'auditeur, dont les sonies *aux oreilles* ne seront potentiellement pas égales. Pour autant, la prise en compte des positions de leurs sources respectives dans l'évaluation de leurs sonies *à la source* pourrait engendrer une forme de constance.

Annexe A

Création des stimuli binauraux utilisés pour les expériences sur la distance

Afin de créer les stimuli sonores des deux salles réverbérantes, des sinus glissants de durée 10 s ont dans un premier temps été diffusés par un haut-parleur Genelec 8040A dans la grande salle de sport ainsi que dans la petite salle de concert. Ces sinus glissants ont été enregistrés dans ces deux salles à l'aide d'un microphone mh acoustics Eigenmike EM32 composé de 32 capsules, placé à une position unique dans les salles lorsque le haut-parleur était situé à une distance de 1 m, 2 m, 4 m, 8 m et 16 m de cette position. Les sinus glissants étaient envoyés au haut-parleur depuis une carte son RME Fireface 800 et les enregistrements du microphone étaient récupérés par l'intermédiaire de l'interface audio propriétaire d'Eigenmike. Le tout était piloté par une session créée sur le logiciel de DAW (*Digital Audio Workstation*) Reaper. La photo représentée figure A.1 montre ce dispositif lors de l'enregistrement effectué dans la grande salle de sport.

FIGURE A.1 – Photos de l'enregistrement des réponses impulsionnelles de la salle de sport. La photo de gauche montre de près le haut-parleur, placé à 1 m du microphone ambisonique. La photo de droite montre le dispositif d'enregistrement, comprenant de gauche à droite la session d'enregistrement sur Reaper, l'interface audio Eigenmike et la carte son RME Fireface 800.

Ces enregistrements ont permis de créer des réponses impulsionnelles de salle (RIR, pour *Room Impulse Response*), qui ont dans un premier temps été débruitées suivant la procédure

décrite par Cabrera *et al.* (2011). Ensuite, l'effet du haut-parleur dans la restitution des sinus glissants a été compensé d'après des filtres d'égalisation calculés sur la base de mesures effectuées en chambre anéchoïque (Salmon *et al.*, 2019). Enfin, les RIR ont été encodées en ambisonie d'ordre élevé (HOA, pour *Higher-Order Ambisonics*) d'ordre quatre. Cette technologie permet de capter et de restituer les sons avec une grande précision spatiale en trois dimensions. D'abord, la sortie des trente-deux capsules du microphone est encodée. Dans le format encodé, un certain nombre de canaux audio (vingt-cinq en HOA d'ordre quatre) décrivent le champ acoustique capté sous forme d'harmoniques sphériques. La figure A.2 représente les diagrammes de directivité de ces harmoniques sphériques. Le premier canal, Y_0^0, correspond à l'ordre zéro et est omnidirectionnel. Chaque ligne inférieure décrit un ordre supérieur, indiqué par l'indice m dans la notation Y_m^n. Cette représentation sous forme d'harmoniques sphériques permet d'effectuer un certains nombres de transformations sur le champ sonore capté. Il est par exemple possible d'effectuer des rotations du champ sonore autour des trois axes x, y et z de la figure A.2. Enfin, ces vingt-cinq canaux sont décodés pour être restitués sur le système de diffusion quel qu'il soit (sur un réseau de haut-parleurs scalable, en binaural, etc.). Une description détaillée de cette technologie est proposée par Lecomte (2016). Une documentation détaillée sur les harmoniques sphériques du format ambisonique utilisé est disponible sur le site internet du fabriquant mh acoustics [1].

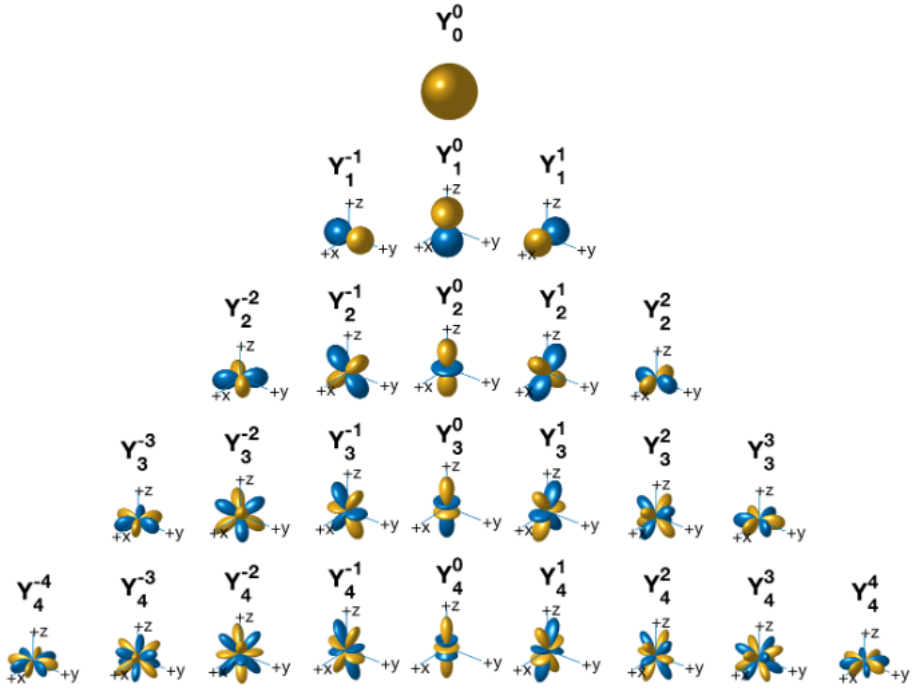


FIGURE A.2 – Harmoniques sphériques représentées sur les vingt-cinq canaux en HOA d'ordre quatre. Les parties de l'espace représentées en jaune correspondent aux valeurs positives, celles représentées en bleu aux valeurs négatives.

Un bruit blanc de durée 200 ms a été convolué avec chacune des dix RIR ainsi créées (cinq distances différentes dans chacune des deux salles). Ce bruit blanc a également été directement

1. Document *Eigenbeam Data* à télécharger à l'adresse : https://mhacoustics.com/download.

encodé en HOA d'ordre quatre pour créer les stimuli anéchoïques. Cinq encodages différents ont été réalisés de cette manière, correspondant aux cinq distances différentes de la source sonore virtuelle. Le niveau était modifié pour réaliser ces cinq distances artificielles, suivant une décroissance de 6 dB par doublement de distance.

Afin de fournir aux auditeurs des conditions d'écoute réalistes bien qu'en restituant les sons au casque, les informations de lacet, de tangage et de roulis (plus communément désignés par leurs traductions anglaises *yaw*, *pitch* et *roll*, respectivement) fournies en temps réel par le HMD étaient utilisées pour effectuer des rotations du champ sonore ambisonique en accord avec les mouvements de la tête des auditeur (*head tracking*). Ces rotations étaient effectuées en temps réel à l'aide du plugin ambiX Rotator (Kronlachner, 2014).

Enfin, les signaux HOA étaient décodés en sons binauraux par l'intermédiaire du plugin IEM BinauralDecoder [2], fonctionnant d'après la méthode *MagLS* décrite par Schorkhuber *et al.* (2018) et utilisant les HRTF de la tête artificielle Neumann KU100 mesurés par Bernschütz (2013). Ces HRTF ont été mesurés à une distance de 3.25 m du centre de la tête artificielle, distance suffisamment éloignée pour que l'ILD ne dépende pas de la distance de la source sonore (Duda et Martens, 1998). Les HRTF ne dépendent de la distance de la source qu'à des distances inférieures à 1 m (Otani *et al.*, 2009) et donc à des distances inférieures à celles testées dans cette étude. Les mêmes HRTF étaient donc utilisées pour les cinq distances différentes de ce test.

Les sons étaient restitués au casque après avoir été filtrés par l'intermédiaire des filtres de compensation de casque fournis par le plugin IEM BinauralDecoder. Les rotations du champ ambisonique ainsi que le décodage en binaural étaient effectués en temps réel sur le logiciel Max/MSP. Les sons étaient délivrés au casque par une carte son RME Babyface Pro. Les niveaux de restitution étaient étalonnés en posant le casque sur une tête artificielle Neumann KU 100 précédemment étalonnée à l'aide d'un pistonphone (Brüel & Kjær Type 4231) diffusant un son pur de fréquence 1 kHz à 94 dB SPL. En définitive, les stimuli sonores virtuels semblaient être émis par une source située à une position fixe (à l'une des cinq distances différentes du point d'écoute), indépendamment des mouvements de tête des auditeurs.

2. La description du plugin ainsi que son téléchargement sont disponibles à l'adresse : https://plugins.iem.at

Annexe B

Mesures effectuées dans les salles

Les niveaux des champs direct et réverbéré ont été mesurés à chacune des cinq distances sur les RIR des deux salles réverbérantes par la méthode expliquée par Zahorik (2002). D'après cette méthode, les premières 2.5 ms de la réponse impulsionnelle constituent le champ direct et le reste du signal constitue le champ réverbéré. La durée de 2.5 ms correspond d'après Møller *et al.* (1995) à la durée des réponses impulsionnelles des HRTF. Les décroissances des champs direct et réverbéré ainsi que le DRR ont été déterminés à chacune des cinq distances étudiées dans chaque salle d'après ces mesures. Ces résultats sont indiqués dans le tableau B.1.

Salle	**Mesure**	1 m	2 m	4 m	8 m	16 m
Salle de sport	Champ direct	0 dB	−5.8 dB	−11.7 dB	−15.9 dB	−23.1 dB
	Champ réverbéré	0 dB	−1.7 dB	−1.2 dB	−7.4 dB	−8.6 dB
	DRR	15.6 dB	11.5 dB	5.1 dB	7.1 dB	1.2 dB
Salle de concert	Champ direct	0 dB	−5.7 dB	−11.3 dB	−16.7 dB	−21.7 dB
	Champ réverbéré	0 dB	−0.7 dB	−2.2 dB	−5.5 dB	−8.7 dB
	DRR	8.8 dB	3.8 dB	−0.3 dB	−2.4 dB	−4.2 dB

Table B.1 – Mesures du DRR ainsi que des décroissances des niveaux du champ direct et du champ réverbéré dans chaque salle à chacune des cinq distances. Les niveaux des champs direct et réverbéré sont exprimés en niveau relatif au niveau à 1 m.

Des régressions non linéaire sur des fonctions logarithmiques de forme $b + a \cdot \log_{10}(d)$ (Zahorik et Wightman, 2001) ont été calculées pour chacun de ces résultats afin d'estimer la décroissance de chacune de ces grandeurs chaque doublement de distance. Ces résultats sont affichés dans le tableau B.2.

Mesure	**Salle de sport**	**Salle de concert**
Champ direct	−5.6 dB	−5.4 dB
Champ réverbéré	−2.3 dB	−2.2 dB
DRR	−3.3 dB	−3.2 dB

Table B.2 – Estimation des décroissances du champ direct, du champ réverbéré et du DRR pour un doublement de distance entre le point d'écoute et la source sonore dans chacune des deux salles.

Bibliographie

R. Aibara, J. T. Welsh, S. Puria et R. L. Goode : Human middle-ear sound transfer function and cochlear input impedance. *Hearing Research*, 152(1-2):100–109, 2001, doi : 10.1016/S0378-5955(00)00240-9.

D. Alais et D. Burr : The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14(3):257–262, 2004, doi : 10.1016/j.cub.2004.01.029.

D. Algom, R. Adam et L. Cohen-Raz : Binaural summation and lateralization of transients : A combined analysis. *The Journal of the Acoustical Society of America*, 84(4):1302–1315, oct. 1988, doi : 10.1121/1.396629.

D. Algom, A. Rubin et L. Cohen-Raz : Binaural and temporal integration of the loudness of tones and noises. *Perception & Psychophysics*, 46(2):155–166, mars 1989, doi : 10.3758/BF03204975.

C. F. Altmann, M. Matsuhashi, M. Votinov, K. Goto, T. Mima et H. Fukuyama : Visual distance cues modulate neuromagnetic auditory N1m responses. *Clinical Neurophysiology*, 123 (11):2273–2280, 2012, doi : 10.1016/j.clinph.2012.04.004.

P. W. Anderson et P. Zahorik : Auditory/visual distance estimation : accuracy and variability. *Frontiers in Psychology*, 5, oct. 2014, doi : 10.3389/fpsyg.2014.01097.

C. R. Andre : *Audiovisual spatial congruence, and applications to 3D sound and stereoscopic video.* Thèse de doctorat, Université de Liège, Belgique, 2013.

J. Andre et S. Rogers : Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Perception & Psychophysics*, 68(3):353–361, 2006, doi : 10.3758/BF03193682.

ANSI S3.21 : Methods for manual pure-tone threshold audiometry, 2004.

ANSI S3.4 : American national standard procedure for the computation of loudness of steady sounds, avr. 2017.

D. H. Ashmead, D. Leroy et R. D. Odom : Perception of the relative distances of nearby sound sources. *Perception & Psychophysics*, 47(4):326–331, 1990, doi : 10.3758/BF03210871.

D. W. Batteau : The role of the pinna in human localization. *Proceedings of the Royal Society B*, 168(1011):158–180, 1967, doi : 10.1098/rspb.1967.0058.

V. Benichoux, M. Rebillat et R. Brette : On the variation of interaural time differences with frequency. *The Journal of the Acoustical Society of America*, 139(4):1810–1821, 2016.

B. Bernschütz : A spherical far field HRIR/HRTF compilation of the Neumann KU 100. *In Proceedings of the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA) Conference on Acoustics*, p. 29, 2013.

L. R. Bernstein et C. Trahiotis : Lateralization of sinusoidally amplitude-modulated tones : Effects of spectral locus and temporal variation. *The Journal of the Acoustical Society of America*, 78(2):514–523, 1985, doi : 10.1121/1.392473.

L. R. Bernstein et C. Trahiotis : Binaural signal detection, overall masking level, and masker interaural correlation : Revisiting the internal noise hypothesis. *The Journal of the Acoustical Society of America*, 124(6):3850–3860, 2008, doi : 10.1121/1.2996340.

L. R. Bernstein, C. Trahiotis et E. L. Hyde : Inter-individual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrow-band or broadband noise. *The Journal of the Acoustical Society of America*, 103(4):2069–2078, 1998, doi : 10.1121/1.421378.

P. Bertelson et M. Radeau : Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, 29(6):578–584, 1981, doi : 10.3758/BF03207374.

G. Berthomieu, V. Koehl et M. Paquier : Influence of interaural time differences on loudness for low-frequency pure tones at varying signal and noise levels. *In Proceedings of Meetings on Acoustics*, vol. 30, p. 050004, 2017, doi : 10.1121/2.0000553.

G. Berthomieu, V. Koehl et M. Paquier : Directional loudness of low-frequency noises actually presented over loudspeakers and virtually presented over headphones. *Journal of the Audio Engineering Society*, 67(9):11, 2019a, doi : 10.17743/jaes.2019.0018.

G. Berthomieu, V. Koehl et M. Paquier : Loudness and distance estimates for noise bursts coming from several distances with and without visual cues to their source. *In Proceedings of the 23rd International Congress on Acoustics*, p. 3897–3904, Aachen, Germany, 2019b, doi : 10.18154/RWTH-CONV-239023.

J. Blauert : Sound localization in the median plane. *Acta Acustica united with Acustica*, 22:205–213, 1969.

J. Blauert : Binaural localization : Multiple images and applications in room-and electroacoustics. *In Localization of Sound : Theory and applications*, p. 65–84. R. W. Gatehouse, Groton, CT : Amphora, 1982.

J. Blauert : Distance hearing and inside-the-head locatedness. *In Spatial Hearing : The Psychophysics of Human Sound Localization*, p. 116. MIT Press, Cambridge, MA, 1997a. ISBN 978-0-262-02413-6.

J. Blauert : *Spatial Hearing : The Psychophysics of Human Sound Localization.* MIT Press, Cambridge, MA, 1997b. ISBN 978-0-262-02413-6.

J. Blauert : Ein versuch zum richtungshören bei gleichzeitiger optischer stimulation [an experiment in directional hearing with simultaneous optical stimulation]. *Acustica*, 23:118–119, 1970.

R. C. Bolles et D. E. Bailey : Importance of object recognition in size constancy. *Journal of Experimental Psychology*, 51(3):222–225, 1956, doi : 10.1037/h0048080.

D. S. Brungart : Auditory localization of nearby sources. III. Stimulus effects. *The Journal of the Acoustical Society of America*, 106(6):3589–3602, 1999, doi : 10.1121/1.428212.

D. S. Brungart, N. I. Durlach et W. M. Rabinowitz : Auditory localization of nearby sources. II. Localization of a broadband source. *The Journal of the Acoustical Society of America*, 106(4):1956–1968, 1999a, doi : 10.1121/1.427943.

D. S. Brungart, N. I. Durlach et W. M. Rabinowitz : Auditory localization of nearby sources. II. Localization of a broadband source. *The Journal of the Acoustical Society of America*, 106(4):1956–1968, 1999b, doi : 10.1121/1.427943.

D. S. Brungart et W. M. Rabinowitz : Auditory localization of nearby sources. Head-related transfer functions. *The Journal of the Acoustical Society of America*, 106(3):1465–1479, 1999, doi : 10.1121/1.427180.

D. S. Brungart et K. R. Scott : The effects of production and presentation level on the auditory distance perception of speech. *The Journal of the Acoustical Society of America*, 110 (1):425–440, juil. 2001, doi : 10.1121/1.1379730.

E. Brunswik : Loudness constancy with distance variant. *In Perception and the representative design of psychological experiments*, p. 70–72. Univ of California Press, 2nd édn, 1956.

R. Butler : Monaural and binaural localization of noise bursts vertically in median sagittal plane. *Journal of Auditory research*, 9(3):230–235, 1969.

R. A. Butler, E. T. Levy et W. D. Neff : Apparent distance of sounds recorded in echoic and anechoic chambers. *Journal of Experimental Psychology : Human Perception and Performance*, 6(4):745, 1980.

S. Buus et M. Florentine : Growth of loudness in listeners with cochlear hearing losses : Recruitment reconsidered. *Journal of the Association for Research in Otolaryngology*, 3(2):120–139, juin 2002, doi : 10.1007/s101620010084.

S. Buus, H. Müsch et M. Florentine : On loudness at threshold. *The Journal of the Acoustical Society of America*, 104(1):399–410, juil. 1998, doi : 10.1121/1.423295.

S. Buus, M. Florentine et T. Poulsen : Temporal integration of loudness, loudness discrimination, and the form of the loudness function. *The Journal of the Acoustical Society of America*, 101(2):669–680, fév. 1997, doi : 10.1121/1.417959.

D. Cabrera, D. Lee, M. Yadav et W. L. Martens : Decay envelope manipulation of room impulse responses : Techniques for auralization and sonification. *In Proceedings of Acoustics 2011*, p. 5, Gold Coast, Australie, 2011.

E. R. Calcagno, E. L. Abregú, M. C. Eguía et R. Vergara : The role of vision in auditory distance perception. *Perception*, 41(2):175–192, 2012, doi : 10.1068/p7153.

S. Carlile : *Virtual Auditory Space : Generation and Applications*. Springer-Verlag, Berlin, Heidelberg, 1996. ISBN 978-3-662-22596-7.

S. Carlile, P. Leong et S. Hyams : The nature and distribution of errors in sound localization by human listeners. *Hearing Research*, 114(1-2):179–196, 1997, doi : 10.1016/S0378-5955(97)00161-5.

B. G. Churcher : A loudness scale for industrial noise measurements. *The Journal of the Acoustical Society of America*, 6(4):216–225, 1935, doi : 10.1121/1.1915740.

B. E. Coutant et G. Westheimer : Population distribution of stereoscopic ability. *Ophthalmic and Physiological Optics*, 13(1):3–7, 1993, doi : 10.1111/j.1475-1313.1993.tb00419.x.

J. F. Culling : Evidence specifically favoring the equalization-cancellation theory of binaural unmasking. *The Journal of the Acoustical Society of America*, 122(5):2803, 2007, doi : 10.1121/1.2785035.

J. E. Cutting et P. M. Vishton : Perceiving layout and knowing distances : The integration, relative potency, and contextual use of different information about depth. *In Perception of space and motion*, p. 69–117. Elsevier, 1995.

P. Damaske et B. Wagener : Richtungshörversuche über einen nachgebildeten kopf [investigations on directional hearing using a dummy head]. *Acustica*, 21:30–35, 1969.

A. Dancer : Le traumatisme acoustique. *Médecine Sciences*, 7(4):357, 1991, doi : 10.4267/10608/4361.

A. Delorme et M. Flückiger : *Perception et réalité : une introduction à la psychologie des perceptions*. Neurosciences & cognition. De Boeck, 2003. ISBN 978-2-89105-807-0.

K. J. Diercks et L. A. Jeffress : Interaural phase and the absolute threshold for tone. *The Journal of the Acoustical Society of America*, 34(7):981–984, 1962, doi : 10.1121/1.1918234.

M. Dietz, M. Klein-Hennig et V. Hohmann : The influence of pause, attack, and decay duration of the ongoing envelope on sound lateralization. *The Journal of the Acoustical Society of America*, 137(2):EL137–EL143, fév. 2015, doi : 10.1121/1.4905891.

N. A. Dodgson : Variation and extrema of human interpupillary distance. *In Proceedings of SPIE*, vol. 5291, 2004, doi : https://doi.org/10.1117/12.529999.

R. O. Duda et W. L. Martens : Range dependence of the response of a spherical head model. *The Journal of the Acoustical Society of America*, 104(5):3048–3058, 1998, doi : 10.1121/1.423886.

N. I. Durlach : Binaural phenomena. *In* E. Carterette et M. P. Friedman, éds : *Handbook of Perception*, vol. IV. Academic Press, New York, NY, 1978.

N. I. Durlach : Equalization and cancellation theory of binaural masking-level differences. *The Journal of the Acoustical Society of America*, 35(8):1206–1218, 1963.

B. A. Edmonds et J. F. Culling : The spatial unmasking of speech : Evidence for better-ear listening. *The Journal of the Acoustical Society of America*, 120(3):1539–1545, sept. 2006, doi : 10.1121/1.2228573.

B. A. Edmonds et J. F. Culling : Interaural correlation and the binaural summation of loudness. *The Journal of the Acoustical Society of America*, 125(6):3865–3870, 2009, doi : 10.1121/1.3120412.

G. Ekman, B. Berglund et U. Berglund : Loudness as a function of the duration of auditory stimulation. *Scandinavian Journal of Psychology*, 7(1):201–208, mars 1966, doi : 10.1111/j.1467-9450.1966.tb01354.x.

M. Epstein et M. Florentine : A test of the equal-loudness-ratio hypothesis using cross-modality matching functions. *The Journal of the Acoustical Society of America*, 118(2):907–913, 2005, doi : 10.1121/1.1954547.

M. Epstein et M. Florentine : Loudness of brief tones measured by magnitude estimation and loudness matching. *The Journal of the Acoustical Society of America*, 119(4):1943–1945, avr. 2006, doi : 10.1121/1.2177592.

M. Epstein et M. Florentine : Binaural loudness summation for speech and tones presented via earphones and loudspeakers. *Ear and hearing*, 30(2):234–237, 2009, doi : 10.1097/AUD.0b013e3181976993.

M. Epstein et M. 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):3981–3988, 2012, doi : 10.1121/1.3701984.

M. O. Ernst et M. S. Banks : Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002, doi : 10.1038/415429a.

H. Fastl et E. Zwicker : Critical bands and excitation. *In Psychoacoustics : Facts and Models*, p. 149–173. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-68888-4.

G. T. Fechner, D. H. Howes et E. G. Boring : *Elements of psychophysics*, vol. 1. Holt, Rinehart and Winston New York, 1966.

W. E. Feddersen, T. T. Sandel, D. C. Teas et L. A. Jeffress : Localization of high-frequency tones. *The Journal of the Acoustical Society of America*, 29(9):988–991, 1957, doi : 10.1121/1.1909356.

P. G. Finlayson et D. M. Caspary : Low-frequency neurons in the lateral superior olive exhibit phase-sensitive binaural inhibition. *Journal of Neurophysiology*, 65(3):598–605, mars 1991, doi : 10.1152/jn.1991.65.3.598.

H. Fletcher : Minimum perceptible changes in frequency and sound pressure level. *In Speech and hearing in communication, 2nd ed*, vol. X, p. 144–152. D. Van Nostrand, Oxford, England, 1953.

H. Fletcher et W. A. Munson : Loudness, its definition, measurement and calculation. *The Journal of the Acoustical Society of America*, 5(2):82–108, 1933, doi : 10.1121/1.1915637.

M. Florentine : Loudness. *In* M. Florentine, A. N. Popper et R. R. Fay, éds : *Loudness*, p. 1–15. Springer New York, New York, NY, 2011. ISBN 978-1-4419-6712-1.

M. Florentine, S. Buus et P. Bonding : Loudness of complex sounds as a function of the standard stimulus and the number of components. *The Journal of the Acoustical Society of America*, 64(4):1036–1040, oct. 1978, doi : 10.1121/1.382062.

M. Florentine, S. Buus et T. Poulsen : Temporal integration of loudness as a function of level. *The Journal of the Acoustical Society of America*, 99(3):1633–1644, 1996, doi : 10.1121/1.415236.

M. Florentine et M. Epstein : To honor Stevens and repeal his law (for the auditory system). *Proceedings of Fechner Day*, 22(1):37–42, 2006.

J.-E. Fournier : *Audiométrie vocale : les épreuves d'intelligibilité et leurs applications au diagnostic, à l'expertise et à la correction prothétique des surdités*. Maloine, 1951.

S. S. Fukusima, J. M. Loomis et J. A. Da Silva : Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology : Human Perception and Performance*, 23(1):86, 1997.

M. B. Gardner : Proximity image effect in sound localization. *The Journal of the Acoustical Society of America*, 43(1):163–163, 1968, doi : 10.1121/1.1910747.

G. Gigerenzer et G. Strube : Are there limits to binaural additivity of loudness ? *Journal of Experimental Psychology : Human Perception and Performance*, 9(1):126–136, 1983, doi : 10.1037/0096-1523.9.1.126.

B. R. Glasberg et B. C. J. Moore : Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1–2):103–138, 1990, doi : 10.1016/0378-5955(90)90170-T.

B. R. Glasberg et B. C. J. Moore : Prediction of absolute thresholds and equal-loudness contours using a modified loudness model. *The Journal of the Acoustical Society of America*, 120(2):585–588, 2006, doi : 10.1121/1.2214151.

B. R. Glasberg et B. C. J. Moore : The loudness of sounds whose spectra differ at the two ears. *The Journal of the Acoustical Society of America*, 127(4):2433–2440, avr. 2010, doi : 10.1121/1.3336775.

J. A. Grange et J. F. Culling : The benefit of head orientation to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 139(2):703–712, fév. 2016, doi : 10.1121/1.4941655.

T. Y. Grechkin, T. D. Nguyen, J. M. Plumert, J. F. Cremer et J. K. Kearney : How does presentation method and measurement protocol affect distance estimation in real and virtual environments ? *ACM Transactions on Applied Perception*, 7(4):1–18, 2010, doi : 10.1145/1823738.1823744.

B. Grothe, M. Pecka et D. McAlpine : Mechanisms of sound localization in mammals. *Physiological Reviews*, 90(3):983–1012, juil. 2010, doi : 10.1152/physrev.00026.2009.

P. Guillon : *Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF.* Thèse de doctorat, Université du Maine, France, 2009.

G. G. Harris : Binaural interactions of impulsive stimuli and pure tones. *The Journal of the Acoustical Society of America*, 32(6):685–692, 1960, doi : 10.1121/1.1908181.

W. M. Hartmann et A. Wittenberg : On the externalization of sound images. *The Journal of the Acoustical Society of America*, 99(6):3678–3688, 1996, doi : 10.1121/1.414965.

B. Haustein et W. Schirmer : Messeinrichtung zur untersuchung des richtungslokalisationsvermögens [a measuring apparatus for the investigation of the faculty of directional localization]. *Hochfrequenztechnik und Elektroakustik*, 79:96–101, 1970.

R. M. Held et N. I. Durlach : Telepresence. *Presence : Teleoperators & Virtual Environments*, 1(1):109–112, 1992, doi : https://doi.org/10.1162/pres.1992.1.1.109.

R. Hellman : Loudness measurements by magnitude scaling : Implication for intensity coding. *In* S. J. B. Jr. et G. A. Gescheider, éds : *Ratio scaling of psychological magnitude : In honor of the memory of S. S. Stevens.* Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1991. ISBN 0-8058-0710-1.

R. Hellman, A. Miśkiewicz et B. Scharf : Loudness adaptation and excitation patterns : Effects of frequency and level. *The Journal of the Acoustical Society of America*, 101(4):2176–2185, 1997.

R. P. Hellman : Growth of loudness in sensorineural impairment : Experimental results and modeling implications. *In Modeling Sensorineural Hearing Loss*, p. 199–212. Taylor & Francis, 1997. ISBN 978-1-317-72937-2.

R. P. Hellman et J. Zwislocki : Monaural loudness function at 1000 cps and interaural summation. *The Journal of the Acoustical Society of America*, 35(6):856–865, 1963, doi : 10.1121/1.1918619.

E. Hendrickx : *Influence de la stéréoscopie sur la perception du son : cas de mixages sonores pour le cinéma en relief.* Thèse de doctorat, Université de Bretagne Occidentale, France, 2015.

E. Hendrickx, M. Paquier, V. Koehl et J. Palacino : Ventriloquism effect with sound stimuli varying in both azimuth and elevation. *The Journal of the Acoustical Society of America*, 138(6):3686–3697, 2015, doi : 10.1121/1.4937758.

E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B. Katz et C. de Boishéraud : Improvement of externalization by listener and source movement using a "binauralized" microphone array. *Journal of the Audio Engineering Society*, 65(7/8):589–599, août 2017a, doi : 10.17743/jaes.2017.0018.

E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B. F. Katz et C. 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):2011–2023, 2017b, doi : 10.1121/1.4978612.

I. J. Hirsh : The influence of interaural phase on interaural summation and inhibition. *The Journal of the Acoustical Society of America*, 20(4):536–544, 1948, doi : 10.1121/1.1906407.

I. P. Howard : *Human visual orientation.* Chichester [Eng.] ; New York : J. Wiley, 1982. ISBN 0471279463. Includes indexes.

L. E. Humes et W. Jesteadt : Models of the effects of threshold on loudness growth and summation. *The Journal of the Acoustical Society of America*, 90(4):1933–1943, oct. 1991, doi : 10.1121/1.401673.

K. Hüttenbrink et H. Hudde : Untersuchungen zur schalleitung durch das rekonstruierte mittelohr mit einem hydrophon : erste ergebnisse. *HNO. Hals-, Nasen-, Ohrenärzte*, 42(1):49–57, 1994.

IEC 60268-13 : Sound system equipment – Part 13 : Listening tests on loudspeakers, 1998.

U. Ingård : A review of the influence of meteorological conditions on sound propagation. *The Journal of the Acoustical Society of America*, 25(3):405–411, 1953, doi : 10.1121/1.1907055.

ISO 226 : Acoustics – Normal equal-loudness-level contours, 2003.

ISO 389-7 : Acoustics – Reference zero for the calibration of audiometric equipment – Part 7 : Reference threshold of hearing under free-field and diffuse-field listening conditions, 2005.

ISO 532-B : Méthode de calcul du niveau d'isosonie, 1975.

W. Jesteadt et L. J. Leibold : Loudness in the laboratory, part I : Steady-state sounds. *In* M. Florentine, A. N. Popper et R. R. Fay, éds : *Loudness*, p. 109–144. Springer New York, New York, NY, 2011. ISBN 978-1-4419-6712-1.

J. Jetzt : Critical distance measurement of rooms from the sound energy spectral response. *The Journal of the Acoustical Society of America*, 65(5):1204–1211, 1979, doi : https://doi.org/10.1121/1.382786.

J. Jouhaneau : Perception de l'espace et immersion - perception visuelle. *Techniques de l'ingénieur*, 2012.

B. Julesz : *Foundations of cyclopean perception.* Foundations of cyclopean perception. U. Chicago Press, Oxford, England, 1971.

S. Jørgensen : Matlab model for loudness perception-directional loudness. Unpublished masters thesis, Technical University of Denmark (DTU), Copenhagen, 2002.

G. Kearney, M. Gorzel, H. Rice et F. Boland : Distance perception in interactive virtual acoustic environments using first and higher order ambisonic sound fields. *Acta Acustica united with Acustica*, 98(1):61–71, jan. 2012, doi : 10.3813/AAA.918492.

K. Keen : Preservation of constant loudness with interaural amplitude asymmetry. *The Journal of the Acoustical Society of America*, 52(4B):1193–1196, 1972, doi : 10.1121/1.1913231.

F. P. Kilpatrick et W. H. Ittelson : The size-distance invariance hypothesis. *Psychological Review*, 60(4):223–231, 1953, doi : 10.1037/h0060882.

H.-Y. Kim, Y. Suzuki, S. Takane et T. Sone : Control of auditory distance perception based on the auditory parallax model. *Applied Acoustics*, 62(3):245–270, 2001, doi : 10.1016/S0003-682X(00)00023-2.

B. A. Kingsbury : A direct comparison of the loudness of pure tones. *Physical Revue*, 29 (4):588–600, avr. 1927, doi : 10.1103/PhysRev.29.588.

R. E. Kirk : *Experimental design : Procedures for the behavioral sciences, 3rd ed.* Thomson Brooks/Cole Publishing Co, Belmont, CA, US, 1995. ISBN 0-534-25092-0 (Hardcover).

O. Klemm : Untersuchungen über die lokalisation von schallreizen iv : über den einfluss des binauralen zeitunterschieds auf die lokalisation. *Archiv für die gesamte Psychologie*, 40:117–145, 1920.

V. Koehl et M. Paquier : Loudness of low-frequency pure tones lateralized by interaural time differences. *The Journal of the Acoustical Society of America*, 137(2):1040–1043, 2015, doi : 10.1121/1.4906262.

V. Koehl, M. Paquier et E. Hendrickx : Effects of interaural differences on the loudness of low-frequency pure tones. *Acta Acustica united with Acustica*, 101(6):1168–1173, 2015, doi : 10.3813/AAA.918909.

V. Koehl, M. Paquier et E. 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):705–708, 2017, doi : 10.3813/AAA.919098.

A. Kohlrausch et S. van de Par : Audio—visual interaction in the context of multi-media applications. *In* J. Blauert, éd. : *Communication Acoustics*, p. 109–138. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. ISBN 978-3-540-22162-3 978-3-540-27437-7.

A. Kolarik, S. Cirstea et S. Pardhan : Discrimination of virtual auditory distance using level and direct-to-reverberant ratio cues. *The Journal of the Acoustical Society of America*, 134(5):3395–3398, 2013, doi : 10.1121/1.4824395.

A. J. Kolarik, B. C. J. Moore, P. Zahorik, S. Cirstea et S. Pardhan : Auditory distance perception in humans : A review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics*, 78(2):373–395, 2016, doi : 10.3758/s13414-015-1015-1.

S. Komiyama : Subjective evaluation of angular displacement between picture and sound directions for HDTV sound systems. *Journal of the Audio Engineering Society*, 37(4):210–214, 1989.

N. Kopčo et B. G. Shinn-Cunningham : Effect of stimulus spectrum on distance perception for nearby sources. *The Journal of the Acoustical Society of America*, 130(3):1530–1541, sept. 2011, doi : 10.1121/1.3613705.

M. Kronlachner : Plug-in suite for mastering the production and playback in surround sound and ambisonics. *Gold-Awarded Contribution to AES Student Design Competition*, 2014.

G. F. Kuhn : Model for the interaural time differences in the azimuthal plane. *The Journal of the Acoustical Society of America*, 62(1):157–167, 1977, doi : 10.1121/1.381498.

A. Kulkarni, S. K. Isabelle et H. S. Colburn : Sensitivity of human subjects to head-related transfer-function phase spectra. *The Journal of the Acoustical Society of America*, 105 (5):2821–2840, 1999, doi : 10.1121/1.426898.

S. Kuwano, H. Fast et S. Namba : Relation between loudness of sounds with under critical duration and L_{Aeq} and L_{AE} (2) in the case of impulsive noise. *In Proc. of 2007 Spring Meeting of the Acoust. Soc. of Japan, Tokyo*, 2007.

S. Kuwano et S. Namba : Loudness in the laboratory, part II : Non-steady-state sounds. *In* M. Florentine, A. N. Popper et R. R. Fay, éds : *Loudness*, p. 145–168. Springer New York, New York, NY, 2011. ISBN 978-1-4419-6712-1.

E. H. A. Langendijk et A. W. Bronkhorst : Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *The Journal of the Acoustical Society of America*, 107(1):528–537, 2000, doi : 10.1121/1.428321.

E. Larsen, N. Iyer, C. R. Lansing et A. S. Feng : On the minimum audible difference in direct-to-reverberant energy ratio. *The Journal of the Acoustical Society of America*, 124 (1):450–461, 2008, doi : 10.1121/1.2936368.

P. Lecomte : *Ambisonie d'ordre élevé en trois dimensions : captation, transformations et décodage adaptatifs de champs sonores.* Thèse de doctorat, Conservatoire National des Arts et Metiers, 2016.

L. J. Leibold, H. Tan, S. Khaddam et W. Jesteadt : Contributions of individual components to the overall loudness of a multitone complex. *The Journal of the Acoustical Society of America*, 121(5):11, 2007.

T. Letowski et S. Letowski : Localization error : Accuracy and precision of auditory localization. *In* P. Strumillo, éd. : *Advances in Sound Localization*, chap. 4. IntechOpen, Rijeka, 2011.

H. Levitt : Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B):467–477, 1971, doi : 10.1121/1.1912375.

J. Lewald, W. H. Ehrenstein et R. Guski : Spatio-temporal constraints for auditory–visual integration. *Behavioural Brain Research*, 121(1-2):69–79, 2001, doi : 10.1016/S0166-4328(00)00386-7.

R. Likert : A technique for the measurement of attitudes. *Archives of Psychology*, 22 140:55–55, 1932.

A. D. Little, D. H. Mershon et P. H. Cox : Spectral content as a cue to perceived auditory distance. *Perception*, 21(3):405–416, 1992, doi : 10.1068/p210405.

J. M. Loomis, R. L. Klatzky, J. W. Philbeck et R. G. Golledge : Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics*, 60(6):966–980, 1998, doi : 10.3758/BF03211932.

D. Loth : Biophysique sensorielle. Acoustique physiologique. *Techniques de l'Ingénieur*, p. 10, 1997.

E. A. Macpherson et J. C. Middlebrooks : Listener weighting of cues for lateral angle : The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, 111(5):2219, 2002, doi : 10.1121/1.1471898.

L. E. Marks : Binaural summation of the loudness of pure tones. *The Journal of the Acoustical Society of America*, 64(1):107–113, juil. 1978, doi : 10.1121/1.381976.

L. E. Marks et M. Florentine : Measurement of loudness, Part I : Methods, problems, and pitfalls. *In* M. Florentine, A. N. Popper et R. R. Fay, éds : *Loudness*, p. 17–56. Springer New York, New York, NY, 2011. ISBN 978-1-4419-6712-1.

J. Marozeau, M. Epstein, M. Florentine et B. Daley : A test of the binaural equal-loudness-ratio hypothesis for tones. *The Journal of the Acoustical Society of America*, 120 (6):3870–3877, déc. 2006, doi : 10.1121/1.2363935.

D. McFadden : Masking-level differences determined with and without interaural disparities in masker intensity. *The Journal of the Acoustical Society of America*, 44(1):212–223, 1968.

D. McFadden, L. A. Jeffress et J. R. Lakey : Differences of interaural phase and level in detection and lateralization : 1000 and 2000 Hz. *The Journal of the Acoustical Society of America*, 52(4B):1197–1206, oct. 1972, doi : 10.1121/1.1913232.

P. McGregor, A. G. Horn et M. A. Todd : Are familiar sounds ranged more accurately ? *Perceptual and Motor Skills*, 61(3_suppl):1082–1082, 1985, doi : 10.2466/pms.1985.61.3f.1082.

D. H. Mershon, D. H. Desaulniers, T. L. Amerson et S. A. Kiefer : Visual capture in auditory distance perception : Proximity image effect reconsidered. *Journal of Auditory Research*, 1980.

D. H. Mershon, D. H. Desaulniers, S. A. Kiefer, T. L. Amerson et J. T. Mills : Perceived loudness and visually-determined auditory distance. *Perception*, 10(5):531–543, 1981, doi : 10.1068/p100531.

D. H. Mershon et L. E. King : Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics*, 18(6):409–415, 1975.

S. Meunier, S. Savel, J. Chatron et G. Rabau : Interindividual differences in directional loudness. *The Journal of the Acoustical Society of America*, 140(4):3268–3268, 2016, doi : 10.1121/1.4970368.

S. Meunier, S. Savel, J. Chatron et G. Rabau : Binaural loudness of moving sources in free field : perceptual measurements versus at-ear levels. *In Proceedings of Fechner Day*, p. 5, Lüneburg - Germany, 2018.

J. C. Middlebrooks et D. M. Green : Directional dependence of interaural envelope delays. *The Journal of the Acoustical Society of America*, 87(5):2149–2162, 1990, doi : 10.1121/1.399183.

J. C. Middlebrooks et D. M. Green : Sound localization by human listeners. *Annual Review Of Psychology*, 42:135–159, 1991.

J. C. Middlebrooks, J. C. Makous et D. M. Green : Directional sensitivity of sound-pressure levels in the human ear canal. *The Journal of the Acoustical Society of America*, 86(1):89–108, 1989.

G. A. Miller : The perception of short bursts of noise. *The Journal of the Acoustical Society of America*, 20(2):160–170, mars 1948, doi : 10.1121/1.1906359.

A. W. Mills : On the minimum audible angle. *The Journal of the Acoustical Society of America*, 30(4):237–246, avr. 1958, doi : 10.1121/1.1909553.

A. W. Mills : Auditory localization. *In* J. V. Tobias, éd. : *Foundations of modern auditory theory*, p. 303–348. Academic Press, New York, NY, 1972.

A. Miśkiewicz, B. Scharf, R. Hellman et C. Meiselman : Loudness adaptation at high frequencies [a]. *The Journal of the Acoustical Society of America*, 94(3):1281–1286, sept. 1993, doi : 10.1121/1.408154.

K. Mohrmann : Lautheitskonstanz im entfernungswechsel. *Zeitschrift fur Psychologie*, 145:145–199, 1939.

B. C. J. Moore et B. R. Glasberg : A revision of Zwicker's loudness model. *Acta Acustica united with Acustica*, 82(2):335–345, 1996.

B. C. J. Moore : Frequency selectivity, masking and the critical band. *In An Introduction to the Psychology of Hearing*, p. 65–126. BRILL, Leiden, Netherlands, 2012a. ISBN 978-1-78052-038-4.

B. C. J. Moore : The nature of sound and the structure and function of the auditory system. *In An Introduction to the Psychology of Hearing*, p. 1–53. BRILL, Leiden, Netherlands, 2012b. ISBN 978-1-78052-038-4.

B. C. J. Moore : Space perception. *In An Introduction to the Psychology of Hearing*, p. 245–250. BRILL, Leiden, Netherlands, 2012c. ISBN 978-1-78052-038-4.

B. C. J. Moore : Speech perception. *In An Introduction to the Psychology of Hearing*, p. 299–332. BRILL, Leiden, Netherlands, 2012d. ISBN 978-1-78052-038-4.

B. C. J. Moore et B. R. Glasberg : Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, 74(3):750–753, sept. 1983, doi : 10.1121/1.389861.

B. C. J. Moore et B. R. Glasberg : Modeling binaural loudness. *The Journal of the Acoustical Society of America*, 121(3):1604–1612, 2007, doi : 10.1121/1.2431331.

B. C. J. Moore, B. R. Glasberg, A. Varathanathan et J. Schlittenlacher : A loudness model for time-varying sounds incorporating binaural inhibition. *Trends in Hearing*, 20, 2016, doi : 10.1177/2331216516682698.

B. C. Moore, B. R. Glasberg et T. Baer : A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, 45(4):224–240, 1997.

B. E. Mulligan, L. S. Goodman, D. P. Gleisner et M. L. Faupel : Steps in loudness summation. *The Journal of the Acoustical Society of America*, 77(3):1141–1154, mars 1985, doi : 10.1121/1.392178.

W. A. Munson : The growth of auditory sensation. *The Journal of the Acoustical Society of America*, 19(4):584–591, 1947, doi : 10.1121/1.1916525.

H. Møller, M. F. Sørensen, D. Hammershøi et C. B. Jensen : Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, 43(5):300–321, 1995.

S. Namba : The relation between loudness and rise time as a function of energy. *The Journal of the Acoustical Society of Japan*, 30:144–150, 1974.

E. C. Odgaard, Y. Arieh et L. E. Marks : Brighter noise : Sensory enhancement of perceived loudness by concurrent visual stimulation. *Cognitive, Affective, & Behavioral Neuroscience*, 4 (2):127–132, 2004.

M. Otani, T. Hirahara et S. Ise : Numerical study on source-distance dependency of head-related transfer functions. *The Journal of the Acoustical Society of America*, 125(5):3253, 2009, doi : 10.1121/1.3111860.

M. Paquier, N. Côté, F. Devillers et V. Koehl : Interaction between auditory and visual perceptions on distance estimations in a virtual environment. *Applied Acoustics*, 105:186–199, 2016, doi : 10.1016/j.apacoust.2015.12.014.

M. Paquier et V. Koehl : Discriminability of the placement of supra-aural and circumaural headphones. *Applied Acoustics*, 93:130–139, 2015, doi : 10.1016/j.apacoust.2015.01.023.

E. Parizet et V. Koehl : Influence of train colour on loudness judgments. *Acta Acustica united with Acustica*, 97(2):347–349, 2011, doi : 10.3813/AAA.918414.

C. Patsouras, T. Filippou et H. Fastl : Influences of color on the loudness judgement. *In Proceedings of Forum Acusticum*, Sevilla, Spain, 2002.

J. Petersen : Estimation of loudness and apparent distance of pure tones in a free field. *Acta Acustica united with Acustica*, 70:5, 1990.

I. Pollack : On the measurement of the loudness of speech. *The Journal of the Acoustical Society of America*, 24(3):323–324, 1952, doi : 10.1121/1.1906900.

E. Ponsot, P. Susini et S. Meunier : A robust asymmetry in loudness between rising- and falling-intensity tones. *Attention, Perception, & Psychophysics*, 77(3):907–920, 2015, doi : 10.3758/s13414-014-0824-y.

R. Preibisch-Effenberger : *Die Schallokalisationsfähigkeit des Menschen und ihre audiometrische Verwendung zur klinischen Diagnostik [The human faculty of sound localization and its audiometric application to clinical diagnostics]*. Thèse de doctorat, Technische Universität, Dresden, 1966.

D. R. Proffitt, M. Bhalla, R. Gossweiler et J. Midgett : Perceiving geographical slant. *Psychonomic Bulletin & Review*, 2(4):409–428, 1995, doi : 10.3758/BF03210980.

R. Pujol : Voyage au centre de l'audition. `http://www.cochlea.eu/`, 2018.

S. Puria, W. T. Peake et J. J. Rosowski : Sound-pressure measurements in the cochlear vestibule of human-cadaver ears. *The Journal of the Acoustical Society of America*, 101(5): 2754–2770, mai 1997, doi : 10.1121/1.418563.

T. Rader, M. Morinaga, T. Matsiu, H. Fastl, S. Kuwano et S. Namba : Crosscultural effects in audio-visual interactions. *In Proceedings of Meeting of the Technical Committee on Noise and Vibration of the Acoustical Society of Japan*, Tokyo, Japan, 2004.

L. Rayleigh : On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(74):214–232, 1907, doi : 10.1080/14786440709463595.

M. Rébillat, X. Boutillon, E. Corteel et B. F. G. Katz : Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments. *ACM Transactions on Applied Perception*, 9(4):1–17, 2012, doi : 10.1145/2355598.2355602.

H. Remmers et H. Prante : Untersuchung zur richtungsabhängigkeit der lautstärkeempfindung von breitbandigen schallen [Investigation of the directional dependence of the loudness of broadband sounds]. *Fortschritte der Akustik–Deutsche Arbeitsgemeinschaft für Akustik (DAGA)*, p. 537–540, 1991.

R. S. Renner, B. M. Velichkovsky et J. R. Helmert : The perception of egocentric distances in virtual environments - A review. *ACM Computing Surveys*, 46(2):1–40, 2013, doi : 10.1145/2543581.2543590.

G. S. Reynolds et S. S. Stevens : Binaural summation of loudness. *The Journal of the Acoustical Society of America*, 32(10):1337–1344, 1960, doi : 10.1121/1.1907903.

D. W. Robinson et R. S. Dadson : A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, 7(5):166–181, mai 1956, doi : 10.1088/0508-3443/7/5/302.

D. W. Robinson et L. S. Whittle : The loudness of directional sound fields. *Acta Acustica united with Acustica*, 10(2):74–80, 1960.

D. E. Robinson et L. A. Jeffress : Effect of varying the interaural noise correlation on the detectability of tonal signals. *The Journal of the Acoustical Society of America*, 35(12):1947–1952, 1963.

L. Robles et M. A. Ruggero : Mechanics of the mammalian cochlea. *Physiological reviews*, 81(3):1305–1352, 2001.

L. D. Rosenblum et C. A. Fowler : Audiovisual investigation of the loudness-effort effect for speech and nonspeech events. *Journal of Experimental Psychology : Human Perception and Performance*, 17:976–985, 1991, doi : https://psycnet.apa.org/doi/10.1037/0096-1523.17.4.976.

M. K. Russell et A. L. Schneider : Sound source perception in a two-dimensional setting : Comparison of action and nonaction-based response tasks. *Ecological Psychology*, 18(3):223–237, 2006, doi : 10.1207/s15326969eco1803_4.

S. Rzhevkin : *A Course of Lectures on the Theory of Sound.* Pergamon Press, Oxford, 1963.

F. Salmon, E. Hendrickx, N. Épain, Q. George et M. Paquier : The influence of the sound source on perceived differences between binaurally rendered sound spaces. *In Proceedings of the AES International Conference on Headphone Technology*, San Francisco, CA, US, 2019.

B. Scharf : Loudness adaptation. *In* J. Tobias et E. Schubert, éds : *Hearing research and theory.* Academic Press, New York, NY, 1983. ISBN 978-0-12-312102-8.

B. Scharf et D. Fishken : Binaural summation of loudness : Reconsidered. *Journal of Experimental Psychology*, 86(3):374–379, 1970, doi : 10.1037/h0030159.

B. Scharf : Dichotic summation of loudness. *The Journal of the Acoustical Society of America*, 45(5):1193–1205, mai 1969, doi : 10.1121/1.1911590.

B. Scharf : Loudness. *In* E. C. Carterette et M. P. Friedman, éds : *Hearing*, Handbook of Perception, p. 187 – 242. Academic Press, 1978. ISBN 978-0-12-161904-6.

C. Schorkhuber, M. Zaunschirm et R. Holdrich : Binaural rendering of ambisonic signals via magnitude least squares. *In Proceedings of the DAGA*, Munich, Germany, 2018.

L. Schwarz : Zur theorie der beugung einer ebenen schallwelle an der kugel. *Akustische Zeitschrift*, 8:91–117, 1943.

E. A. G. Shaw et J. E. Piercy : Physiological noise in relation to audiometry. *The Journal of the Acoustical Society of America*, 34(5):745–745, mai 1962, doi : 10.1121/1.1937329.

R. N. Shepard : *Mind sights : Original visual illusions, ambiguities, and other anomalies, with a commentary on the play of mind in perception and art.* W H Freeman/Times Books/ Henry Holt & Co, New York, NY, US, 1990. ISBN 0-7167-2134-1 (Hardcover) ; 0-7167-2133-3 (Paperback).

S. Shigenaga : The constancy of loudness and of acoustic distance. *Bulletin of the Faculty of Literature of Kyushu University*, 9:289–333, 1965.

V. P. Sivonen : Directional loudness and binaural summation for wideband and reverberant sounds. *The Journal of the Acoustical Society of America*, 121(5):2852–2861, mai 2007, doi : 10.1121/1.2717497.

V. P. Sivonen et W. 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):2965–2980, 2006, doi : 10.1121/1.2184268.

V. P. Sivonen et W. Ellermeier : Binaural loudness for artificial-head measurements in directional sound fields. *Journal of the Audio Engineering Society*, 56(6):452–461, 2008.

V. P. Sivonen et W. Ellermeier : Binaural Loudness. *In* M. Florentine, A. N. Popper et R. R. Fay, éds : *Loudness*, p. 169–197. Springer New York, New York, NY, 2011. ISBN 978-1-4419-6712-1.

V. P. Sivonen, P. Minnaar et W. Ellermeier : Effect of direction on loudness in individual binaural synthesis. *In Proceedings of the AES 118th Convention*, p. 6512, Barcelona, Spain, 2005.

A. M. Small, J. F. Brandt et P. G. Cox : Loudness as a function of signal duration. *The Journal of the Acoustical Society of America*, 34(4):513–514, avr. 1962, doi : 10.1121/1.1918157.

D. R. Soderquist et R. D. Shilling : Loudness and the binaural masking level difference. *Bulletin of the Psychonomic Society*, 28(6):553–555, 1990.

S. S. Stevens : A scale for the measurement of a psychological magnitude : loudness. *Psychological Review*, 43(5):405–416, 1936, doi : 10.1037/h0058773.

S. S. Stevens : The measurement of loudness. *The Journal of the Acoustical Society of America*, 27(5):815–829, 1955, doi : https://doi.org/10.1121/1.1908048.

S. S. Stevens : The direct estimation of sensory magnitudes : Loudness. *The American Journal of Psychology*, 69(1):1, 1956, doi : 10.2307/1418112.

S. S. Stevens et M. Guirao : Loudness, reciprocality, and partition scales. *The Journal of the Acoustical Society of America*, 34(9B):1466–1471, 1962, doi : 10.1121/1.1918370.

M. R. Stinson et B. W. Lawton : Specification of the geometry of the human ear canal for the prediction of sound-pressure level distribution. *The Journal of the Acoustical Society of America*, 85(6):2492–2503, 1989, doi : 10.1121/1.397744.

P. Susini et S. McAdams : Loudness asymmetry ratings between accelerating and decelerating car sounds. *The Journal of the Acoustical Society of America*, 123(5):3307–3307, 2008, doi : 10.1121/1.2933739.

P. Susini, S. McAdams et B. K. Smith : Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings. *Acta Acustica united with Acustica*, 93 (4):623–631, 2007.

Y. Suzuki, K. Abe, K. Ozawa et T. Sone : Factors for perceiving sound environments and the effects of visual and verbal information on these factors. *In Contributions to Psychological Acoustics (A. Schick et al. Eds), BIS Oldenburg*, p. 209–232, 2000.

Y. Suzuki et H. Takeshima : Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America*, 116(2):918–933, août 2004, doi : 10.1121/1.1763601.

E. Terhardt : On the perception of periodic sound fluctuations (roughness). *Acta Acustica united with Acustica*, 30:13, 1974.

W. R. Thurlow et C. E. Jack : Certain determinants of the "ventriloquism effect". *Perceptual and Motor Skills*, 36(3_suppl):1171–1184, 1973, doi : 10.2466/pms.1973.36.3c.1171.

T. H. Townsend et D. P. Goldstein : Suprathreshold binaural unmasking. *The Journal of the Acoustical Society of America*, 51(2B):621–624, 1972, doi : 10.1121/1.1912884.

C. Trahiotis et L. R. Bernstein : Lateralization of bands of noise and sinusoidally amplitude-modulated tones : Effects of spectral locus and bandwidth. *The Journal of the Acoustical Society of America*, 79(6):1950–1957, 1986, doi : 10.1121/1.393202.

C. W. Tyler : Spatial organization of binocular disparity sensitivity. *Vision Research*, 15 (5):583–590, 1975, doi : 10.1016/0042-6989(75)90306-5.

J. Udesen, T. Piechowiak et F. Gran : The effect of vision on psychoacoustic testing with headphone-based virtual sound. *Journal of the Audio Engineering Society*, 63(7/8):552–561, 2015, doi : 10.17743/jaes.2015.0061.

J. Van Soest : Rictingshooren bij sinusoorrnige geluidstrillingen. *Physica*, 9(271-282), 1929.

M. Vannier : *Sonie de champs acoustiques stationnaires en situation d'écoute dichotique*. Thèse de doctorat, INSA Lyon, 2015.

N. F. Viemeister et S. P. Bacon : Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones. *The Journal of the Acoustical Society of America*, 84(1):172–178, 1988, doi : 10.1121/1.396961.

E. M. von Hornbostel et M. Wertheimer : *Über die Wahrnehmung der Schallrichtung*. Sitzungsber. Akad. Wiss., Berlin, 1920.

M. Wallace, G. Roberson, W. Hairston, B. Stein, J. Vaughan et J. Schirillo : Unifying multisensory signals across time and space. *Experimental Brain Research*, 158(2), 2004, doi : 10.1007/s00221-004-1899-9.

H. Wallach : The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4):339–368, 1940, doi : 10.1037/h0054629.

D. H. Warren : Spatial localization under conflict conditions : Is there a single explanation ? *Perception*, 8(3):323–337, 1979, doi : 10.1068/p080323.

R. M. Warren : Anomalous loudness function for speech. *The Journal of the Acoustical Society of America*, 54(2):390–396, 1973, doi : 10.1121/1.1913590.

T. C. Weerts et W. R. Thurlow : The effects of eye position and expectation on sound localization. *Perception & Psychophysics*, 9(1):35–39, 1971, doi : 10.3758/BF03213025.

R. B. Welch : Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. *In* G. Aschersleben, T. Bachmann et J. Müsseler, éds : *Cognitive Contributions to the Perception of Spatial and Temporal Events*, vol. 129 de *Advances in Psychology*, p. 371 – 387. North-Holland, 1999.

R. B. Welch et D. H. Warren : Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88(3):638–667, 1980, doi : 10.1037/0033-2909.88.3.638.

R. G. Wettschureck : The absolute difference limen of directional perception in the median plane under conditions of both, natural hearing and hearing with artificial-head-system. *Acta Acustica united with Acustica*, 28(4):197–208, 1973.

R. Wettschurek : Über unterschiedsschwellen beim richtungshören in der medianebene [on difference thresholds in connection with directional hearing in the median plane]. *Gemeinschaftstagung für Akustik und Schwingungstechnik, Berlin*, p. 385–388, 1970.

S. Whilby, M. Florentine, E. Wagner et J. Marozeau : Monaural and binaural loudness of 5- and 200-ms tones in normal and impaired hearing. *The Journal of the Acoustical Society of America*, 119(6):3931–3939, juin 2006, doi : 10.1121/1.2193813.

M. G. Wisniewski, E. Mercado, K. Gramann et S. Makeig : Familiarity with speech affects cortical processing of auditory distance cues and increases acuity. *PLoS ONE*, 7(7):e41025, juil. 2012, doi : 10.1371/journal.pone.0041025.

B. G. Witmer et M. J. Singer : Measuring presence in virtual environments : A presence questionnaire. *Presence : Teleoperators and Virtual Environments*, 7(3):225–240, 1998, doi : 10.1162/105474698565686.

J. Wittmann : Beitrage zur analyse des horens bei dikotischer reizaufnahme. *Archiv für die gesamte Psychologie*, 51(1925):21–122, 1925.

R. S. Woodworth : Auditory space perception. *In Experimental Psychology*, p. 349 – 361. Holt, Rinehart and Winston, New York, NY, 1954.

G. K. Yates : Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions. *Hearing Research*, 50(1-2):145–162, 1990.

G. K. Yates : Cochlear structure and function. *In Hearing*, p. 41–74. Brian C.J. Moore, Academic, San Diego, CA, 1995.

W. A. Yost : The masking-level difference and overall masker level : Restating the internal noise hypothesis. *The Journal of the Acoustical Society of America*, 83(4):1517–1521, 1988.

P. Zahorik et F. L. Wightman : Loudness constancy with varying sound source distance. *Nature Neuroscience*, 4(1):78–83, 2001, doi : 10.1038/82931.

P. Zahorik : Estimating sound source distance with and without vision :. *Optometry and Vision Science*, 78(5):270–275, 2001, doi : 10.1097/00006324-200105000-00009.

P. Zahorik : Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America*, 111(4):1832–1846, 2002, doi : 10.1121/1.1458027.

P. Zahorik : Auditory and visual distance perception : The proximity-image effect revisited. *The Journal of the Acoustical Society of America*, 113(4):2270–2270, 2003, doi : https://doi.org/10.1121/1.4780524.

P. Zahorik : Auditory distance perception in humans : A summary of past and present research. *Acta Acustica united with Acustica*, 91:12, 2005.

E. Zwicker : Ein verfahren zur beredinung der lautstärke. *Acta Acustica united with Acustica*, 10(4):304–308, 1960.

E. Zwicker : Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*, 33(2):248–248, fév. 1961, doi : 10.1121/1.1908630.

E. Zwicker, G. Flottorp et S. S. Stevens : Critical band width in loudness summation. *The Journal of the Acoustical Society of America*, 29(5):548–557, mai 1957, doi : 10.1121/1.1908963.

E. Zwicker et U. T. Zwicker : Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution. *The Journal of the Acoustical Society of America*, 89(2):756–764, fév. 1991, doi : 10.1121/1.1894635.

E. Zwicker et G. B. Henning : On the effect of interaural phase differences on loudness. *Hearing Research*, 53(1):141–152, 1991.

E. Zwicker et B. Scharf : A model of loudness summation. *Psychological Review*, 72(1):3–26, 1965, doi : 10.1037/h0021703.

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

UBO Université de Bretagne Occidentale

Titre : Influence de la position d'une source sur le niveau sonore perçu

Mots clés : Sonie, Azimut, Distance, Interactions audiovisuelles

Résumé : Dans notre environnement quotidien, les sons nous parviennent depuis différentes distances et différentes directions. La sonie, qui désigne le niveau sonore perçu, dépend de la position relative (en distance et en direction) de la source sonore par rapport à l'auditeur. Cette thèse comporte deux parties distinctes s'intéressant à ces relations, l'une liant la sonie à l'azimut d'une source sonore et l'autre à la distance d'une source sonore frontale. La première partie étudie plus précisément les effets de sonie directionnelle liés aux différences interaurales de temps pour des sons basses fréquences. Ces différences, qui dépendent de l'azimut de la source et qui sont maximales lorsque la source se situe sur le côté de l'auditeur, provoquent une augmentation de sonie pour les sons basses fréquences. Cette augmentation de sonie est étudiée notamment en présence de bruit et en fonction de la validité écologique des conditions de présentation des sons. La seconde partie de cette thèse étudie deux types distincts de sonie (l'un correspondant au niveau sonore perçu « aux oreilles » et l'autre au niveau sonore perçu « à la source ») en fonction de la distance d'une source sonore en présence ou non d'indices visuels. Cette partie s'intéresse à des sons non-familiers (des bruits blancs) ainsi qu'à des sons familiers (des signaux de parole), ces derniers fournissant des informations aux auditeurs sur la puissance de leur source par le biais de leur timbre.

Title : Influence of the position of a sound source on loudness

Keywords : Loudness, Azimuth, Distance, Audiovisual interactions

Abstract : In our daily environment, we perceive sounds coming from different distances and different directions. The loudness of these sounds depends on the relative position (in distance and direction) of their sources. This thesis is structured in two distinct parts studying these relations. One part focuses on links between loudness and sound source azimuth and the other on links between loudness and frontal sound source distance. The first part focuses more precisely on directionnal effects due to interaural time differences for low-frequency sounds. These differences, which depend on the sound source azimuth and which are maximal when the source is located at the side of the listener, lead to an increase in loudness for low-frequency sounds. This loudness increase is studied when sounds are presented in noise and in presentation conditions with different ecological validity. The second part of this thesis focuses on two distinct kinds of loudness (one being the loudness “at the ears” and the other the loudness “at the source”) as a function of the distance of a sound source, with and without visual cues to the latter. This part studies unfamiliar (white noises) and familiar (speech) stimuli, the latters providing the listeners with information about their source power thanks to their timbre.