



HAL
open science

Perception de l'harmonicité et fusion harmonique chez l'adulte normo-entendant

Caroline Duprat

► **To cite this version:**

Caroline Duprat. Perception de l'harmonicité et fusion harmonique chez l'adulte normo-entendant. Sciences cognitives. 2018. dumas-02071297

HAL Id: dumas-02071297

<https://dumas.ccsd.cnrs.fr/dumas-02071297>

Submitted on 18 Mar 2019

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.

Année Universitaire 2017-2018

COLLÈGE SANTÉ

MÉMOIRE (UE 7.5) PRÉSENTÉ POUR L'OBTENTION DU

Certificat de Capacité d'Orthophoniste – Grade Master

Par Caroline DUPRAT

Née le 15 mai 1994 à Brive-la-Gaillarde

**Perception de l'harmonicité et fusion harmonique
chez l'adulte normo-entendant**

Sous la direction de Monsieur le Dr. Damien BONNARD

Soutenu le 26 juin 2018

Membres du jury :

Madame le Dr. Delphine LAMARGUE HAMEL, Orthophoniste, Responsable pédagogique du département d'orthophonie, Neurocentre Magendie, *Présidente du jury*

Monsieur le Pr. René DAUMAN, ORL, PU-PH, CHU de Bordeaux, *Rapporteur*

Madame Nathalie MICHAS, Orthophoniste, *Examinatrice*

Année Universitaire 2017-2018

COLLEGE SANTÉ

MÉMOIRE (UE 7.5) PRÉSENTÉ POUR L'OBTENTION DU
Certificat de Capacité d'Orthophoniste – Grade Master

Par Caroline DUPRAT

Née le 15 mai 1994 à Brive-la-Gaillarde

**Perception de l'harmonicité et fusion harmonique
chez l'adulte normo-entendant**

Sous la direction de Monsieur le Dr. Damien BONNARD

Soutenu le 26 juin 2018

Membres du jury :

Madame le Dr Delphine LAMARGUE-HAMEL, Orthophoniste, Responsable pédagogique du département d'orthophonie, Neurocentre Magendie, *Présidente du jury*

Monsieur le Pr René DAUMAN, ORL, PU-PH, CHU de Bordeaux, *Rapporteur*

Madame Nathalie MICHAS, Orthophoniste, *Examinatrice*

Titre : Perception de l'harmonicité et fusion harmonique chez l'adulte normo-entendant

Résumé :

Nous sommes capables de distinguer deux sons complexes (une voyelle ou une note de musique) produits en même temps, grâce à un phénomène acoustique particulier : la fusion harmonique. En d'autres termes, l'auditeur est capable d'extraire une hauteur unique d'un ensemble de composantes harmoniques d'une même source sonore.

L'objectif de ce mémoire est de comprendre l'origine du codage de la hauteur des sons complexes dans le système auditif en étudiant la détection d'inharmonicité et donc le phénomène de fusion harmonique, à travers deux expériences. Deux théories sont classiquement proposées pour expliquer le codage de la hauteur ; l'une repose sur des mécanismes spatiaux, l'autre sur des mécanismes temporels. Nous étudions notamment l'hypothèse d'un mécanisme temporel à l'origine dans ce codage.

Les résultats de l'expérience 1 indiquent que lorsque le système auditif est contraint d'utiliser une information temporelle, on retrouve une asymétrie dans une tâche de détection d'inharmonicité. Néanmoins, ces résultats pourraient aussi être expliqués par l'intervention d'un mécanisme spatial acquis avant la naissance jusqu'aux premiers mois de vie. Dans l'expérience 2, l'asymétrie est retrouvée dans toutes les conditions sauf dans celle où le niveau sonore est bas et les sons sont antiphasiques. Cela indique que le système binaural à lui seul est capable de détecter une inharmonicité. Ceci constitue une donnée nouvelle pour la compréhension de la fusion harmonique. Notre étude démontre que les mécanismes temporels jouent un rôle central dans les phénomènes de détection d'inharmonicité, sans exclure d'autres types de mécanismes associés.

Ces résultats ouvrent de nouvelles perspectives pour la recherche fondamentale et pour la pratique clinique en orthophonie. Il est notamment possible d'envisager une rééducation auditive basée sur un entraînement à la détection d'inharmonicité chez les patients atteints de surdité ou porteurs d'implants cochléaires, dans l'optique d'améliorer leurs capacités de fusion harmonique et ainsi leur compréhension de la parole dans un environnement bruyant.

Mots-clés :

Harmonicité, Fusion harmonique, Hauteur, Codage temporel.

Title : Harmonicity perception and harmonic fusion in normal hearing adults

Abstract :

We are able to distinguish between two complex sounds (a vowel or a music note) produced at the same time, thanks to a particular acoustic phenomenon : the harmonic fusion. In other words, the listener is able to extract a single pitch from a set of harmonic components from the same sound source.

The aim of this thesis is to understand the origin of pitch coding of complex sounds in the auditory system based on the detection of inharmonicity, and therefore the phenomenon of harmonic fusion, through two experiments. Two theories are commonly proposed to explain this coding ; one based on spatial clues and the other on temporal clues. We hypothesized that a temporal mechanism is involved in this coding.

Results of experiment 1 indicates that when the auditory system is forced to use temporal information, we find an asymmetry in an inharmonicity detection task. On the other hand, these results could also be explained by the intervention of a spatial mechanism acquired before the birth until the first months of life. In experiment 2, the asymmetry of experiment 1 is found under all conditions except when the sound level is low and the sounds are antiphase. This means that binaural system alone is able to detect an inharmonicity. This is new data, which adds to our understanding of harmonic fusion. Our study demonstrates that temporal mechanisms play the main role in inharmonicity detection phenomenon, without excluding other kinds of associated mechanisms.

With our results in mind, fundamental research will continue to investigate this topic. Our research also has implications for the clinical practice of speech therapy. For example, it is possible to imagine auditory rehabilitation based on inharmonicity detection training in patients with hearing loss or cochlear implants, in order to improve their harmonic fusion capabilities and thus their speech understanding in a noisy environment.

Keywords :

Harmonicity, Harmonic fusion, Pitch, Temporal theory.

Remerciements

Je remercie le Dr Damien Bonnard pour son accompagnement chaleureux tout au long de mon mémoire, et son écoute attentive lors de mes questionnements. J'ai bénéficié d'un encadrement bienveillant et rassurant, qui m'a permis d'avancer le plus sereinement possible dans ce travail.

Je remercie également le Dr Laurent Demany, qui a permis la réalisation de ce projet et m'a offert l'opportunité de découvrir le monde de la recherche.

Je remercie Sylvie Lautissier pour son écoute et ses conseils dans ce projet de mémoire depuis son commencement. Elle a su m'encourager dans les moments de doutes pour aller jusqu'au bout de ma démarche.

J'adresse mes remerciements aux membres du jury, Dr Delphine Lamargue-Hamel, Pr René Dauman et Mme Nathalie Michas pour avoir accepté de participer à ma soutenance.

Je remercie les sujets de l'expérience, des étudiants en orthophonie ou en médecine essentiellement, pour leur coopération et leur assiduité.

Je remercie l'ensemble de l'équipe pédagogique du département d'orthophonie.

Mme Lamargue-Hamel, pour ses encouragements, sa bienveillance et son engouement. Elle a soutenu mon projet de mémoire qui n'est possible que depuis la masterisation des études d'orthophonie.

Mme Makni pour son investissement, les coordinatrices de stage et tous les intervenants qui ont proposé au fil de ces cinq ans des enseignements dans leur spécialité.

Je tiens à remercier les orthophonistes qui m'ont accueillie lors des différents stages pendant mes études. J'ai pu partager leur expérience du métier et leur expertise dans des domaines de l'orthophonie qui m'intéressent particulièrement aujourd'hui.

Je remercie particulièrement, Noémie qui fut un soutien fort et ce depuis le début de mes études, Victor pour ses encouragements, Guillaume et Victor pour leur investissement dans la relecture. Je n'oublie pas mes amies, *Apolline, Catlina, Laure, Laurie et Mathilde*, notre bande de six, unies pour le meilleur et pour le pire.

Enfin, je remercie mes parents et mon frère pour leur soutien inconditionnel pendant cette période charnière dans mes études.

Sommaire

Introduction	7
I / Système auditif	8
1) La cochlée	8
2) Traitement rétrocochléaire du signal : les voies auditives	9
3) Traitement central du signal.....	10
4) Système binaural.....	11
II / Codage de la hauteur dans le système auditif	12
1) Données physiques.....	12
a) <i>Fréquence</i>	12
b) <i>Hauteur</i>	12
c) <i>Analyse spectrale</i>	12
d) <i>Périodicité</i>	13
2) Codage de la hauteur des sons	13
a) <i>Codage de la hauteur d'un son pur</i>	14
➤ <i>Théorie spatiale</i>	14
➤ <i>Théorie temporelle</i>	14
b) <i>Codage de la hauteur d'un son complexe harmonique (cf annexe 3)</i>	15
➤ <i>Théorie spatiale</i>	15
➤ <i>Théorie temporelle</i>	16
3) Codage de la hauteur par le système binaural	18
4) Battements de consonnance imparfaite.....	18
III / Harmonicité et fusion harmonique	19
1) Harmonicité.....	19
2) Fusion harmonique	20
a) <i>Définition de la fusion harmonique</i>	20
b) <i>Etudes réalisées</i>	20
3) Détection d'inharmonicité	23
4) Données neurophysiologiques.....	23
a) <i>Traitement des sons</i>	23
b) <i>Détection d'inharmonicité</i>	24
c) <i>Fusion harmonique</i>	24
IV / Intelligibilité de la parole	24
1) Acoustique des voyelles	24
2) Structure temporelle fine et enveloppe temporelle d'un son.....	25
a) <i>Enveloppe temporelle</i>	25
b) <i>Structure temporelle fine</i>	25
c) <i>Rôle de ces paramètres dans la compréhension de la parole</i>	26
3) Fusion harmonique et intelligibilité de la parole.....	26
4) Surdité neurosensorielle et implant cochléaire.....	27
Problématique et hypothèse	28
Méthodologie générale	29
1) Sujets	29
2) Conditions expérimentales et matériel	29
3) Etapes de calibrage.....	30
a) <i>Audiogramme</i>	30
b) <i>Détermination des seuils de détection des sons</i>	30
4) Analyse des résultats	31

Expérience 1 : détection d’inharmonicité dans un stimulus constitué de deux groupes d’harmoniques résolues et irrésolues	32
1) Méthode.....	32
2) Résultats	34
a) Résultats (moyennes de tous les sujets).....	34
b) Résultats individuels.....	35
3) Discussion	36
Expérience 2 : détection d’inharmonicité dans un stimulus constitué de sons purs résolus dont le premier harmonique peut être déphasé interaurealement. 38	38
1) Méthode.....	38
2) Résultats	40
a) Résultats (moyennes de tous les sujets).....	40
b) Résultats individuels.....	41
3) Discussion	42
Discussion générale	44
Perspectives	47
Conclusion	50
Bibliographie	51
Figures.....	55
Annexes	56

Table des annexes

Annexe 1 : Lexique.....	56
Annexe 2 : Conversions.....	57
Annexe 3 : Codage de la hauteur d’un son complexe par le système auditif.....	58

Introduction

La perception auditive soulève des interrogations depuis des dizaines d'années. En effet, même avec un audiogramme normal, certains sujets ressentent une gêne pour comprendre la parole dès qu'ils se trouvent dans un environnement bruyant.

Le décodage de la parole dans le silence est effectué grâce à la reconnaissance des phonèmes et aux fonctions psycho-acoustiques sous-jacentes. La reconnaissance des phonèmes s'effectue en trois étapes (Dupont & Lejeune, 2010).

La 1^{ère} étape est *l'analyse acoustique* : l'auditeur traite l'information auditive suivant la structure temporelle et spectrale du signal acoustique. La 2^{ème} étape est *l'analyse phonétique* qui consiste à traiter ces informations acoustiques en spécifiant les traits distinctifs des sons langagiers. La 3^{ème} étape est *l'analyse phonémique* qui consiste à filtrer les informations sans valeur distinctive dans la langue et à activer le phonème correspondant aux traits phonétiques perçus. Cette reconnaissance des phonèmes, nécessaire au décodage de la parole, est sous-tendue par des fonctions psycho-acoustiques : la discrimination auditive (analyser et différencier les fréquences, en fait, les sons de la parole), la sélectivité (sélectionner ce que l'on veut entendre quel que soit l'environnement sonore), la séparativité (séparer les groupes de mots ou les phrases pour rendre le discours de notre interlocuteur intelligible et significatif) et la mémoire (stocker l'information auditive en mémoire à court terme puis en mémoire à long terme, pour reconnaître les mots et leur donner un sens).

Ces mêmes mécanismes entrent en jeu dans le traitement de l'information auditive en milieu bruyant. La tâche est bien plus difficile dans cette condition. En effet, la compréhension de la parole dans le bruit est une difficulté majeure rencontrée au quotidien par de nombreux sujets. Les personnes porteuses d'un implant cochléaire sont généralement satisfaites des améliorations prodiguées, toutes s'accordent sur les difficultés majeures de perception et de compréhension de la parole dans le bruit, contrairement à des personnes normo-entendantes.

Dans l'optique de ce mémoire, notre questionnement s'est d'abord porté sur l'intelligibilité de la parole dans des milieux bruyants chez les adultes porteurs d'implants cochléaires. Cependant, des difficultés d'ordre matériel et la complexité de tester des patients porteurs d'implants cochléaires (types de surdités, marques d'implants, stratégies de codage, réserve cognitive individuelle, éventuelle prise en charge en orthophonie, intensité de ce suivi, éventuelle formation musicale, etc.), ont orienté notre questionnement vers l'analyse acoustique de la reconnaissance

des phonèmes dans un environnement sonore complexe. Nous ne traiterons pas des aspects de la reconnaissance de la parole relatifs aux indices sémantiques (la signification des mots précédents et suivants, du thème), ni des indices syntaxiques (règles grammaticales), ni des indices périphériques (identité du locuteur, environnement d'écoute, capacités internes...), même si ces aspects peuvent avoir une importance considérable, en particulier dans un environnement bruyant. Nous nous concentrerons sur l'analyse acoustique des sons complexes, dans l'objectif de lever le voile sur un élément fondamental qui permet de comprendre la perception auditive chez les personnes normo-entendantes et par extension celle des personnes avec un implant cochléaire : la fusion harmonique. En psycho-acoustique, ce phénomène perceptuel joue un rôle important dans l'analyse des scènes auditives de la vie quotidienne (environnement bruyant) où les auditeurs doivent différencier des sources sonores concomitantes.

La compréhension de la perception de l'harmonicité qui sous-tend la fusion harmonique pourrait nous permettre d'enrichir les données fondamentales de la littérature concernant le codage de la hauteur des sons complexes dans le système auditif et d'envisager des applications cliniques pour la rééducation des personnes atteintes de surdité neurosensorielle ou porteuses d'implants cochléaires.

I / Système auditif

1) La cochlée

La cochlée, organe sensoriel de l'oreille interne, est un tube spiralé qui est le siège de l'organe acoustique récepteur : l'organe de Corti. La fonction de la cochlée est de convertir les vibrations sonores mécaniques en impulsions neuronales que le cerveau peut interpréter. L'organe de Corti contient deux types de cellules ciliées : les cellules ciliées externes (CCE) et les cellules ciliées internes (CCI). Les mouvements mécaniques transmis à la platine de l'étrier provoquent des mouvements liquidiens dans la cochlée, qui entraînent une vibration à un endroit précis de la membrane basilaire correspondant à la fréquence de stimulation. Par leurs contractions, les CCE amplifient l'énergie fournie par la vibration et entraînent la dépolarisation des CCI qui génèrent des potentiels d'action dans les fibres afférentes du nerf auditif avec lesquelles elles font synapse. Dès cette étape et jusqu'au cortex auditif, les stimuli sont traités au niveau fréquentiel notamment par l'organisation tonotopique particulière de la cochlée, la base codant les fréquences aiguës et l'apex les fréquences graves (Danziger & Alamowitch, 2016; Dumont, 2008). La cochlée est

responsable de l'encodage des signaux envoyés au cerveau, qui s'effectue de deux manières très différentes : spatialement et temporellement (Hartmann, 2013).

La membrane basilaire est constituée de filtres auditifs passe-bandes (Fletcher, 1940). Un filtre auditif laisse passer un nombre restreint de fréquences dans une bande plus ou moins large autour d'une fréquence centrale. Chaque stimulation de la membrane basilaire par un son pur présenté à la cochlée entraîne une réponse dans un unique filtre. La largeur de bande de chaque filtre est d'environ 10 à 15 % de la fréquence centrale du filtre. La largeur absolue (en Hz) de chaque filtre augmente donc avec la fréquence centrale. Par conséquent, plus on se déplace vers les hautes fréquences à la base de la cochlée, plus ces filtres sont larges, donc moins ils sont sélectifs (cf annexe 3).

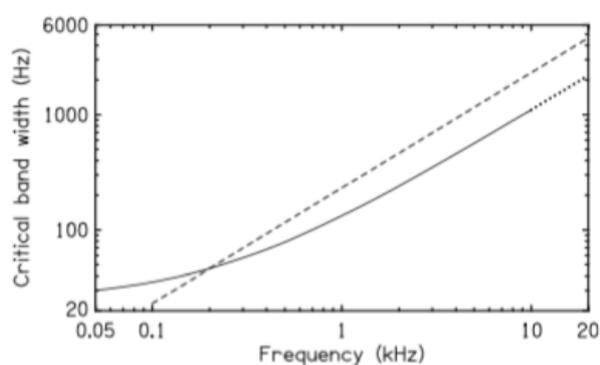


Figure 1 : Largeur de la bande passante en fonction de la fréquence centrale du filtre.

Dans le cas d'un son complexe harmonique, les premières harmoniques (les plus graves, de la 1^{ère} à la 8^{ème}) sont traitées une à une par des filtres auditifs distincts. Ces harmoniques sont alors dites *résolues*. Au contraire, à partir de la 8^{ème} harmonique (fréquences plus aiguës), les filtres auditifs sont de moins en moins sélectifs et laissent passer plusieurs harmoniques consécutives. Ces harmoniques, qui ne sont donc plus bien différenciées spatialement, sont dites *irrésolues* (Bonnard, thèse 2013).

2) Traitement rétrocochléaire du signal : les voies auditives

Le nerf auditif chemine dans le conduit auditif interne et pénètre dans le tronc cérébral au niveau de la protubérance. Le message bio-électrique chemine ainsi dans les voies auditives par des relais qui permettent aux sons de parvenir au cortex auditif primaire (Hartmann, 2013; Golestani, 2017; Dumont, 2008).

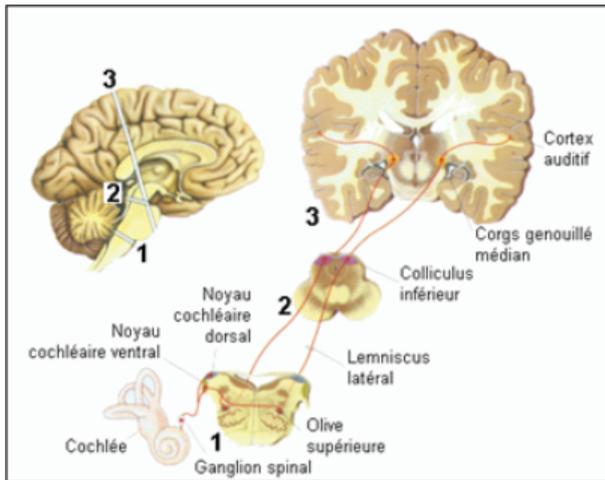


Schéma 1 : Schéma des voies auditives
 (adapté de Bear, Connors, & Paradiso, 1999)

Le premier relais est le *noyau cochléaire* (collection de cellules neurales) du tronc cérébral. Ensuite, les voies auditives se projettent vers le cortex de manière bilatérale avec une dominance controlatérale. L'information arrive au deuxième relais, le *complexe olivaire supérieur* dans le tronc cérébral. Les fibres atteignent ensuite le troisième relais, le *colliculus inférieur* dans le mésencéphale. Avant de gagner le cortex auditif, les fibres se dirigent vers le quatrième relais, dans le thalamus au niveau du *corps genouillé médian*. Du thalamus, l'information parvient au *cortex auditif*. En plus des voies ascendantes pré-citées, il y a des voies descendantes qui suivent le même trajet, permettant de moduler l'activité des centres sous-jacents et qui constituent ainsi une forme de rétroaction. La richesse de cette structure en innervations ascendantes et descendantes permet la constitution de boucles auditives thalamo-corticales (Dumont, 2008).

3) Traitement central du signal

Après un traitement sous-cortical d'un son ou de la parole entendue, la première région du cortex qui s'active est le cortex auditif primaire (lobe temporal). Il s'étale sur les deux tiers médians du gyrus de Heschl. L'information se répand par la suite au cortex auditif secondaire, comprenant le planum tempore, puis au cortex auditif associatif (bilatéralement le long du cortex temporal supérieur moyen et inférieur). Parmi les aires associatives, l'aire de Wernicke, importante pour la compréhension de la parole et l'aire de Broca, plutôt recrutée lors de la production de la parole. En réalité, il y a un recoupement entre les fonctions de production et de perception de ces régions. La communication entre ces aires se fait via le faisceau arqué. Le traitement central est effectué sous forme de réseaux (Golestani, 2017).

4) Système binaural

Deux ondes périodiques de même fréquence sont dites *en phase (homophasiques)* si leurs vibrations atteignent leur maximum et leur minimum simultanément (1).

Ces deux ondes sont dites *en opposition de phase (antiphasiques)* si l'une atteint son maximum lorsque l'autre atteint son minimum (2) (Fischetti, 2004).

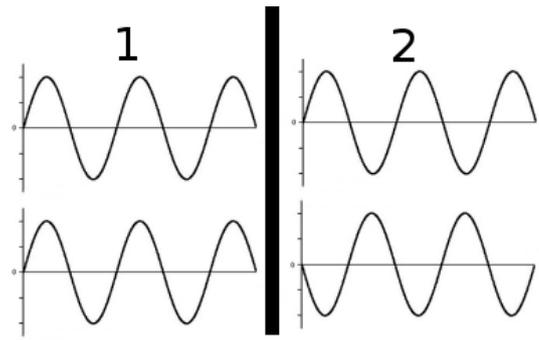


Figure 2 : Ondes en phase et en opposition de phase

Le déphasage consiste donc à introduire un décalage temporel entre les deux stimuli équivalents envoyés dans chacune des deux oreilles.

Le Binaural Masking Level Difference (BMLD) est une mesure de l'amélioration de la détectabilité d'un son pur qui peut se produire dans des conditions d'écoute binaurale. C'est la différence de seuil de détection du son pur (en dB) entre le cas où le son pur et le masque ont les mêmes relations de phase et de niveau sonore dans les deux oreilles (N_0S_0)* et le cas où les relations interaurales de phase et/ou de niveau sonore du son pur et du masque sont différentes ($N_0S\pi$)** (Moore, 2004).

*Dans la condition N_0S_0 (N pour noise et S pour son), un son pur présenté aux deux oreilles simultanément (bruit homophasique) est plongé dans le même bruit pour mesurer son seuil de détection dans ce bruit (intensité à partir de laquelle ce son est perçu dans le bruit).

**Dans la condition $N_0S\pi$, le bruit est homophasique et le son pur est en opposition de phase sur les deux oreilles (antiphasique, déphasé de 180° ou π radians).

Le seuil est plus bas (meilleur) dans la deuxième condition car le déphasage permet de mieux percevoir le son. La différence interaurale entre N et S crée l'amélioration du seuil (Gebhardt, Goldstein, & Robertson, 1972). Le BMLD est donc la différence entre le seuil de détection d'un son pur obtenu dans la condition N_0S_0 et le seuil obtenu dans la condition $N_0S\pi$. La valeur du BMLD est de 15 dB pour les basses fréquences (autour de 500 Hz) et diminue de 2-3 dB pour les fréquences au-dessus de 1500 Hz (Durlach & Colburn, 1978 cités par Moore en 2004).

Le système binaural est capable de détecter à lui seul un son antiphasique à condition qu'il ait un niveau sonore très faible. Dans des expériences psycho-acoustiques, le BMLD est ainsi utilisé pour déterminer une fourchette d'intensités dans laquelle un son pur n'est détectable que par le système binaural.

Ainsi, une des implications du phénomène de BMLD est que la détection et la discrimination de sons, parole comprise, sont améliorées quand le son et le masque ne coïncident pas dans l'espace, comme dans le phénomène de « cocktail party ».

II / Codage de la hauteur dans le système auditif

1) Données physiques

a) Fréquence

La fréquence est une dimension physique du son. La fréquence fondamentale d'un son complexe périodique est nommée f_0 . C'est la fréquence de répétition de l'onde sonore au cours du temps. La fréquence de sa 1^{ère} harmonique est $f_1 = 2f_0$, sa 2^{ème} harmonique est $f_2 = 3f_0$, etc. Les fréquences audibles s'étendent de 20 Hz à 20 000 Hz. Dans la cochlée, elles sont réparties sur la membrane basilaire selon une échelle logarithmique : entre 100 et 200 Hz, il y a la même distance qu'entre 1000 et 2000 Hz. En musique, les rapports de fréquences sont exprimés en tons et demi-tons. En acoustique, cette dimension est traduite en rapports de fréquences exprimés en *cents* (cf annexe 1 pour les conversions) (Fischetti, 2004).

b) Hauteur

L'ANSI (1994) définit la hauteur tonale comme « l'attribut de la sensation auditive permettant de classer les sons sur une échelle tonale s'étendant des graves vers les aigus. ». Plack et al. (2005) définissent la hauteur comme la sensation dont la variation est associée à une mélodie musicale. Cette dimension est la base de la mélodie dans la musique et de l'émotion dans la parole. Sans la hauteur, la musique ne serait que du rythme et de l'intensité, la parole serait alors monotone et robotique. La caractéristique physique corrélée à cette donnée perceptive est la *fréquence*.

c) Analyse spectrale

La décomposition d'un son en une succession d'harmoniques constitue le *spectre* de ce son (Fischetti, 2004). Le physicien Joseph Fourier a montré qu'il est possible de décomposer un son complexe périodique en une somme de sons purs sinusoïdaux, appelés *harmoniques*, dont les fréquences sont toutes des multiples d'une même fréquence fondamentale f_0 . Effectuer une *analyse spectrale* consiste à étudier la décomposition de sons complexes en sons purs d'après le

théorème de Fourier. Cette représentation mathématique de l'onde sonore est pertinente dans la mesure où la cochlée se comporte grossièrement comme un analyseur de spectre.

d) Périodicité

La périodicité est un phénomène physique où l'onde sonore se répète à l'identique en respectant une période temporelle (en seconde ou milliseconde). Il existe un rapport direct entre la période et la fréquence de répétition d'une onde sonore, caractérisé par la formule T (période) = $1/f$ (fréquence). Par exemple, une onde complexe avec des composantes à 200 et 600 Hz a une période de $1/200$ seconde. Au bout de $1/200$ s (5 ms), la composante à 200 Hz a subi exactement un cycle, et celle à 600 Hz en a subi exactement 3. Les deux composantes sont alors de nouveau à la phase initiale (Hartmann, 2013).

2) Codage de la hauteur des sons

Plusieurs études ont tenté d'expliquer le codage de la hauteur dans le système auditif. Une étude de Demany (1985) recherche les effets de l'entraînement à la discrimination des fréquences chez les auditeurs normo-entendants. La discrimination de fréquences à 200 Hz a été testée avant et après l'entraînement. Quatre groupes d'auditeurs ont reçu une formation dans quatre régions de fréquences différentes : 200, 360, 2500 et 6000 Hz. Il a été constaté que la formation à 200, 360 et 2500 Hz a fourni à tous une amélioration comparable de la performance de discrimination à 200 Hz alors que la formation à 6000 Hz a fourni moins d'amélioration. Ce résultat est cohérent avec l'idée que la discrimination de fréquences et la perception de la hauteur sont médiées par différents processus à des fréquences élevées (supérieures à 5000 Hz) et basses (inférieures à 5000 Hz). Une deuxième étude révèle que l'entraînement à la discrimination de fréquences améliore plus cette dernière lorsque les sons complexes contiennent des harmoniques résolues que lorsqu'ils sont composés uniquement d'harmoniques irrésolues (Grimault et al., 2003). Une troisième étude suggère que deux mécanismes jouent un rôle dans la discrimination de fréquences, l'un basé sur des changements dans le modèle d'excitation (un mécanisme spatial) et un basé sur le verrouillage de phase dans le nerf auditif (un mécanisme temporel) (Sek & Moore, 1995).

Au final, deux théories de codage de la hauteur des sons dans le système auditif (spatial et temporel) ont été proposées par plusieurs équipes de recherche, mais à ce jour il n'y a toujours pas de consensus (Colin D. , 2016).

a) Codage de la hauteur d'un son pur

Un *son pur* est caractérisé par une onde sinusoïdale, qui a une fréquence unique. Ce type de son est artificiel : une telle pureté ne se retrouve pas dans la nature. Il permet de réaliser des études en psycho-acoustique.

➤ *Théorie spatiale*

Dans la cochlée, la membrane basilaire est sélective pour la fréquence. Un son pur fait vibrer la membrane basilaire à un endroit précis correspondant à sa fréquence. Cette vibration provoque l'excitation des CCI caractéristiques de la fréquence du stimulus. En effet, chaque CCI réagit à une bande très étroite de fréquences, et vient stimuler une région spécifique du cortex auditif par l'intermédiaire du nerf auditif (Oxenham, 2013). La sélectivité en fréquences des neurones est appelée *analyse tonotopique*. Cela signifie que les différentes fréquences excitent des endroits légèrement différents dans le cerveau. Il est alors naturel de supposer que la sensation de hauteur est le résultat de l'analyse tonotopique. Par conséquent, selon cette théorie spatiale, nous entendons différentes hauteurs (graves ou aiguës) en raison de différents neurones excités (différentes localisations corticales).

➤ *Théorie temporelle*

Une CCI envoie un potentiel d'action sur une phase particulière de l'onde dans la cochlée. Ainsi, l'excitation d'une CCI et les potentiels d'action des neurones qui lui sont connectés se synchronisent avec le son entrant, c'est ce qu'on appelle le *verrouillage de phase*.

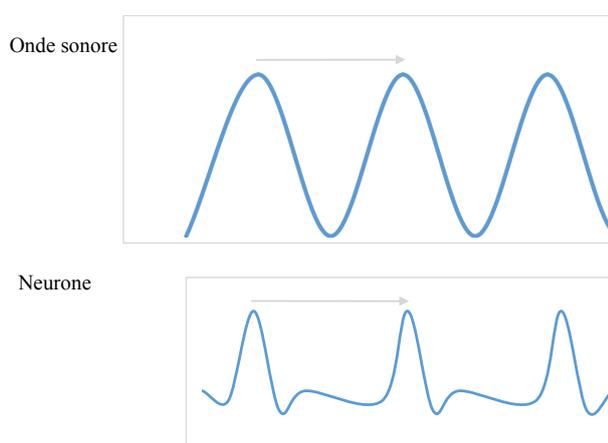


Figure 3 : Verrouillage de phase : synchronisation des potentiels d'action avec la période de l'onde sonore entrante.

L'intervalle de temps entre deux potentiels d'action est un multiple entier d'une période : 2, 3, 4 périodes par exemple. Ainsi, nous pouvons déterminer la période du son et donc sa hauteur (car $T_{période} = \frac{1}{F} \rightarrow F = \frac{1}{T}$). Cependant, le nerf auditif ayant besoin d'une période de repos entre

chaque potentiel d'action, ce codage s'effectue parfaitement jusqu'à une fréquence maximale de 1000 Hz pour un son pur. La perception fine des fréquences est dégradée à partir de ce seuil. Au-delà de 5000 Hz, le verrouillage de phase n'est plus possible car les potentiels d'action sont trop proches. Ce sont les *limites du verrouillage de phase*. Les informations spatiales sont alors nécessaires même si elles sont imprécises car les filtres auditifs sont larges à la base de la cochlée et sont donc peu sélectifs pour la fréquence. Un son pur de haute fréquence (à partir de 5000 Hz) est donc préférentiellement codé spatialement (tonotopie cochléaire) car on ne peut pas avoir de potentiels d'actions successifs assez rapides pour coder temporellement une haute fréquence (Hartmann, 2013).

b) Codage de la hauteur d'un son complexe harmonique (cf annexe 3)

La plupart des sons que nous rencontrons sont des *sons complexes*, correspondant à la somme de plusieurs sons purs. Ils ont donc plusieurs composantes de fréquences différentes. Si le son complexe est périodique, les composantes sont dites *harmoniques* et leurs fréquences (f_1 , f_2 , etc.) sont des multiples entiers de la fréquence fondamentale f_0 . Par exemple, un son de fondamentale 150 Hz aura ses 2^{ème} et 3^{ème} harmoniques à 300 et 450 Hz (multiples entiers de 150 Hz). Ce son est dit *harmonique*. Si au contraire le son complexe est apériodique, ses composantes n'ont a priori aucun rapport harmonique car elles ne sont pas toutes multiples d'une même fréquence. Le son est alors dit *inharmonique*.

➤ ***Théorie spatiale***

Après des années de recherche et des études récentes sur des personnes porteuses d'implants cochléaires, nous savons que cette théorie du codage spatial de la hauteur est capable de prendre en compte de nombreux aspects de la perception de la hauteur, mais ne peut expliquer le codage des harmoniques irrésolues dans les sons complexes. En effet, un problème est rencontré lorsque deux sons purs de fréquences proches (irrésolues) sont entendus simultanément. Les fréquences mettent en vibration la même région de la membrane basilaire. Comme la largeur de bande de chaque filtre auditif est d'environ 10 à 15 % de la fréquence centrale du filtre, au-delà de la 8^{ème} harmonique, les harmoniques voisines ne peuvent plus être séparées (elles sont irrésolues) et interagissent (*cf* théorie temporelle du codage des sons complexes) (Roederer, 2008).

Finalement, la théorie spatiale permet de justifier le codage des harmoniques résolues d'un son complexe, les indices spatiaux n'étant pas pertinents pour le traitement des harmoniques irrésolues.

Modèle sous-tendant la théorie spatiale : modèle de reconnaissance de patterns

D'après le modèle de reconnaissance de patterns pour la perception de la hauteur (Shamma & Klein, 2000), dès les premières expériences auditives d'un sujet, peut-être même prénatales, lorsque nous entendons un son nous associons inconsciemment son spectre à un gabarit (un motif spectral) stocké dans notre mémoire. Le gabarit qui correspond le mieux aux harmoniques du son entendu est celui qui donne la sensation de la hauteur. La f_0 serait alors la fréquence la plus basse du gabarit (Hartmann, 2013). Ce modèle se base sur l'idée que la hauteur est donnée par les harmoniques résolues du son complexe.

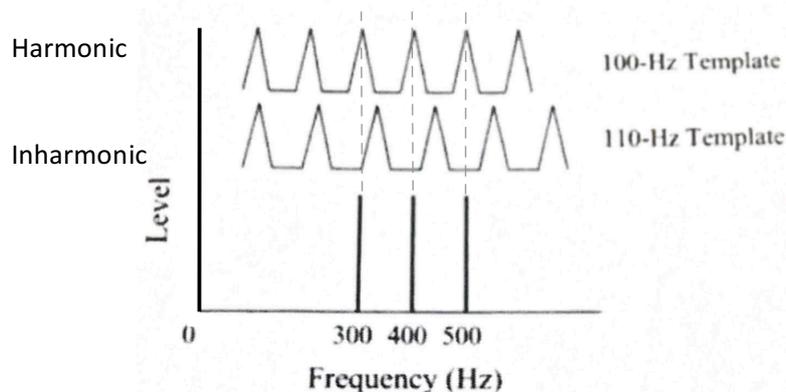


Figure 4 : Modèle de reconnaissance de patterns

(Bonnard, thèse 2013)

Ce modèle a toutefois été critiqué (Bonnard, thèse 2013). Tout d'abord, aucun support neuronal précis de ce type de gabarit n'a été mis en évidence. Ensuite, la question de la génération de ces gabarits au niveau central se pose. Une réponse possible à cette question serait l'apprentissage progressif mais des études chez les bébés démontrent que la sensibilité à l'harmonicité apparaît dès 3 mois (Demany & Armand, 1984; Clarkson & Rogers, 1995). De plus, ce modèle ne considère que les harmoniques résolues dans la détermination de la hauteur. Il ne prend donc pas en compte les harmoniques irrésolues qui peuvent à elles seules permettre une discrimination de hauteurs. Enfin, un son complexe excite la membrane basilaire qui ne montre pas un seul maximum. Il existe donc plusieurs maxima. Le plus grand maximum ne correspond pas nécessairement à la fondamentale du son complexe. Pourtant, la hauteur perçue correspond à la fondamentale du son (Moore, 2004). Si ce modèle à lui seul ne peut pas justifier l'extraction de la hauteur d'un son complexe, c'est qu'il existe probablement un mécanisme différent.

➤ *Théorie temporelle*

La théorie temporelle permet d'expliquer ces phénomènes. Une illustration de l'intérêt de cette théorie est le phénomène de la fondamentale absente. La gamme fréquentielle entière n'est pas indispensable à la perception de la parole en raison d'un phénomène psychophysique appelé

résidu de hauteur décrit par Helmholtz au XIX^e siècle.

Par exemple, le téléphone laisse passer les fréquences entre 300 et 3400 Hz (Fischetti, 2004), donc nous ne devrions pas entendre la voix d'un interlocuteur avec une fondamentale de 200 Hz. C'est pourtant le cas ! Cela indique que la perception de la hauteur des sons harmoniques ne repose pas uniquement sur la présence d'une excitation dans la zone de la membrane basilaire dédiée à la fréquence fondamentale (il n'y a donc pas seulement un codage spatial) (Patterson, 1969; Licklider, 1954; Rauschecker & Shannon, 2002).

Codage des harmoniques résolues des sons complexes :

Schématiquement, chaque harmonique résolue stimule une fibre ou un groupe restreint de fibres nerveuses différentes. D'après la théorie temporelle, les neurones dont la fréquence caractéristique est résolue se verrouillent en phase avec l'harmonique correspondante. Le modèle temporel rendrait compte des bonnes performances dans les basses fréquences (résolues) par un codage précis de la hauteur jusqu'aux limites traditionnellement admises du verrouillage de phase (4000 à 5000 Hz), en complément d'un codage spatial très pertinent pour ces fréquences résolues.

Codage des harmoniques irrésolues des sons complexes :

Dans le cas des harmoniques irrésolues (qui sont donc a priori des hautes fréquences), une information temporelle est également disponible. Le système auditif ne peut plus traiter ces harmoniques séparément, mais des informations issues de leur interaction sont utilisables. En effet, l'interaction entre les harmoniques irrésolues (à partir de la 8^{ème} harmonique) à la sortie du filtre auditif est un indice supplémentaire qui permet d'extraire la hauteur des sons complexes. Cette interaction donne une onde complexe dont l'enveloppe temporelle a la même périodicité que la fondamentale du son complexe. Les potentiels d'action sont déclenchés par les pics les plus saillants de l'onde complexe, les intervalles entre potentiels d'action sont donc égaux à la période de fluctuations de l'enveloppe ($T_{période} = \frac{1}{F}$).

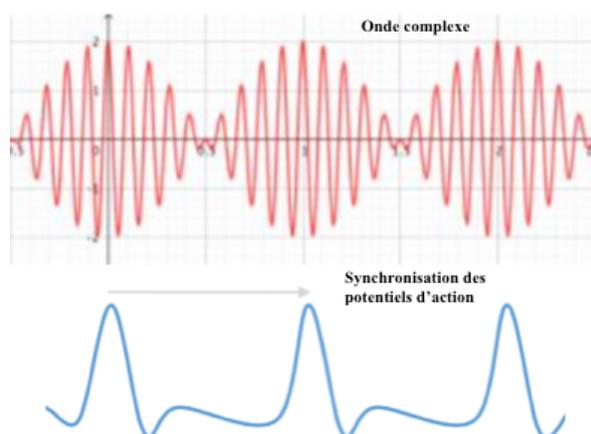


Figure 5 : Verrouillage de phase : synchronisation des potentiels d'action avec la période de l'onde complexe résultant de l'interaction des harmoniques irrésolues.

Ainsi, le neurone dont la fréquence caractéristique se situe dans un filtre auditif très peu sélectif de la fréquence (zone irrésolue) se verrouille en phase avec les fluctuations de l'enveloppe de l'onde complexe synchrones de f_0 . Le système auditif peut donc déduire la hauteur d'un groupe d'harmoniques irrésolues grâce à une information temporelle.

D'après des études de Moore et al. (2003 et 2006), entre la 8^{ème} et la 12^{ème} harmonique, il y a une certaine sensibilité à la structure temporelle fine de l'onde complexe qui implique une représentation précise de la période de l'onde et donc de sa fréquence fondamentale. A partir de la 12^{ème} harmonique, il y a une sensibilité aux variations d'amplitude de l'enveloppe car il y a une moindre sensibilité à la structure temporelle fine (Bonnard, thèse 2013).

Pour résumer, la hauteur d'un son complexe peut être déterminée par des indices temporels et spatiaux pour les harmoniques résolues et majoritairement par des indices temporels pour les harmoniques irrésolues.

Aucune des deux théories n'a été définitivement démontrée comme prédominante dans le codage de la hauteur des sons complexes.

3) Codage de la hauteur par le système binaural

Le système binaural apporte majoritairement des informations spatiales dans le codage de la hauteur. Il a toutefois besoin d'information temporelle pour fonctionner (*cf* déphasage). Les informations temporelles sont présentes mais utilisées pour déterminer la différence de phase interaurale et non la hauteur. Un neurone du système binaural va interpréter le délai temporel qui arrive des deux oreilles pour déterminer une hauteur. Ce neurone est situé à un endroit précis, d'où l'indice essentiellement spatial délivré par le système binaural pour le codage de la hauteur. Finalement, une fois le traitement des informations temporelles effectué, le résultat est représenté sous forme spatiale (tonotopique) (Wallace, Shackleton, & Palmer, 2002).

4) Battements de consonnance imparfaite

Si l'on superpose deux ondes de fréquences f_1 et f_2 légèrement différentes (irrésolues), il y a un décalage de phase continu entre les deux composantes, qui donne naissance perceptivement à une sensation de battement ou de vibration. Cette vibration possède une fréquence constante mais modulée en amplitude (fluctuation de l'enveloppe temporelle). Le son perçu est donc alternativement plus ou moins fort. Ce sont les *battements* : l'interaction entre les composantes au sein d'un même filtre auditif produit des battements détectables par les sujets dès qu'une composante est légèrement déplacée par rapport à l'harmonicité.

Les battements sont donc perçus lorsque deux sons simultanés ont des fréquences proches (non résolues). Ils disparaissent lorsque ces fréquences sont plus éloignées (fréquences résolues qui appartiennent à deux filtres auditifs distincts) ou lorsqu'elles sont égales (Fischetti, 2004).

Moore, Glasberg et Peters (1985) ont déterminé la valeur minimale de déplacement d'une composante d'un son complexe nécessaire pour détecter une inharmonicité : pour les composantes de rang 4 à 6, il suffit de 1,1 % de la fréquence de l'harmonique concernée ; pour les composantes de rangs plus élevés (moins bien résolues), leur interaction au sein d'un même filtre auditif produit des battements. Ils sont détectables dès lors qu'une composante est légèrement déplacée par rapport à l'harmonicité.

Les *battements de consonnance imparfaite* peuvent être présents en cas d'inharmonicité : les conditions qui créent ces battements sont l'absence de bruit, une intensité forte et l'interaction de plusieurs composantes au sein d'un même filtre auditif. Pour étudier la perception de l'harmonicité (par le biais de la détection d'inharmonicité), il est impératif de proposer des conditions optimales pour éviter la création de battements ; ces battements pouvant être un indice d'inharmonicité pour le sujet et ainsi biaiser l'expérience car ils empêchent toute interprétation sur le mécanisme de fusion harmonique.

III / Harmonicité et fusion harmonique

1) Harmonicité

L'harmonicité est la relation physique qu'entretiennent les composantes harmoniques des sons complexes. Elle survient lorsque deux composantes appartiennent à un même schéma spectral (rapports de fréquences entiers et multiples de la fondamentale). Un cas d'amusie (déficit musical de détection de la hauteur des sons) a permis d'établir que le jugement de la consonance ne reposait pas uniquement sur l'absence de perception des battements créés par la présence de composantes inharmoniques dans le même filtre auditif (Cousineau, McDermott, & Peretz, 2012). Depuis quelques années, un second phénomène, qui peut par ailleurs correspondre avec l'absence de battements, fait davantage consensus : la sensibilité à l'harmonicité, reposant sur la fusion harmonique (Langner & Ochse, 2006; McDermott, Lehr, & Oxenham, 2010).

2) Fusion harmonique

a) Définition de la fusion harmonique

La *fusion harmonique* est un phénomène de regroupement perceptif de composantes sonores ayant des relations harmoniques entre elles en une seule unité auditive. Elle est sous-tendue par le phénomène physique d'harmonicité. Chaque son musical est accompagné d'une multitude d'harmoniques. Lorsque l'on écoute deux sons complexes simultanément, la cochlée ne connaît pas d'emblée le nombre réel de sources sonores distinctes. Leurs fréquences, qui stimulent la membrane basilaire, sont alors entremêlées. Le système auditif les analyse et tente de les faire appartenir à un même schéma spectral grâce à l'harmonicité afin de leur associer une unique fréquence fondamentale. Ce phénomène est l'un de ceux qui nous permet de discriminer deux sons simultanés (Moore, 2004).

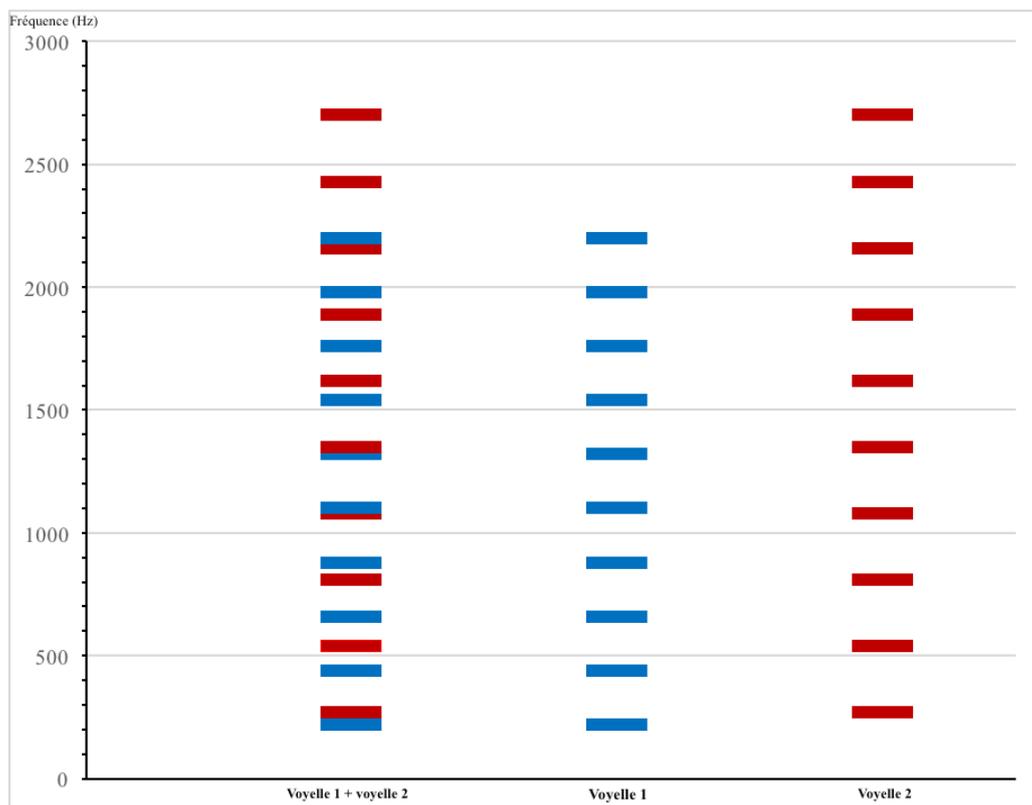


Figure 6 : *Représentation de la fusion harmonique – cas de deux voyelles entendues simultanément*
Les fréquences d'un son complexe (à gauche) sont en fait ici la superposition des premières harmoniques de deux voyelles de fréquence fondamentale 200 Hz (voyelle 1) et 250 Hz (voyelle 2).

b) Etudes réalisées

Des expériences psychoacoustiques suggèrent que la fusion harmonique repose sur des indices temporels (Meddis & Hewitt, 1992, cités par Moore en 2004).

Les études suivantes révèlent que la détection d'inharmonicité est asymétrique pour les stimuli constitués de composantes résolues ; elle est plus facilement détectée lorsque la variation de l'inharmonicité est négative ; en d'autres termes, il est plus facile de détecter qu'un intervalle de fréquences est inharmonique quand il est plus petit que la valeur harmonique correcte que lorsqu'il est plus grand. L'origine de cette asymétrie n'est pas connue à ce jour, mais elle témoigne probablement de phénomènes importants en rapport avec la fusion harmonique, et doit être prise en compte dans les modèles théoriques de la perception des sons complexes.

Une étude de Demany, Semal et Carlyon (1991) utilise des paires de sons purs modulés en fréquence dans un bruit rose (*cf* annexe 2). Chaque son est présenté à 25 dB SL. Le rapport de fréquences entre ces deux sons est constant. Soit il est égal à 2 (une octave) soit il est légèrement inférieur (inharmonicité négative) ou légèrement supérieur (inharmonicité positive) à l'octave. La tâche consiste à trouver la paire où le rapport est différent de 2 en utilisant un indice de dissonance. Les résultats montrent que lorsque l'inharmonicité est négative, les sujets ont de meilleures performances que lorsqu'elle est positive (on appelle cela une asymétrie d'inharmonicité).

En 2013, Bonnard et al. utilisent des sons complexes composés de 4 sons purs simultanés espacés régulièrement en fréquences sur une échelle logarithmique. Pour chaque essai, 4 sons complexes sont présentés. Dans 3 d'entre eux, le rapport de fréquences est égal à 2. Dans le son complexe cible, le rapport de fréquences est différent de 2 (légèrement supérieur ou légèrement inférieur). La tâche consiste à trouver ce son cible inharmonique. Une procédure adaptative est utilisée pour mesurer le seuil de détection de l'inharmonicité. Les résultats montrent qu'au niveau sonore le plus faible, le seuil est meilleur quand l'inharmonicité est négative que lorsqu'elle est positive. Cette asymétrie d'inharmonicité diminue et la performance s'améliore lorsque que le niveau sonore augmente. Quand le niveau sonore est haut, les indices de battements peuvent être utilisés grâce aux interactions des sons dans la cochlée, ce qui pourrait rendre compte de cette diminution de l'asymétrie.

Une étude ultérieure de Bonnard et al. (2017) utilise des sons complexes composés de 3 sons purs simultanés, également espacés régulièrement sur une échelle logarithmique. Les sons sont présentés à 5 ou 15 dB SL dans un bruit particulier appelé *Threshold Equalizing Noise* ou TEN (Moore et al., 2000) (pour en savoir plus sur ce type de bruit *cf* annexe 2). A chaque essai, 4 sons complexes sont présentés. Dans 3 d'entre eux, le rapport de fréquences est égal à 2. Dans le son cible, le rapport de fréquences est différent de 2. La tâche consiste à trouver le son complexe cible

inharmonique. Les principaux seuils des 12 sujets normo-entendants sont très peu affectés par le niveau de sensation (SL) des sons, mais sont significativement meilleurs quand l'inharmonicité est négative (rapport de fréquences inférieur à 2) que lorsqu'il est positif. Cette asymétrie n'était pas observée dans un groupe de 8 sujets atteints de surdité neurosensorielle d'origine cochléaire.

Borchert, Micheyl et Oxenham (2011) utilisent des stimuli constitués de 2 groupes d'harmoniques simultanés, dans des régions spectrales séparées (0-700 Hz et 1150-3500 Hz). Ces deux groupes ont soit la même fondamentale et le stimulus est harmonique ($200 \text{ Hz} \pm 3$ demi-tons), soit des fondamentales différentes et le stimulus est inharmonique. Les composantes de chaque groupe sont partiellement résolues. Dans chaque essai, 2 stimuli sont présentés, l'un est harmonique, l'autre ne l'est pas. La tâche consiste à trouver le stimulus inharmonique. Chez 28 sujets normo-entendants, les stimuli inharmoniques sont légèrement mieux détectés lorsque l'inharmonicité est négative que lorsqu'elle est positive ($p = 0.002$).

Une étude de Bonnard et al. (2016) utilise des sons complexes composés de 2 sons purs simultanés, également espacés régulièrement sur une échelle logarithmique. Les sons sont présentés à 15 dB SL dans du TEN. Dans chaque essai, 2 sons complexes sont présentés, avec des rapports de fréquences différents. La tâche consiste à choisir le son complexe perceptivement le plus consonant, sur la base d'un jugement subjectif. Toutes les paires possibles parmi une gamme de rapports de fréquences autour d'une octave sont testées. Le taux de choix est ensuite calculé pour chaque rapport de fréquence. Les résultats montrent que chez 20 sujets normo-entendants, les sons complexes sont perçus comme plus consonants lorsque l'inharmonicité est positive (supérieure à une octave) que lorsqu'elle est négative.

Pour des stimuli constitués de sons purs résolus, la fusion harmonique et la détection d'inharmonicité peuvent se baser sur des indices temporels ou spatiaux.

Les travaux de Carlyon, Demany et Semal (1992) ont montré que les auditeurs normo-entendants sont capables de détecter un décalage entre les fréquences fondamentales de deux groupes d'harmoniques simultanés, un groupe de basses fréquences constitué d'harmoniques largement résolues et un groupe de hautes fréquences contenant seulement des harmoniques irrésolues. Cela suggère que la détection d'inharmonicité peut être basée sur des indices purement temporels, du moins pour les harmoniques irrésolues.

Ces différents travaux ne permettent pas de conclure quel mécanisme (spatial ou temporel) est prédominant dans la fusion des harmoniques.

3) Détection d'inharmonicité

La valeur minimale de déplacement d'une composante résolue d'un son complexe nécessaire pour détecter une inharmonicité (seuil de détection d'inharmonicité) est de 1,1 % de la fréquence de l'harmonique correcte. Le seuil de détection d'une inharmonicité correspond à la sensibilité absolue à l'harmonicité. Si une harmonique irrésolue est décalée par rapport à l'harmonicité, cela donne naissance à des battements de consonnance imparfaite qui nuisent à l'interprétation de la fusion harmonique.

La hauteur du son complexe varie proportionnellement avec le décalage d'une composante (jusqu'à 2 à 3 % de sa valeur harmonique). Au-delà de cette valeur, le décalage n'a plus d'importance sur la hauteur du son complexe et l'auditeur entendra deux sons distincts. Ainsi, les phénomènes de fusion harmonique et d'extraction de la hauteur sont distincts mais bien intriqués (Moore, Glasberg, & Peters, 1985).

4) Données neurophysiologiques

Les données neurophysiologiques permettent de vérifier les modèles établis à partir des données psycho-acoustiques.

a) Traitement des sons

La technique d'EEG (électro-encéphalographie) permet d'observer la réponse des différents relais de la voie auditive centrale, même lors de l'écoute passive, et donc de mesurer des seuils de détection et des latences de réponse en excluant des biais tels qu'une attention plus soutenue ou une mémoire de travail plus efficiente. Le tronc cérébral induit des réponses électriques qui ressemblent fortement aux sons ayant évoqué ces réponses, ce qui permet de mieux comprendre comment le système auditif central se représente ces sons (Kolinsky, Peretz, & Grandjean, 2017).

On retrouve une organisation tonotopique des sons graves et aigus au niveau du cortex auditif. Une étude de Patterson et al. (2002) suggère qu'il y a un centre de traitement de la hauteur dans la partie antérolatérale du gyrus de Heschl. A ce niveau, l'activation est antérieure et latérale pour les basses fréquences et postérieure et médiane pour les hautes fréquences. Une même organisation fréquentielle est enregistrée au niveau des aires auditives secondaires. Cette organisation se retrouve au niveau du planum temporale également. Ces organisations sont plus facilement

identifiables au niveau de l'hémisphère droit. Effectivement, la capacité à différencier deux mélodies est effectuée grâce au lobe fronto-temporal droit (Platel, 2010).

b) Détection d'inharmonicité

Le noyau cochléaire est un relais important dans la détection d'inharmonicité (Sinex & Li, 2007) : du nerf jusqu'à ce relais, la réponse neuronale suit principalement les composantes individuelles du son complexe, harmonique ou pas ; après ce relais, la réponse neuronale reflète les fluctuations lentes de l'enveloppe liées à l'interaction des composantes inharmoniques.

Une étude utilise les Potentiels Evoqués Auditifs (PEA) pour révéler la détection des stimuli inharmoniques : l'inharmonicité était reflétée par une onde particulière. Il pourrait donc exister un mécanisme automatique dans le système auditif central qui permet de comparer un stimulus inharmonique aux gabarits internes (Alain, Schuler, & McDonald, 2002; Micheyl & Oxenham, 2010).

c) Fusion harmonique

Aucune donnée à ce jour ne met clairement en évidence un mécanisme neuronal ou un niveau particulier du système auditif en jeu pour la fusion harmonique.

IV / Intelligibilité de la parole

1) Acoustique des voyelles

Tous les phonèmes voisés (voyelles et consonnes sonores) sont des sons complexes périodiques (et donc harmoniques) qui ont une fréquence fondamentale (f_0). Cette fréquence détermine le fondamental laryngé, qui correspond à la fréquence de vibration des cordes vocales lors de la production de ces phonèmes. Elle génère des variations prosodiques qui contribuent à l'identification du sexe, de l'âge, de l'identité du locuteur et de la signification du message prononcé. Le fondamental laryngé se trouve dans un registre grave, soit entre 80 et 250 Hz chez l'homme (moyenne à 110 Hz), entre 150 et 350 Hz chez la femme (moyenne à 220 Hz) et au-delà de 250 Hz chez l'enfant. Les voyelles sont graves (200-4000 Hz) et les consonnes peuvent être plus aiguës et atteindre les 10000 Hz pour le /f/ par exemple. Les voyelles sont composées de *formants*. Ce sont des bandes de fréquences correspondant aux résonances des cavités du conduit vocal. Par exemple, la fréquence des deux premiers formants F_1 et F_2 permet de distinguer

approximativement les différentes voyelles du langage (Calcus, 2017; Kolinsky & Hoonhorst, 2017; Fischetti, 2004; Renard, 2008). La zone de fréquences la plus sollicitée dans l'aire conversationnelle se situe entre 2000 Hz et 4000 Hz, elle contient les formants qui permettent de différencier certains phonèmes (Dumont, 2008).

2) Structure temporelle fine et enveloppe temporelle d'un son

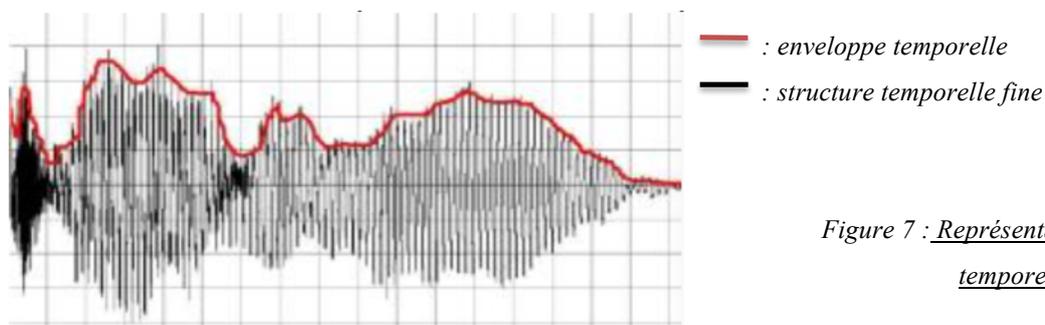


Figure 7 : Représentation de l'information temporelle d'un son complexe

Deux types d'informations temporelles sont extraites par notre système perceptif.

a) *Enveloppe temporelle*

Sur la figure 7, la forme d'onde (ou enveloppe temporelle, en rouge) est la représentation des fluctuations lentes de l'amplitude du signal au cours du temps. Ces modulations déterminent les traits acoustiques d'intensité, de durée et d'attaque. Les modulations de 3-4 Hz sont les plus pertinentes pour la perception de la parole (via une oreille saine ou un implant cochléaire) car elles correspondent au rythme syllabique et contribuent à la segmentation en syllabes.

L'enveloppe temporelle assure aussi la constance de la perception de la parole en cas de modification du contenu fréquentiel (présence de bruit ou perte de la sélectivité fréquentielle dans une surdité). Elle sert à coder la hauteur des composantes irrésolues d'un son complexe.

b) *Structure temporelle fine*

La structure temporelle fine est un paramètre de l'onde sonore qui correspond à des fluctuations d'amplitude rapides du signal au cours du temps. Elle joue un rôle majeur dans la compréhension de la parole dans le bruit (Colin C. , 2017) et sert à la perception de la hauteur des sons complexes en codant les composantes résolues majoritairement. Le codage de la structure temporelle fine dans le nerf auditif est effectué grâce au verrouillage de phase des cycles de l'onde, mais il n'est pas possible au-delà d'une fréquence de 4000 à 5000 Hz.

c) Rôle de ces paramètres dans la compréhension de la parole

Afin d'examiner le rôle de l'enveloppe temporelle et de la structure fine dans la perception de la parole, des chercheurs ont conservé seulement l'enveloppe temporelle globale ou la structure temporelle fine du son dans des tâches de compréhension de la parole. Shannon et al. (1995) ont présenté à des adultes normo-entendants des phrases dont l'enveloppe temporelle était préservée mais dont la structure temporelle fine était altérée et réduite à seulement 4, 3, 2 ou 1 bandes de fréquences. Avec seulement 3 ou 4 bandes de fréquences, l'intelligibilité atteignait pratiquement 100 %. Avec 1 ou 2 bandes de fréquences, l'intelligibilité était altérée mais restait au-dessus du niveau du hasard. Les résultats montrent que l'enveloppe temporelle et la structure temporelle fine sont nécessaires pour une bonne intelligibilité mais aussi que l'intégrité de l'enveloppe temporelle conduit à une très bonne intelligibilité des mots dans le silence malgré la suppression de bandes de fréquences.

3) Fusion harmonique et intelligibilité de la parole

Une audition normale ne garantit pas une compréhension optimale de la parole dans les environnements auditifs quotidiens. Dans la plupart des scènes auditives, plusieurs sources sonores (des personnes qui parlent, de la musique, des verres qui tintent...) émettent des sons concurrents qui s'additionnent pour former un mélange acoustique complexe : d'une certaine façon, toutes les fréquences qui atteignent le tympan « se mélangent ». Il s'agit donc pour l'auditeur de démêler cet ensemble complexe pour attribuer chaque flux auditif à sa source sonore. Lorsque plusieurs personnes parlent en même temps, il peut être difficile de distinguer et de comprendre la parole de chacun. C'est le problème appelé Cocktail Party, identifié par Cherry en 1953. Les auditeurs utilisent en fait plusieurs indices. Naturellement et facilement, ils effectuent une analyse de la scène auditive pour entendre l'un des flux en particulier : isoler un stimulus selon ses caractéristiques acoustiques (fréquence, intensité, timbre), grouper plusieurs objets auditifs sur une échelle de temps plus longue. Plus les objets sonores sont similaires, plus le phénomène de masquage auditif entre en jeu, et plus la tâche est compliquée (Calcus, 2017).

La capacité d'orienter son attention vers certains attributs sensoriels est souvent limitée par la présence des catégories perceptuelles existantes chez l'individu. Par exemple, il est très difficile de percevoir les phonèmes d'une nouvelle langue lorsque le système auditif est spécialisé pour filtrer les sonorités de sa langue maternelle (Goldstone & Hendrickson, 2010). Cette perception catégorielle est la conséquence d'un mécanisme perceptuel : la fusion. Pour des raisons d'efficacité, les systèmes perceptuels cherchent à regrouper les différents attributs sensoriels en

unités plus simples et globales. Par exemple, la couleur contient de nombreux attributs sensoriels perceptibles. Comme ils se présentent naturellement en co-occurrence et qu'il n'est généralement d'aucune utilité de savoir les distinguer, la plupart des gens les regroupent naturellement et, en conséquence, ne parviennent plus à les percevoir distinctement (Garner, 1976). La fusion des attributs sensoriels est un mécanisme essentiel qui permet à l'individu de continuer à bien percevoir son environnement en présence de variations (Fournier, 2017).

Le phénomène de fusion harmonique joue probablement un rôle très important pour ségréguer les différentes sources sonores en milieu bruyant ou multi-locuteur, où, à un instant donné, deux voyelles sont émises simultanément par deux locuteurs distincts. Chacune donnera au niveau de la cochlée de l'auditeur son propre pattern d'excitation en rapport avec sa propre fréquence fondamentale et la série d'harmoniques qui en découle. La fusion harmonique permet de réattribuer correctement les composantes correspondant à chacune des sources car elles ont des relations harmoniques entre elles (Fournier, 2017). La fusion harmonique nous permet donc de différencier deux sources sonores (donc les voix de deux personnes) (Carlyon & Gockel, 2008).

Pour illustrer le phénomène de fusion harmonique, prenons l'exemple d'une étude de Brokx & Nootboom (1982) sur des sujets normo-entendants, présentée par Colin dans sa thèse (2016). Les résultats montrent que lorsque deux personnes parlent en même temps, il est plus facile de suivre la conversation de l'un lorsque les deux personnes ont des fréquences fondamentales laryngées différentes. Cet effet peut s'expliquer par les différences théoriques du codage de la hauteur. Lorsque les fréquences des deux voix sont différentes, les harmoniques résolues de ces deux voix ont des fréquences différentes et excitent des zones différentes de la membrane basilaire (codage spatial). Ceci permet au cerveau de séparer les harmoniques des deux voix et d'attribuer à chaque interlocuteur seulement les composantes ayant des relations harmoniques. Les harmoniques les plus hautes, elles, conduisent à des patterns d'excitation de la membrane basilaire et ces formes d'ondes diffèrent de par leur vitesse de répétition puisque les deux interlocuteurs ont des fréquences fondamentales laryngées différentes ($T = 1 / f$). Le système auditif peut être capable de détecter les différences de vitesse de répétition pour améliorer la séparation des deux voix.

4) Surdit  neurosensorielle et implant cochl aire

Comme nous venons de le voir, chez les auditeurs normo-entendants, l'enveloppe temporelle est essentielle pour une bonne compr hension de la parole dans le silence. Dans le bruit, c'est

l'information issue de la structure temporelle fine qui semble être cruciale (Zeng et al., 2004; Lorenzi et al., 2006; Dincer D'Alessandro et al., 2017).

En cas de lésion endocochléaire, il y a une *baisse de la sélectivité fréquentielle* et une *dégradation de la structure temporelle fine*. Le 1^{er} phénomène est lié à une augmentation de la taille des filtres auditifs notamment dans les basses fréquences (résolues). Des harmoniques résolues chez un auditeur normo-entendant ne le seront plus forcément chez une personne avec une lésion. L'implant cochléaire fonctionne sur le principe suivant : le son est capté par un microphone puis découpé en plusieurs bandes de fréquences (dont le nombre varie en fonction du nombre d'électrodes implantées, entre 12 et 22 électrodes) selon le principe du théorème de Fourier, et le signal de chacune de ces bandes passe ensuite dans un circuit d'extraction de l'enveloppe temporelle. Cette étape peut toutefois s'avérer plus compliquée lorsque le signal de parole est produit à débit rapide, en particulier chez les personnes âgées qui sont très sensibles aux aspects temporels du signal de parole (Colin C. , 2017). Par ailleurs, la structure temporelle fine, nécessaire pour déterminer la hauteur d'un son complexe en codant les harmoniques résolues et certaines harmoniques irrésolues, est également dégradée en cas de lésion endocochléaire. L'implant cochléaire n'en permettant pas une bonne restitution, les informations de hauteur musicale peuvent être altérées et la perception de la parole dans le bruit difficile. Ces difficultés peuvent également résulter de l'altération du mécanisme de fusion harmonique.

Problématique et hypothèse

Problématique :

Sur quels indices repose la fusion harmonique ?

Hypothèse :

Nous posons l'hypothèse que la détection d'inharmonicité nécessite la présence d'indices temporels et que la fusion harmonique repose sur un codage temporel de la hauteur.

Méthodologie générale

Les éléments suivants sont communs aux deux expériences réalisées dans ce mémoire.

1) Sujets

Pour réaliser nos expériences, nous recrutons une population d'adultes musiciens avec une audition normale.

Les sujets sont recrutés sur le campus universitaire. Ils doivent signer un consentement de participation à une étude de recherche expérimentale sur la personne humaine. Ils devront être disponibles pour 11 séances au total d'environ 1h-1h30 chacune.

Tous nos sujets sont étudiants, musiciens, avec une audition normale. Certains sujets de l'expérience 1 sont recrutés également pour l'expérience 2. Nous avons un nombre restreint de sujets par expérience (12 sujets) mais nous privilégions la précision des mesures pour chaque sujet et pour chaque condition expérimentale. Cette méthodologie est communément utilisée dans les études en psycho-acoustique.

2) Conditions expérimentales et matériel

L'expérience se déroule au laboratoire de l'équipe « Perception auditive » de l'Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA, UMR CNRS 5287, Université de Bordeaux), en présence du Dr Laurent Demany et moi-même.

Le matériel est disponible au laboratoire : deux cabines insonorisées, le programme Matlab, un ordinateur, une carte son, un casque, un audiomètre et un multimètre.

Pour chacune des étapes qui vont suivre, le sujet passe individuellement dans une cabine insonorisée avec un casque sur les deux oreilles. Il est face à un ordinateur où il doit choisir le stimulus qu'il juge correct en fonction de la tâche demandée.

3) Etapes de calibrage

a) Audiogramme

Un audiogramme tonal dans le silence et par voie aérienne est réalisé au préalable selon la méthodologie utilisée en pratique clinique pour vérifier que les seuils de chaque sujet sont normaux (compris entre 0 et 20 dB).

b) Détermination des seuils de détection des sons

La deuxième étape de l'expérimentation consiste à mesurer les seuils de détection d'un bruit (TEN) dans le silence. Le bruit peut être composé d'harmoniques résolues (groupe bas) ou irrésolues (groupe haut). Ce programme est adaptatif : à chaque fois que le sujet détecte bien le bruit, le niveau de présentation est abaissé. Inversement, lorsqu'il se trompe cela signifie que le sujet détecte mal le bruit, alors le niveau est augmenté. Le programme Matlab a été conçu pour déterminer le seuil de détection de ce bruit à l'issue de 14 réponses erronées. A chaque réponse, le sujet a un feedback visuel « oui » / « non ». La f_{\min} du TEN (limite fréquentielle inférieure du TEN) est 50 Hz et sa f_{\max} est 6000 Hz.

A l'issue de la 2^{ème} étape, nous avons déterminé les paramètres suivants :

- Paramètre Z = amplitude du TEN à son seuil de détection ;
- Paramètre C = puissance du groupe bas à son seuil de détection dans le silence ;
- Paramètre D = puissance du groupe haut à son seuil de détection dans le silence.

La troisième étape consiste à mesurer les seuils de détection d'un son complexe harmonique composé uniquement d'harmoniques résolues ou d'harmoniques irrésolues, et de f_0 fixe à 141 Hz (moyenne géométrique de l'intervalle 100-200 Hz choisi dans l'expérience principale) dans le silence ou dans le bruit. Ce programme est également adaptatif, avec feedback. Nous utilisons le paramètre Z (amplitude du bruit à son seuil de détection) de chaque sujet calculé dans l'étape précédente. Le niveau de sensation du bruit dans lequel est présenté le signal est de 40 dB SL, soit 40 dB au-dessus du seuil de détection du TEN.

A l'issue de la 3^{ème} étape, nous obtenons les paramètres suivants :

- Paramètre A = $y \times 10^9$ = puissance du groupe bas à son seuil de détection en présence de TEN à 40 dB SL

- Paramètre $B = y \times 10^9$ = puissance du groupe haut à son seuil de détection en présence de TEN à 40 dB SL

4) Analyse des résultats

Pour analyser les données brutes récupérées sur le programme Matlab, nous utilisons Excel pour le calcul des moyennes et écarts-types en pourcentages de bonnes réponses et seuils en cents, ainsi que pour la conception des graphiques. Le logiciel NCSS version 9 est utilisé pour réaliser les ANOVA.

Rappels des résultats attendus : nous nous attendons à trouver la même asymétrie d'inharmonicité dans les deux expériences, nous permettant de penser que le mécanisme à l'origine de la fusion harmonique est d'ordre temporel dans le système auditif.

Expérience 1 : détection d'inharmonicité dans un stimulus constitué de deux groupes d'harmoniques résolues et irrésolues

1) Méthode

Une fois les mesures effectuées, nous pouvons commencer la première expérience. La consigne qui s'affiche à l'écran est : « Quel son est le plus dissonant ? ». Il s'agit pour le sujet de choisir parmi deux stimuli celui des deux qui est inharmonique. Nous vérifions la bonne compréhension de la consigne en proposant un feedback visuel lors des essais préliminaires de familiarisation avec l'expérience.

Cette expérience évalue la détection d'une inharmonicité entre un groupe d'harmoniques « basses » (résolues) et un groupe d'harmoniques « hautes » (irrésolues) (cf Figure 8). Dans chaque essai, 2 stimuli sont présentés. Chaque stimulus est un son complexe constitué d'une fondamentale f_0 tirée au hasard entre 100 et 200 Hz (les f_0 s des 2 groupes varient intra-essai), des harmoniques résolues jusqu'à 500 Hz (rang 4 maximum) et des harmoniques non résolues comprises entre 2500 et 5000 Hz (rangs 18 à 35). Dans l'un d'eux, la f_0 du groupe d'harmoniques irrésolues (groupe haut) est la même que celle du groupe d'harmoniques résolues (groupe bas). Dans l'autre stimulus, la f_0 irrésolue diffère de la f_0 résolue de plus ou moins 300 ou 600 cents (cf annexe 2) (l'intervalle et le signe de la différence sont fixés dans des blocs de 60 essais).

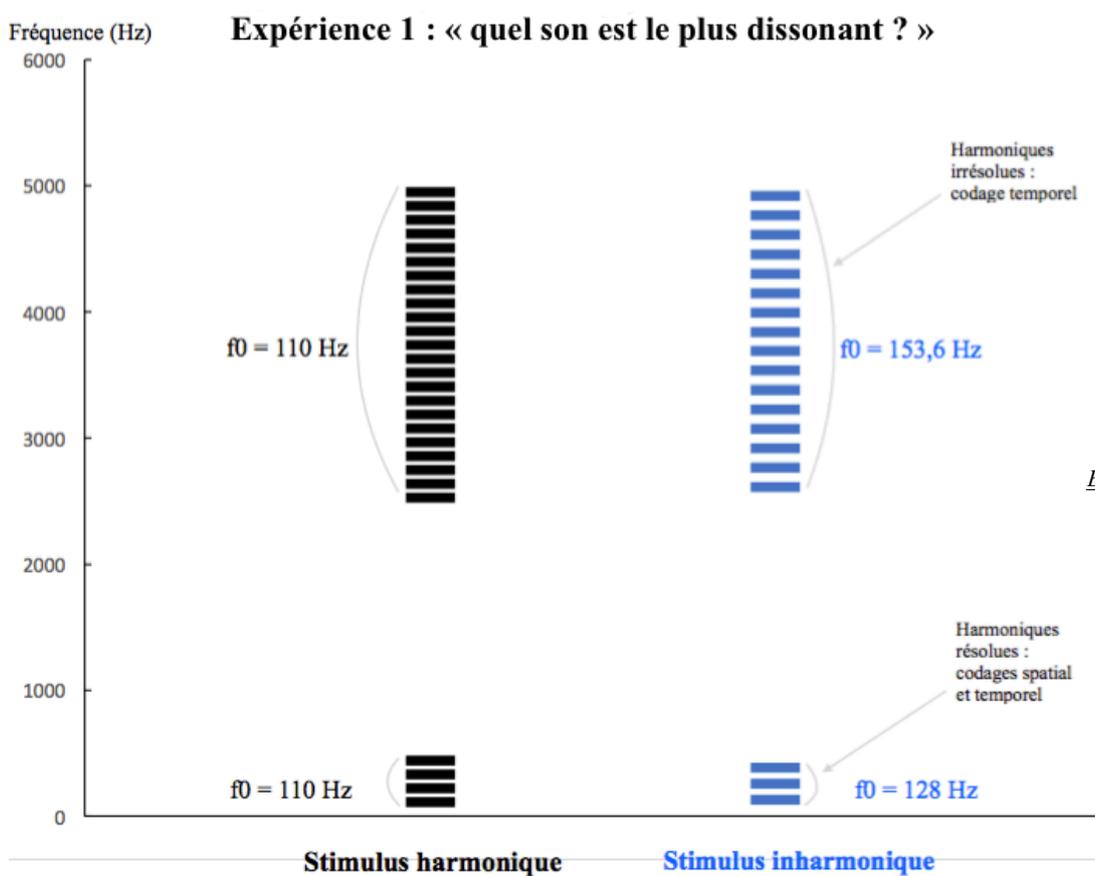


Figure 8 :
Exemple de stimuli –
expérience 1.

La tâche consiste à identifier le stimulus inharmonique, prétendument « le plus dissonant ». Quand le son est harmonique, les 2 groupes d'harmoniques qui le composent ont la même fondamentale et le sujet n'entend qu'un seul son. Lorsque le son est inharmonique, alors les fondamentales des deux groupes d'harmoniques non résolues et résolues sont décalées d'une demi-octave (soit 600 cents) ou d'un quart d'octave (soit 300 cents) vers le haut ou vers le bas (inharmonicité négative ou inharmonicité positive).

L'inharmonicité à détecter est de +300 cents, -300 cents, +600 cents, ou -600 cents (4 conditions). Les harmoniques basses sont résolues (rang maximum 4). Les harmoniques hautes sont irrésolues (rang harmonique minimum 13). Les 2 groupes d'harmoniques d'un stimulus sont présentés au même niveau de sensation : 20, 40 ou 60 dB SL (3 conditions). La durée du stimulus est de 400 ms. Les 12 sujets effectuent 240 essais dans chacune des 12 conditions expérimentales (3 niveaux de sensation sonore x 2 tailles d'inharmonicité x 2 signes de variation de l'inharmonicité).

Le signal présenté à 20 dB SL est présenté dans du TEN à 40 dB SL pour masquer éventuellement les produits de distorsion. La condition à 20 dB SL de cette expérience est parfaite en acoustique (condition plus difficile, intensité faible, dans du bruit pour limiter au maximum les biais). Les deux conditions à 40 dB SL et 60 dB SL ressemblent aux conditions de l'EEG (FFR, Frequency Following Response, mode de recueil d'EEG pour savoir quelles structures répondent à la fréquence fondamentale). Ces conditions sont moins idéales en acoustique mais elles sont plus réalistes. Ces conditions nous permettent d'observer si l'asymétrie persiste encore dans ces conditions de FFR.

Au total dans l'expérience 1, il y a 12 conditions contenant chacune 60 essais. Chaque sujet effectue 4 sessions pour chaque condition donc il y a 240 essais par condition et par sujet.

- | | |
|---|--|
| - Condition 1 : 20 dB SL dans TEN à 40 dB SL ; - 300 cents | - Condition 6 : 40 dB SL silence ; + 300 cents |
| - Condition 2 : 20 dB SL dans TEN à 40 dB SL ; + 300 cents | - Condition 7 : 40 dB SL silence ; - 600 cents |
| - Condition 3 : 20 dB SL dans TEN à 40 dB SL ; - 600 cents | - Condition 8 : 40 dB SL silence ; + 600 cents |
| - Condition 4 : 20 dB SL dans TEN à 40 dB SL ; + 600 cents | - Condition 9 : 60 dB SL silence ; - 300 cents |
| - Condition 5 : 40 dB SL silence ; - 300 cents | - Condition 10 : 60 dB SL silence ; + 300 cents |

- **Condition 11** : 60 dB SL silence ; -
600 cents

- **Condition 12** : 60 dB SL silence ; +
600 cents

Les résultats sont affichés à la fin de chaque condition sous la forme d'un score sur 60, auquel le sujet n'a pas accès.

2) Résultats

a) Résultats (moyennes de tous les sujets)

Expérience 1 : 20 dB dans le bruit

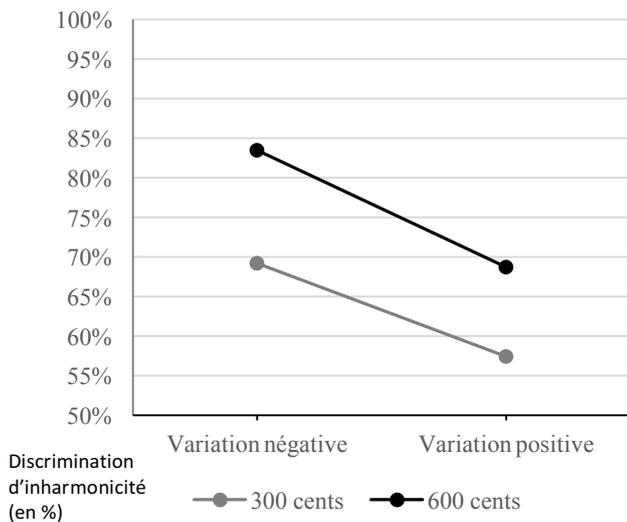


Figure 9

Expérience 1 : 40 dB dans le silence

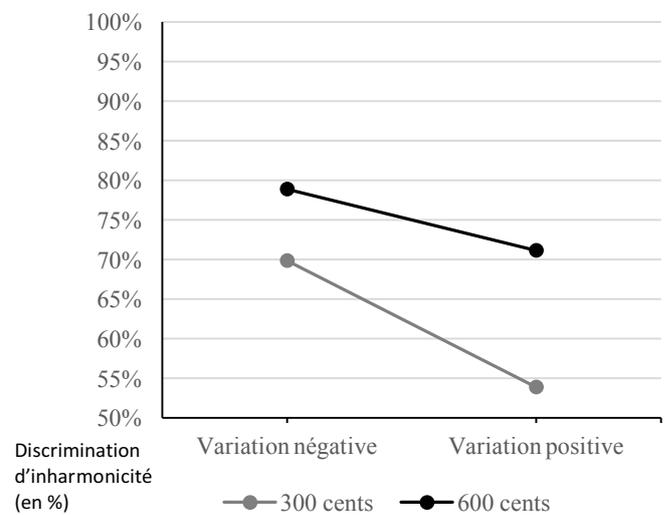


Figure 10

Expérience 1 : 60 dB dans le silence

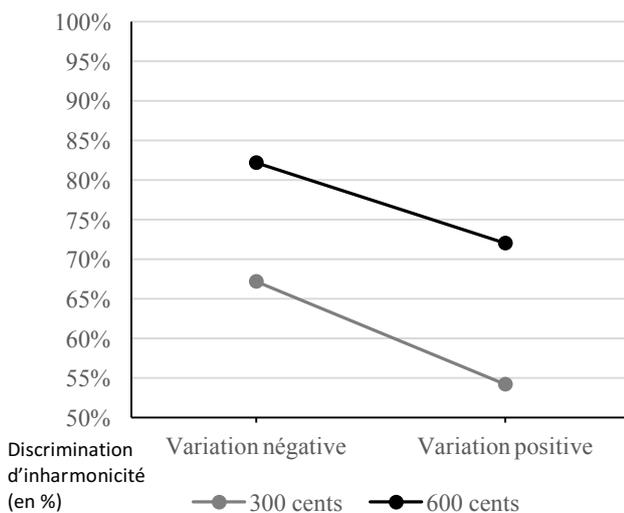


Figure 11

En ordonnée, l'origine correspond au niveau du hasard.

Une ANOVA révèle des effets significatifs de la taille de l'inharmonicité ($p < 0.001$) et de son signe ($p = 0.001$), mais pas du niveau de sensation ($p = 0.49$). L'interaction des 3 conditions

expérimentales est aussi significative ($p = 0.003$). Nous relevons un effet significatif du signe de l'inharmonicité : les variations négative et positive ne sont pas également détectables. La variation négative améliore la performance. Il y a alors à une asymétrie de détection d'inharmonicité (en faveur de l'inharmonicité négative). Nous relevons également un effet significatif de la taille de l'inharmonicité : les sujets sont meilleurs à 600 cents qu'à 300.

b) Résultats individuels

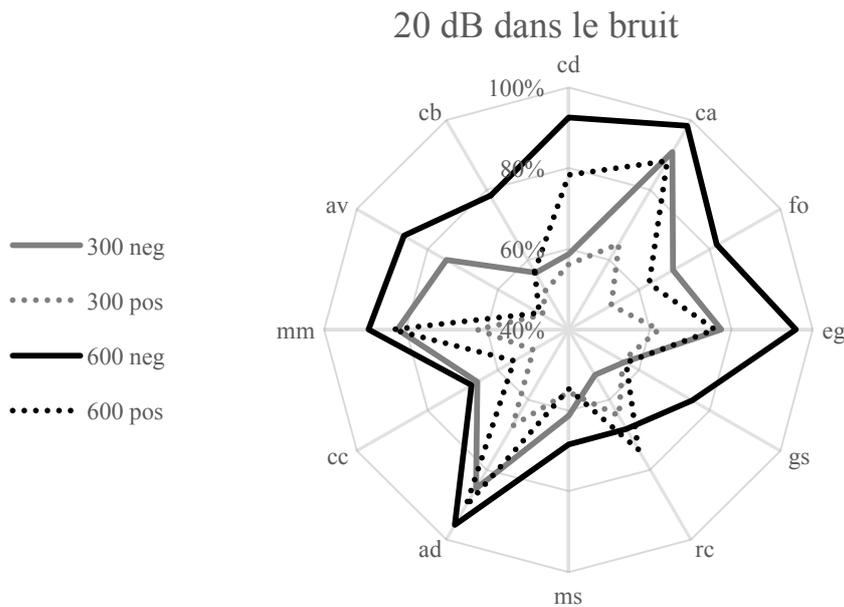


Figure 12

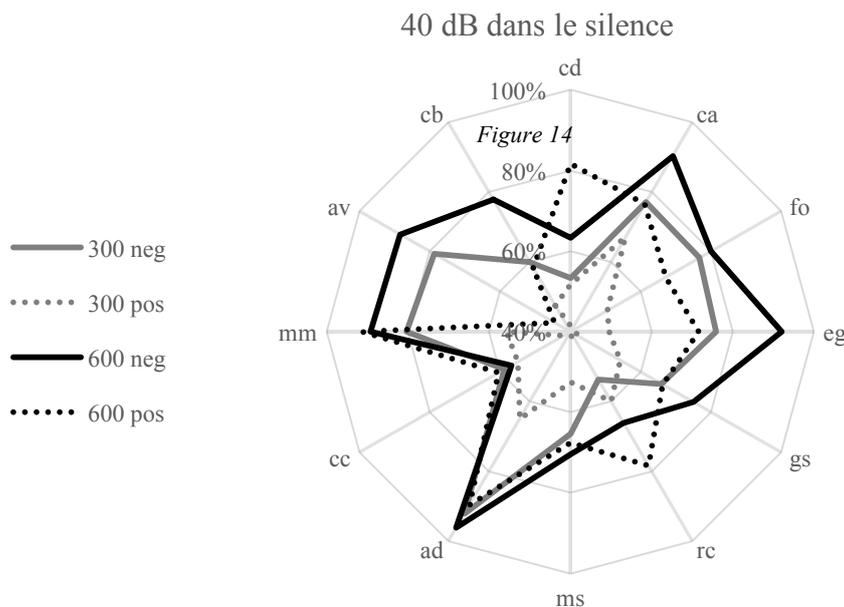


Figure 13

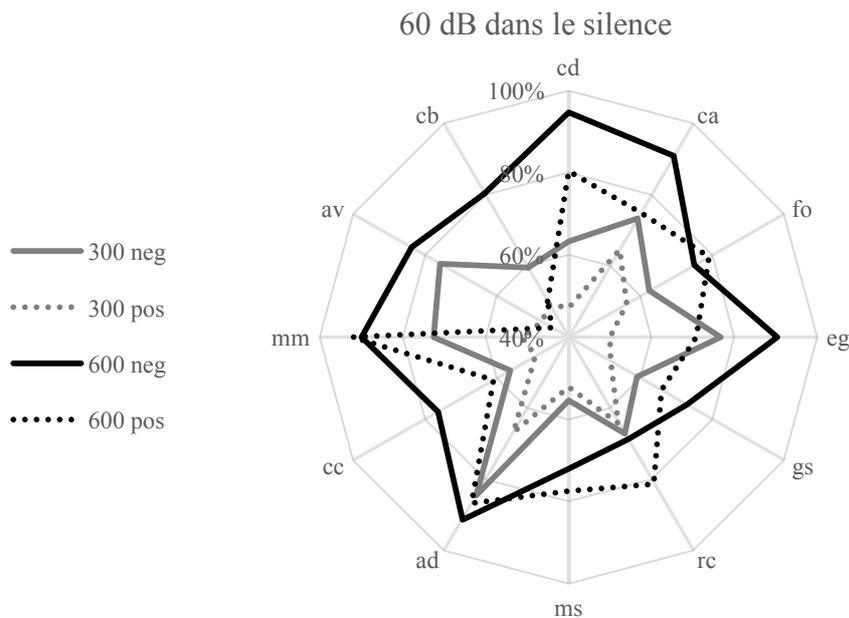


Figure 14

11 des 12 sujets montrent clairement un effet du signe de l'inharmonicité dans la direction attendue (négative) dans les trois conditions. Le sujet « rc » montre un petit effet dans la direction opposée (il détecte mieux l'inharmonicité lorsque la variation est positive).

3) Discussion

L'expérience 1 qui teste la détection d'inharmonicité dans un stimulus constitué de sons purs résolus et irrésolus révèle une asymétrie : la détection d'inharmonicité est plus facile lorsque la variation d'inharmonicité est négative. Ces résultats sont concordants avec une étude de Bonnard et al. (2017). Dans notre expérience, le niveau de sensation sonore n'a pas d'effet significatif sur la détection d'inharmonicité.

Le sujet « rc » qui montre des résultats allant à l'encontre de la tendance générale dans l'expérience 1 a pu être gêné par un manque d'attention, ou bien révèle un fonctionnement auditif qui ne va pas dans le sens de notre hypothèse.

Plusieurs études citées par Moore (2004) expriment l'idée que les changements de perception de hauteur en fonction de l'intensité s'accordent avec la théorie spatiale du codage de la hauteur. En revanche, bien que la théorie temporelle indique que les neurones se synchronisent avec la forme d'onde entrante, elle dit aussi que ces intervalles de temps entre les potentiels d'action doivent être mesurés dans le système auditif. Il est possible que le mécanisme en jeu dans

cette analyse temporelle soit affecté par les neurones actifs et par le nombre de potentiels d'action dans ces neurones, qui dépendent eux-mêmes du niveau d'intensité.

Il est possible que la détection d'inharmonicité dans cette expérience soit sous-tendue par un mécanisme de codage spatial de la hauteur. Demany (1985) a montré que l'inharmonicité est perçue chez les bébés de 3 mois. Ce résultat irait en faveur d'une innéité de la perception de l'harmonicité et donc de la théorie temporelle du codage de la hauteur dans le système auditif, car la théorie spatiale repose sur l'apprentissage de gabarits. En revanche, nous pouvons aussi envisager que des patterns aient été appris in utero jusqu'à 3 mois. Cela pourrait aussi expliquer la réussite dans la détection d'inharmonicité de cette expérience, et cela irait en faveur d'une origine spatiale du codage de la hauteur, en admettant que ces gabarits ne soient pas symétriques, c'est-à-dire que les bébés détectent plus facilement une inharmonicité négative que positive.

Expérience 2 : détection d'inharmonicité dans un stimulus constitué de sons purs résolus dont le premier harmonique peut être déphasé interaurealement

1) Méthode

Cette expérience mesure la détection d'inharmonicité d'un stimulus constitué d'un groupe de trois sons purs simultanés : un son pur H1 situé à un intervalle proche d'une octave d'un deuxième son pur H2, qui est lui-même à exactement une octave du troisième son pur (H4).

Expérience 2 : « Quel son est le plus dissonant ? »

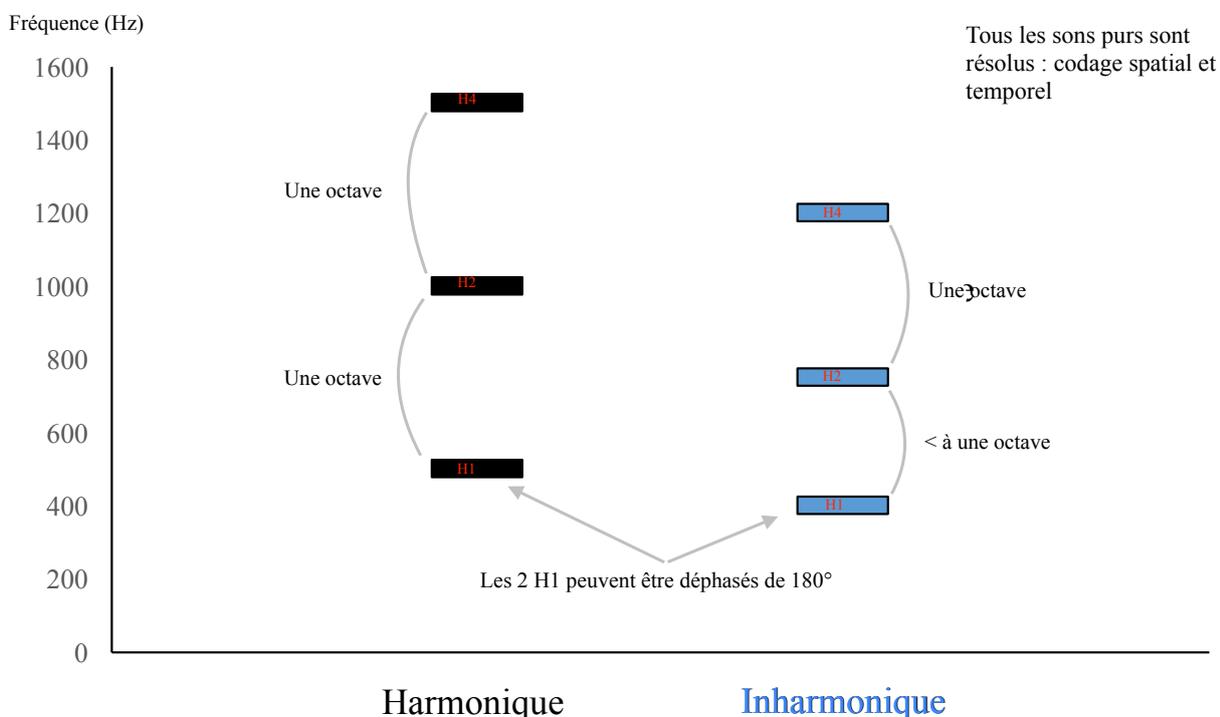


Figure 15 : Exemple de stimuli - expérience 2.

Le son complexe formé est plongé dans du TEN induisant le phénomène de BMLD. On mesure avec une procédure adaptative la variation d'inharmonicité (soit positive soit négative) tout juste détectable. Le TEN est homophasique, de même que H2 et H4. La limite fréquentielle inférieure de la 1^{ère} harmonique est 300 Hz, sa limite supérieure est 600 Hz. H1 a un déphasage de 0° (homophasique) ou 180° (π radians, antiphasique). Le seuil en homophasique et en antiphasique est mesuré pour chaque sujet préalablement. Les trois composantes de chaque paire de stimuli sont présentées soit à 8 soit à 18 dB SL. Par exemple pour un sujet, si le seuil de détection d'un son pur

homophasique est de 25 dB SPL, avec un BMLD de 12 dB, le seuil de détection d'un son pur antiphasique sera de 13 dB SPL. A partir de ces seuils, on ajoute soit 8 dB SL soit 18 dB SL.

-Dans la condition à 8 dB SL, un son homophasique sera perçu à 33 dB SPL et un son antiphasique à 21 dB SPL. Comme 21 dB SPL < 25 dB SPL, la condition « 8 dB SL + son antiphasique » nous permet de présenter des sons qui ne sont détectables que par le système binaural.

-Dans la condition à 18 dB SL, un son homophasique sera perçu à 43 dB SPL et un son antiphasique à 31 dB SPL. Comme 31 dB SPL > 25 dB SPL, la condition « 18 dB SL + son antiphasique » nous permet de présenter des sons détectés monoralement et binauralement.

Donc, dans la condition « 8 dB SL + son antiphasique », avec un BMLD de 12 dB, le son est détectable uniquement par le système binaural, et dans la condition « 18 dB SL + son antiphasique », le son est détectable par le par le système monaural. Les résultats pourraient être différents dans ces deux conditions si notre hypothèse est vérifiée. C'est pourquoi les sujets recrutés pour cette 2^{ème} expérience doivent avoir un BMLD de 12 dB minimum.

Il s'agit d'une procédure adaptative sans feedback, on procède par essais-erreurs. Il faut attendre 14 changements pour obtenir le seuil (la plus petite inharmonicité détectable, en cents).

Il y a 8 conditions, 3 paramètres dans chaque condition (déphasage 0 ou 180°, SL 8 ou 18, variation d'inharmonicité -1 ou +1). Chaque sujet effectue 4 séances au cours desquelles chaque condition est effectuée 2 fois. Au total, chaque condition est réalisée 8 fois par sujet, et comme il y a 12 sujets, chaque condition est réalisée 96 fois (12x8).

- **Condition 1 :**
 - Déphasage 0°
 - 8 dB SL
 - -1
- **Condition 2 :**
 - Déphasage 0°
 - 8 dB SL
 - +1
- **Condition 3 :**
 - Déphasage 0°
 - 18 dB SL
 - -1
- **Condition 4 :**
 - Déphasage 0°
 - 18 dB SL
 - +1
- **Condition 5 :**
 - Déphasage 180°
 - 8 dB SL
 - -1
- **Condition 6 :**
 - Déphasage 180°
 - 8 dB SL
 - +1
- **Condition 7 :**
 - Déphasage 180°
 - 18 dB SL
 - -1
- **Condition 8 :**
 - Déphasage 180°
 - 18 dB SL
 - +1

2) Résultats

a) Résultats (moyennes de tous les sujets)

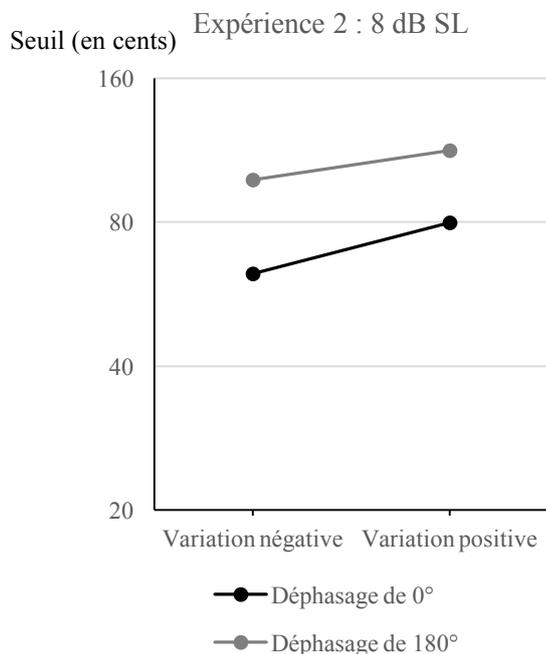


Figure 16

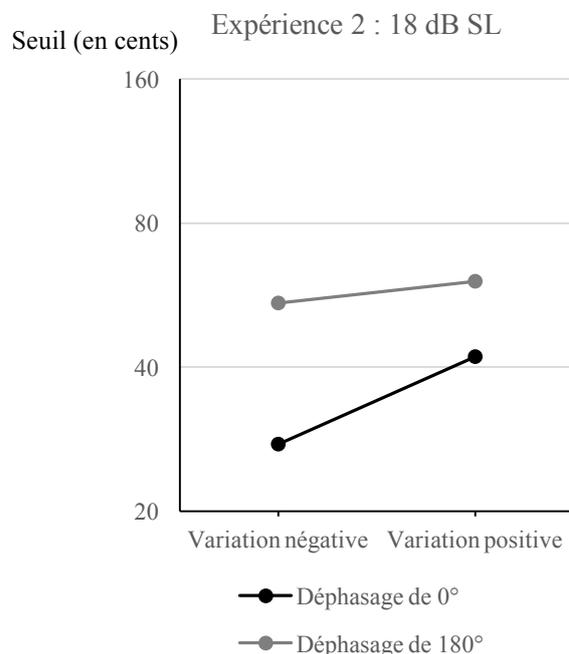


Figure 17

Plus le seuil est bas, plus l'intervalle nécessaire pour détecter une inharmonicité est petit et meilleur est le sujet dans cette tâche.

Dans les conditions à 8 dB SL et 18 dB SL, les sujets détectent mieux l'inharmonicité lorsque la variation est négative et dans la condition sans déphasage interaural.

Une ANOVA révèle des effets significatifs du niveau de sensation ($p < 0.001$), du déphasage ($p < 0.001$) et du signe de l'inharmonicité ($p < 0.001$). L'interaction des conditions expérimentales « niveau de sensation » et « signe de l'inharmonicité » est aussi significative ($p = 0.008$). Nous relevons un effet significatif du signe de l'inharmonicité : les variations négative et positive ne sont pas également détectables. La variation négative améliore la performance. Il y a alors à une asymétrie de détection d'inharmonicité en faveur de l'inharmonicité négative, sauf dans la condition « 8 dB SL + déphasage de 180° ». Nous relevons un effet significatif du déphasage : l'absence de déphasage interaural améliore la performance. Nous relevons également un effet significatif du niveau de sensation : la condition 18 dB SL améliore les seuils.

b) Résultats individuels

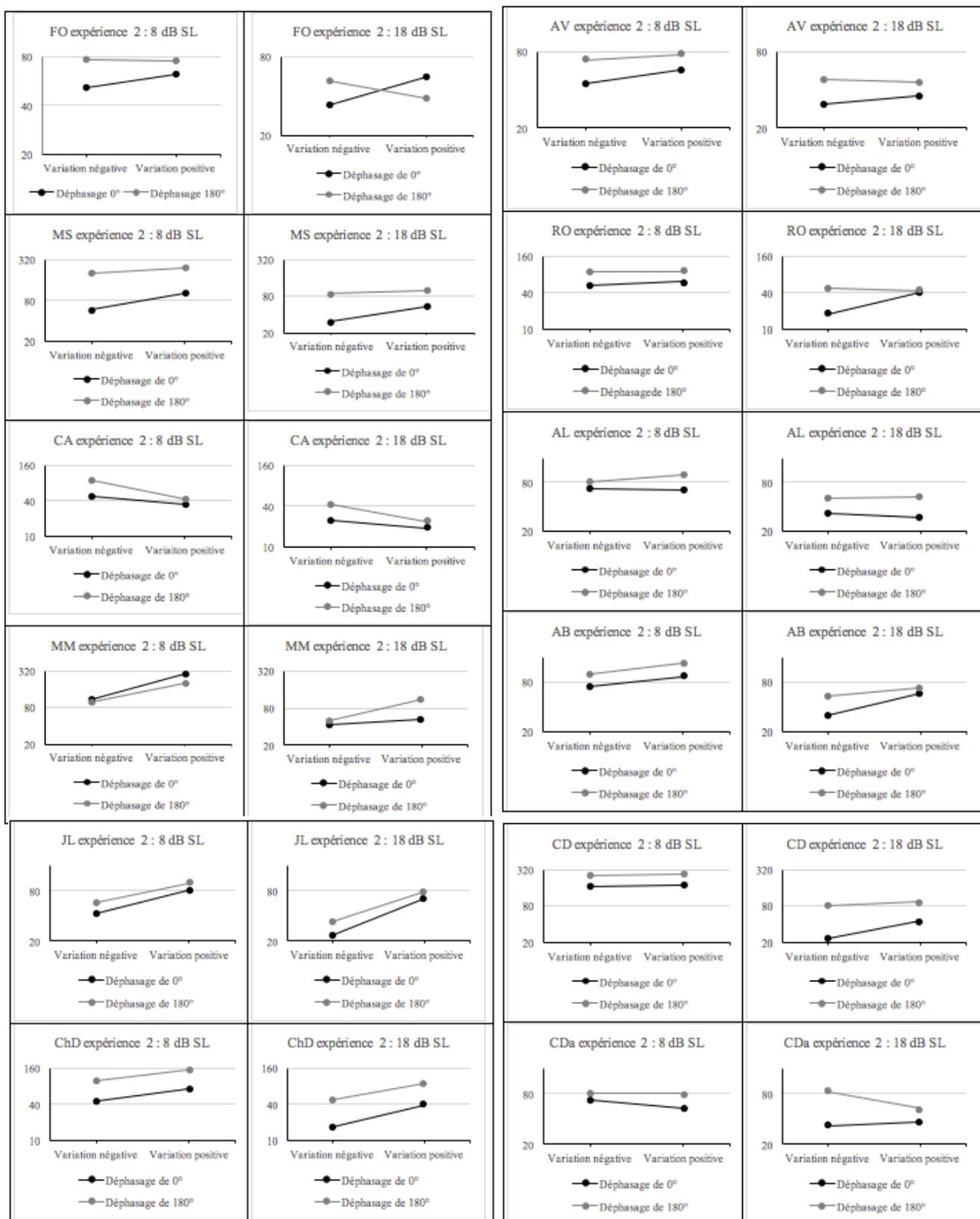


Figure 18

10 des 12 sujets ont des seuils plus bas dans la condition de variation d'inharmonicité négative (sauf les sujets « cda » et « ca »). Tous les sujets ont des seuils plus bas dans la condition

à 18 dB SL. 11 des 12 sujets ont plus de facilité à détecter une inharmonicité en l'absence de déphasage. En effet, les seuils sont plus bas dans les conditions de déphasage (sauf le sujet « fo »).

3) Discussion

Dans cette expérience, pour valider notre hypothèse du codage temporel de la hauteur, le résultat attendu était que les seules conditions où l'asymétrie n'est pas retrouvée soient celles où le niveau de sensation est bas (8 dB SL) et l'écoute antiphasique (déphasage de 180°) (conditions 5 et 6) car seuls les indices spatiaux seraient en jeu dans ces conditions.

Les performances sont moins bonnes dans la condition à 8 dB SL qu'à 18 dB SL et ceci est probablement lié au fait suivant (Bonnard et al., 2017) : avec une faible intensité sonore, l'encodage périphérique des harmoniques est dégradé. Cela se voit car le seuil de détection d'inharmonicité est plus grand. Avec un fort niveau d'intensité sonore, la détection d'inharmonicité va être limitée par les mécanismes centraux (notre capacité).

Les seuils de détection d'inharmonicité sont plus bas (meilleurs) lorsque la variation d'inharmonicité est négative. Seuls les sujets « cda » et « ca » sont meilleurs lorsque la variation d'inharmonicité est positive. Cela pourrait être expliqué si l'on connaissait l'origine de cette asymétrie. Le déphasage rend plus difficile la détection d'inharmonicité en faisant intervenir le système binaural. Ceci est concordant avec notre hypothèse car la détection d'inharmonicité dans les conditions de déphasage fait intervenir le système binaural qui dispose essentiellement d'indices spatiaux. En revanche, l'asymétrie constatée sur les courbes dans la condition « 8 dB SL + déphasage de 180° » entre variation positive et variation négative de l'inharmonicité n'est pas significative. L'asymétrie de l'expérience 1 n'est donc pas retrouvée dans toutes les conditions de l'expérience 2. Les seuils sont meilleurs lorsque les sons sont homophasiques mais l'inharmonicité reste tout de même détectée lorsque seul le système binaural peut apporter une information.

Nous avons mené cette expérience en prenant garde à la perception des battements de consonance imparfaite, qui peuvent être un indice d'inharmonicité biaisant nos résultats. En effet, les conditions expérimentales ont été particulièrement étudiées dans cette optique : stimuli de faible niveau sonore (entre 8 et 18 dB SL) présentés dans du TEN. Les sujets utilisaient donc l'harmonicité comme repère.

Au seuil de détection d'un son pur dans le bruit, il faut moins d'intensité si le son est antiphasique que s'il est homophasique. Dans l'expérience 2, nous avons pensé travailler à niveau

de sensation sonore constant pour les conditions homophasique et antiphasique, soit à 0 dB SL. Ce point est discutable. Si le bruit est beaucoup plus faible que le son (rapport signal/bruit de 60 dB), la sensation sonore va dépendre seulement du niveau physique du son et le bruit n'aura plus beaucoup d'importance. Donc le son antiphasique va paraître plus faible puisque son niveau physique va paraître plus faible. Par exemple, 60 dB SL en homophasique vont être perçus plus forts que 60 dB SL en antiphasique. Ce phénomène-là se fait sentir dans l'expérience 2. Nous pouvons alors penser que les mauvaises performances de détection d'inharmonicité dans les conditions en antiphasique sont dues à une moins bonne perception du stimulus.

Dans la condition « 8 dB SL + antiphasique », le son n'est pas détectable par le système monaural donc il n'y a pas d'information directement exploitable en périphérie. L'information n'est que dans le système central, où le codage temporel est dégradé par rapport à la périphérie. La pente de la condition « 8 dB SL + antiphasique » n'est pas significativement la même que pour « 8 dB SL + homophasique », donc il n'y a pas la même asymétrie dans cette condition que dans les autres. Pour certains sujets, les sons étaient à peine détectables par le système monaural. Dans la condition « 18 dB SL + antiphasique », cela correspondait à 4 ou 5 dB SL pour un sujet qui a un seuil homophasique bas et un BMLD fort. Ainsi, pour certains sujets, dans la condition « 18 dB SL + antiphasique », le système binaural pouvait répondre à la place du système monaural. Ceci pourrait constituer un biais dans l'interprétation de cette expérience.

Les sons homophasiques étaient alors objectivement plus forts que les sons antiphasiques, ce qui pourraient expliquer que les seuils soient plus bas (donc meilleurs) dans les conditions homophasiques.

Discussion générale

L'objectif de cette étude était de comprendre le codage de la hauteur des sons complexes et donc le mécanisme de fusion harmonique dans le système auditif. Pour cela nous avons réalisé deux expériences. Si la détection d'inharmonicité est basée essentiellement sur les mêmes mécanismes temporels pour des stimuli constitués uniquement de sons purs résolus et pour des stimuli constitués de sons purs résolus et irrésolus, alors il faut s'attendre à ce que la détection d'inharmonicité montre une asymétrie similaire dans les deux cas.

Les résultats montrent l'implication du mécanisme temporel dans le codage de la hauteur des sons complexes, même si d'autres indices peuvent être utilisés par le système auditif.

Il existe parfois une grande variabilité inter-individuelle dans les résultats. Pour tenter de l'expliquer, il est important de comprendre les mécanismes cognitifs en jeu dans la fusion harmonique.

La fusion harmonique est le phénomène à la source de la perception de la parole. Ce mécanisme complexe du système auditif est sollicité à l'extrême lorsque l'on est dans un environnement bruyant. La fusion harmonique est un phénomène perceptif. Lorsque ce phénomène est dégradé, le sujet rencontre des difficultés de compréhension de la parole, exacerbées dans le bruit. Cependant, il faut voir le lien entre ce phénomène de fusion harmonique et les capacités internes de chacun. Effectivement, deux personnes avec une audition normale et des processus de bas niveau opérationnels (sensibilité à l'harmonicité et fusion harmonique par exemple) peuvent ressentir une gêne pour comprendre la parole dans le bruit ou non. Cette gêne peut être très différente selon l'individu. Les origines de ces difficultés sont bien illustrées par les différents types de *masquage auditif*.

La perception de la parole dans le bruit requiert une combinaison adéquate de processus sensoriels (voies auditives périphériques : cochlée) et de traitements cognitifs centraux (cortex auditif primaire et aires associatives). Dans un environnement bruyant, il arrive parfois que l'on n'entende pas un signal auditif donné. Ce phénomène de masquage s'observe quand la perception du signal d'intérêt est altérée par la présence d'un autre son, appelé *masque*. On distingue essentiellement deux types de masquage, selon que l'interférence induite par le masque apparaît au niveau périphérique ou central.

Masquage dû à l'interférence périphérique :

Un son faible est masqué par une bande de bruit intense si la fréquence du son et les fréquences du bruit tombent dans le même filtre auditif. Masquer signifie que le son faible ne peut pas être entendu à cause de l'intensité du bruit. Si le son faible et la bande de bruit intense sont séparés en fréquence par plus d'un filtre auditif, alors il n'y a plus de masquage du tout (Hartmann, 2013). Pour comprendre l'influence de l'interférence périphérique sur la perception de la parole dans le bruit, nous prendrons l'exemple d'une conversation dans un club de jazz. Une grande partie des difficultés rencontrées dans cette scène auditive complexe vient de la superposition d'énergies contenues dans les signaux auditifs simultanés. En effet, si la cible (parole du locuteur) et le masque de parole (musique, conversations voisines) partagent des composantes fréquentielles, ils risquent de tomber dans le même filtre auditif de la cochlée, ce qui entravera la perception de la cible. On appelle cela un masquage énergétique. Les premières études sur le sujet ont révélé une propriété importante du système auditif : il est sensible aux composantes fréquentielles du masque (i.e., à son contenu spectral). Par exemple, l'intelligibilité d'un signal de parole présenté simultanément avec un bruit blanc modulé en fréquence ou en amplitude est meilleure que lorsque le bruit de fond est continu. Ainsi, les auditeurs semblent bénéficier des fluctuations spectro-temporelles dans le bruit de fond (les « creux », ou vallées du bruit de fond fluctuant) où le rapport signal-bruit est plus favorable. Un « démasquage » de la parole efficace reflète alors la bonne capacité des auditeurs à analyser les caractéristiques du signal de parole au niveau périphérique du système auditif (Calcutt, 2017). Cette capacité est dépendante de chacun mais n'est pas contrôlable.

Masquage dû à l'interférence centrale :

Quelle est l'influence de l'interférence centrale sur la perception de la parole dans le bruit ? Plusieurs études se sont intéressées à la notion du « cocktail-party » formulée par Cherry (1953) comme suit : « Comment reconnaissons-nous ce qu'une personne dit quand d'autres parlent en même temps ? ». L'interférence centrale peut se résumer ainsi : la fréquence du signal de parole est distincte de celle du masque (bruit environnant), cependant l'auditeur n'arrive pas à distinguer les deux. On parle alors de masquage informationnel. Plus l'auditeur a d'indices de ségrégation (localisation spatiale, nombre de sons interférents, taille de la zone protégée ou zone de fréquences qui ne concerne que la cible) et une disponibilité cognitive (attention, fatigue, expertise musicale, langue étrangère), plus il réussira à dépasser le masquage et inversement. Habituellement, nous nous attendons à recevoir une information sonore à la fois. Cette information sonore, ou son complexe, se démarque de l'arrière-plan formé par d'autres sons complexes. La sélection d'un son

complexe parmi d'autres impose des contraintes à l'attention, mais l'attention peut également influencer la perception d'un son.

Ce masquage ne peut pas être attribué à la superposition spectrale entre cible et masque. Il est en effet influencé par des facteurs cognitifs de haut niveau. Il résulte une interférence au niveau des voies auditives centrales. L'individu peut donc influencer sa compréhension dans le bruit en agissant sur son attention par exemple.

Finalement, dans une situation de conversations simultanées rencontrée quotidiennement, les performances de compréhension de la parole d'un sujet donné dépendent du masquage énergétique et du masquage informationnel. Des études confirment que la difficulté principale rencontrée par des personnes dont le système périphérique réalise a priori une fusion harmonique efficiente dans les situations de type « cocktail party » est due à une interférence au niveau des voies auditives centrales attribuable au masquage informationnel (Calcus, 2017).

Dans ces deux expériences, l'expérience précoce de la musique, les différences individuelles et le bagage culturel ont pu aussi jouer un rôle important dans ce type de jugement (Moore, 2004).

Perspectives

Les résultats des expériences réalisées suggèrent que la fusion harmonique (et donc le codage de la hauteur des sons complexes dans le système auditif), s'appuie sur des indices temporels, mais pas seulement. D'autres expériences sont en cours pour éclaircir les mécanismes en jeu, qui constituent un enjeu majeur dans le domaine de la psycho-acoustique, permettant d'expliquer un élément fondamental de l'audition et par extension l'origine de la compréhension de la parole. Il est possible d'envisager des applications de ces résultats au niveau de la recherche fondamentale et pour la pratique clinique.

Pour la recherche fondamentale :

Une étude effectuée dans le laboratoire de l'équipe « Perception auditive » de l'INCLIA (UMR CNRS 5287, Université de Bordeaux) est en cours pour préciser les mécanismes en jeu dans le codage de la hauteur et la fusion harmonique : l'expérience consiste à faire entendre aux sujets des stimuli harmoniques dans une illusion de continuité. Cette condition entraîne un codage du son uniquement au niveau central, donc une absence d'information temporelle (limitée à la périphérie). En théorie, comme les sujets sont moins performants pour détecter l'inharmonicité pendant la continuité, la sensibilité à l'harmonicité devrait être réduite quand le son est amputé de l'information temporelle, et l'asymétrie devrait alors disparaître. A ce jour, les résultats non finalisés montrent de la réussite dans la détection d'inharmonicité même en l'absence de l'information temporelle. Ils iraient donc dans le sens de la théorie du codage spatial de la hauteur, de la prédominance de l'information tonotopique sur l'information temporelle. Cependant, l'information temporelle a également un impact dans la détection d'inharmonicité, puisque sans elle, la détection d'inharmonicité est plus difficile.

Une asymétrie d'inharmonicité a été l'indice que nous avons utilisé pour justifier l'existence d'un mécanisme commun dans la détection d'inharmonicité dans ces deux expériences. Cette asymétrie, bon repère pour traduire un fonctionnement du système auditif, reste toutefois inexplicée. Pourquoi la détection d'inharmonicité est-elle plus facile lorsque la variation de l'inharmonicité est négative (inférieure à l'intervalle harmonique) ?

Dans l'expérience 1, nous avons montré une même asymétrie entre les différents niveaux d'intensité sonore (20 dB SL dans le bruit, 40 dB SL dans le silence et 60 dB SL dans le silence).

Au niveau de la recherche fondamentale, ces résultats vont nous permettre de tester des sujets en recueillant une FFR (Frequency Following Response) afin d'approfondir ces nouvelles données. La FFR est un potentiel évoqué généré par un stimulus auditif périodique. Elle reflète l'activité neuronale en réponse à un stimulus auditif. Cette réponse est souvent verrouillée en phase avec les cycles de la forme d'onde du stimulus ou avec l'enveloppe des stimuli périodiques. Cette méthode se sert d'une information temporelle exclusivement. Les deux dernières conditions de l'expérience 1 (40 et 60 dB SL dans le silence) reflètent l'environnement de passation de cet examen et les données sont concordantes avec la condition à 20 dB SL. On peut s'attendre à recueillir la même asymétrie avec un examen objectif tel que la FFR.

Pour la pratique clinique :

L'expérience 2 montre que la détection d'inharmonicité est facilitée en l'absence de déphasage interaural. Nous savons qu'un léger déphasage peut intervenir dans certaines surdités asymétriques. Améliorer la synchronisation des informations binaurales dans les appareils (prothèse ou implant) pourrait avoir un effet bénéfique sur la détection d'inharmonicité donc sur la fusion harmonique et ainsi sur les confusions phonétiques et la compréhension en milieu bruyant. Des systèmes de communication de plus en plus performants permettent maintenant aux aides auditives conventionnelles et aux implants cochléaires d'échanger en temps réel des informations entre l'appareil droit et gauche.

Dans une étude récente de Bonnard et al. (2017), les sujets porteurs d'implants cochléaires, les plus performants dans des tâches de détection d'inharmonicité le sont aussi pour comprendre la parole dans le bruit. Au cours de notre étude, nous avons constaté que les sujets normo-entendants ne montraient pas spécifiquement d'effets d'apprentissage dans ce type de tâches. On pourrait alors envisager une rééducation orthophonique pour les patients porteurs d'un implant cochléaire, basée sur l'entraînement à la détection d'inharmonicité, dans le but d'améliorer leur compréhension de la parole dans le bruit.

Les travaux de Bonnard et al. (2013 et 2017) sur des patients présentant une surdité neurosensorielle sont en faveur d'une dégradation concomitante de la discrimination fréquentielle et de la sensibilité à l'harmonicité. La dégradation de la sensibilité à l'harmonicité et l'altération du mécanisme de fusion harmonique sont la conséquence de l'altération du mécanisme de discrimination fréquentielle. Un entraînement à la discrimination fréquentielle aurait un impact positif sur la compréhension du langage oral, grâce à l'affinement du traitement de la prosodie. Le résultat de plusieurs études est en faveur d'un transfert positif des effets d'un entraînement musical

sur le traitement du langage (Patel & Iversen, 2007). Nous pourrions proposer une rééducation orthophonique pour entraîner la discrimination fréquentielle.

Il existe des logiciels conçus pour la rééducation auditive (Audiolog[®]) qui proposent notamment des entraînements sur des tâches de reconnaissance de bruits avec des propositions en liste fermée, de différenciation de durées, d'intensités ou de hauteurs, ou encore de reconnaissance d'instruments de musique. La création de ce type d'application est envisageable pour entraîner la détection d'inharmonicité et la discrimination fréquentielle. Ces cibles de travail, plus axées sur l'acoustique, pourraient compléter les tâches déjà existantes. Elles contribueraient également à l'amélioration de la compréhension de la parole dans le bruit des personnes atteintes de surdité neurosensorielle ou porteuses d'implants cochléaires.

Conclusion

La compréhension de la parole dans le bruit est une question d'actualité. De nombreuses personnes (normo-entendantes, presbycousiques, ou appareillées) s'accordent à dire que la parole devient plus difficile à comprendre dans un milieu bruyant. Malgré les progrès incontestables des appareils auditifs conventionnels et des implants cochléaires, cette limite reste à ce jour l'une des plaintes principales des patients.

La fusion harmonique est l'un des phénomènes perceptifs acoustiques à l'origine de la compréhension de la parole en milieu multi-source. L'objectif de ce mémoire était de comprendre le fonctionnement de ce phénomène, en déterminant quel mécanisme permet de coder la hauteur des sons complexes dans le système auditif.

Nous avons posé l'hypothèse que la présence d'indices temporels est essentielle dans le codage de la hauteur. Nous avons réalisé deux expériences de détection d'inharmonicité. La première faisait intervenir les deux types de codage communément admis (spatial et temporel) et permettait d'envisager les mécanismes temporels comme responsables de ce codage. La deuxième expérience intégrait le système binaural, qui dispose majoritairement d'informations spatiales, pour tenter d'isoler un mécanisme.

Les résultats montrent que les mécanismes temporels jouent un rôle capital dans la détection d'inharmonicité et donc dans le codage de la hauteur. Toutefois, il s'avère que le système binaural seul est aussi capable de détecter une inharmonicité, même si la tâche est plus difficile. Il semblerait donc que d'autres mécanismes interviennent dans ce type de tâche dans une moindre mesure.

Ce travail sera poursuivi par d'autres expériences en recherche fondamentale, qui viseront à éclaircir le rôle des indices temporels dans le codage de la hauteur des sons complexes et à expliquer l'asymétrie retrouvée dans notre étude. Une perspective clinique s'ouvre également pour la rééducation orthophonique, avec notamment la création d'une application permettant un entraînement à la détection d'inharmonicité et à la discrimination fréquentielle pour améliorer la compréhension de la parole dans le bruit chez les personnes atteintes de surdité neurosensorielle ou porteuses d'un implant cochléaire.

Bibliographie

Alain, C., Schuler, B., & McDonald, K. (2002). Neural activity associated with distinguishing concurrent auditory objects. *J. Acoust. Soc. Am.*, 111(2), 990-995.

ANSI. (1994). *American National Standard Acoustical Terminology*, ANSI S1.1.1994. New York: American National Standards Institute.

Bonnard, D. (2013). *Thèse de médecine : La perception de l'harmonicité des sons en présence ou en l'absence de surdit  neurosensorielle*. Bordeaux.

Bonnard, D., Dauman, R., Semal, C., & Demany, L. (2016). Harmonic fusion and pitch affinity : is there a direct link ? *Hear. Res.*, 333, 247-254.

Bonnard, D., Dauman, R., Semal, C., & Demany, L. (2017). The Effect of Cochlear Damage on the Sensitivity to Harmonicity. *Ear & Hearing*, 38(1), 85-93.

Bonnard, D., Micheyl, C., Semal, C., Dauman, R., & Demany, L. (2013). Auditory discrimination of frequency ratios : the octave singularity. *J. Exp. Psychol. Hum. Percept. Perform.*, 39(3), 788-801.

Borchert, E., Micheyl, C., & Oxenham, A. (2011). Perceptual grouping affects pitch judgments across time and frequency. *J. Exp. Psychol. Hum. Percept. Perform.*, 37(1), 257-269.

Brokx, J.P., & Nootboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, 10(1), 23-26.

Calcus, A. (2017). MOOC Neurosciences : audition, langage et musique.

Carlyon, R., & Gockel, H. (2008). Effects of harmonicity and regularity on the perception of sound sources. in *Auditory perception of sound sources* from Yost W.A., Fay R.R., Popper A.N., 191-213.

Carlyon, R., Demany, L., & Semal, C. (1992). Detection of across-frequency differences in fundamental frequency. *J. Acoust. Soc. Am.*, 91(1), 279-292.

Cherry, E. (1953). Some experiments on the recognition of speech with one and two ears. *J. Acoust. Soc. Am.*, 25, 975-979.

Clarkson, M., & Rogers, E. (1995). Infants require low-frequency energy to hear the pitch of the missing fundamental. *J. Acoust. Soc. Am.*, 98(1), 148-154.

Colin, C. (2017). MOOC Neurosciences : audition, langage et musique.

Colin, D. (2016). *Thèse de neurosciences : Influence de la surdit  neurosensorielle sur la perception de la hauteur tonale*. Lyon.

Cousineau, M., McDermott, J., & Peretz, I. (2012). The basis of musical consonance as revealed by congenital amusia. *Proc. Natl. Acad. Sci. USA.*, 109(48), 19858-19863.

- Danziger, N., & Alamowitch, S. (2016). *Neurologie ECN 9^{ème} édition*. Le référentiel Med-line.
- Demany, L. (1985). Perceptual learning in frequency discrimination. *J. Acoust. Soc. Am.*, 78(3), 1118-1120.
- Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *J. Acoust. Soc. Am.*, 76(1), 57-66.
- Demany, L., Semal, C., & Carlyon, R. (1991). On the perceptual limits of octave harmony and their origin. *J. Acoust. Soc. Am.*, 90(6), 3019-3027.
- Dincer D'Alessandro, H., Ballantyne, D., Boyle, P., De Seta, E., DeVincentiis, M., & Mancini, P. (2017). Temporal Fine Structure Processing, Pitch, and Speech Perception in Adult Cochlear Implant Recipients. *Ear and Hearing*.
- Dumont, A. (2008). *Orthophonie et Surdit   : Communiquer, comprendre, parler*. Elsevier Masson.
- Dupont, M., & Lejeune, B. (2010). *R  ducation de la boucle audio-phonatoire, chez les adultes sourds porteurs d'un implant cochl  aire*. Masson.
- Durlach, N., & Colburn, H. (1978). Binaural phenomena. Dans E. Carterette, & M. Friedman, *Handbook of Perception, 4.*, New York.
- Fischetti, A. (2004). *Initiation    l'acoustique*. Belin Sup.
- Fletcher, H. (1940). Auditory Patterns. *Reviews of modern physics, 12*, 47-65.
- Fournier, G. (2017). Perception des intervalles harmoniques et apprentissage perceptuel. *Revue musicale OICRM, 4(1)*.
- Garner, W. (1976). Adequate and inadequate stimulus information : comments on "some considerations of two alleged kinds of selective attention" by Keren. *J. Exp. Psychol. Gen.*, 105(4), 376-377.
- Gebhardt, C., Goldstein, D., & Robertson, R. (1972). Frequency discrimination and the MLD. *J. Acoust. Soc. Am.*, 51(4), 1228-1232.
- Goldstone, R., & Hendrickson, A. (2010). Categorical perception. Dans J. a. Wiley, *Wiley Interdisciplinary Reviews : Cognitive Science, 1(1)*, 69-78.
- Golestani, N. (2017). MOOC Neurosciences : audition, langage et musique.
- Grimault, N., Micheyl, C., Carlyon, R., Bacon S.P., Collet, L., (2003). Perceptual learning in frequency discrimination and amplitude modulation rate discrimination, and generalization to fundamental frequency discrimination. *Hear. Res.*, 184, 41-50.
- Hartmann, W. (2013). *Principles of musical acoustics*. New York, Etats-Unis: Springer.
- Kolinsky, R., & Hoonhorst, I. (2017). MOOC Neurosciences : audition, langage et musique.
- Kolinsky, R., Peretz, I., & Grandjean, D. (2017). MOOC Neurosciences : audition, langage et

musique.

Langner, G., & Ochse, M. (2006). The neural basis of pitch and harmony in the auditory system. *Musicae Scientiae*, 185-208.

Licklider, J. (1954). "Periodicity" pitch and "place" pitch. *J. Acoust. Soc. Am.*, 26, 945-950.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, H., & Moore, B. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc. Natl. Acad. Sci. USA.*, 103(49), 18866-18869.

McDermott, J., Lehr, A., & Oxenham, A. (2010). Individual Differences Reveal the Basis of Consonance. *Curr. Biol.*, 20(11), 1035-1041.

Meddis, R., & Hewitt, M. (1992). Modeling the identification of concurrent vowels with different fundamental frequencies. *J. Acoust. Soc. Am.*, 91(1), 233-245.

Micheyl, C., & Oxenham, A. (2010). Pitch, harmonicity and concurrent sound segregation : Psychoacoustical and neurophysiological findings. *Hear. Res.*, 266(1-2), 36-51.

Moore, B. (2004). *An Introduction to the Psychology of Hearing (5th edition)*. Elsevier.

Moore, B., Glasberg, B., & Peters, R. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *J. Acoust. Soc. Am.*, 77, 1853-1860.

Moore, B., Glasberg, B., Flanagan, H., & Adams, J. (2006). Frequency discrimination of complex tones; assesing the role of the component resolvability and temporal fine structure. *J. Acoust. Soc. Am.*, 119(1), 480-490.

Moore, B., Huss, M., Vickers, D., Glasberg, B., & Alcantara, J. (2000). A test for the diagnosis of dead regions in the cochlea. *Br J Audiol*, 34(4), 205-224.

Moore, G., & Moore, B. (2003). Perception of the low pitch of frequency-shifted complexes. *J. Acoust. Soc. Am.*, 113, 977-985.

Oxenham, A. (2013). The Perception of Musical Tones. Dans D. Deutch, *The Psychology of Music*, 3^{ème} édition, London: Elsevier., 1-33.

Patel, A., & Iversen, J. (2007). The linguistic benefits of musical abilities. *Trends. Cogn. Sci.*, 11(9), 369-372.

Patterson, R. (1969). Noise masking of a change in residue pitch. *J. Acoust. Soc. Am.*, 45(6), 1520-1524.

Patterson, R., Uppenkamp, S., Johnsrude, I., & Griffiths, T. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron.*, 36(4), 767-776.

Plack, C., Oxenham, A., Fay, R., & Popper, A. (2005). *Pitch : neural coding and perception*. New York : Springer.

Platel, H. (2010). Anatomie fonctionnelle de la perception et de la mémoire musicale (chap. 13). Dans H. Platel, *Le cerveau musicien*. De Boeck Supérieur.

- Rauschecker, J., & Shannon, R. (2002). Sending sound to the brain. *Science*, 295(5557), 1025-1029
- Renard, X. (2008). *Précis d'audioprothèse, Production, Phonétique acoustique et Perception de la parole*. Elsevier Masson.
- Roederer, J. (2008). *The Physics and Psychophysics of Music. An Introduction, 4^{ème} édition*. New York : Springer.
- Sek, A., & Moore, B. (1995). Frequency discrimination as a function of frequency, measured in several ways. *J. Acoust. Soc. Am.*, 97(4), 2479-2486.
- Shamma, S., & Klein, D. (2000). The case of the missing pitch templates : How harmonic templates emerge in the early auditory system. *J. Acoust. Soc. Am.*, 107, 2631-2644.
- Shannon, R., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Spec Recognition with Primarily Temporel Cues. *Science*, 270(5234), 303-304.
- Sinex, D., & Li, H. (2007). Responses of inferior colliculus neurons to double harmonic tones. *J. Neurophysiol.* , 98(6), 3171-3184.
- Wallace, M., Shackleton, T., & Palmer, A. (2002). Phase-locked responses to pure tones in the primary auditory cortex. *Hear Res.*, 172(1-2), 160-171.
- Zeng, F., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y., Chen H., (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues. *J. Acoust. Soc. Am.*, 116(3), 1351-1354.

Figures

Figure 1 : Largeur de la bande passante en fonction de la fréquence centrale du filtre.....	10
Figure 2 : Ondes en phase et en opposition de phase.....	11
Figure 3 : Verrouillage de phase : synchronisation des potentiels d'action avec la période de l'onde sonore entrante.....	14
Figure 4 : Modèle de reconnaissance de patterns.....	16
Figure 5 : Verrouillage de phase : synchronisation des potentiels d'action avec la période de l'onde complexe résultant de l'interaction des harmoniques irrésolues.....	17
Figure 6 : Représentation de la fusion harmonique – cas de deux voyelles entendues simultanément.....	20
Figure 7 : Représentation de l'information temporelle d'un son complexe.....	25
Figure 8 : Exemple de stimuli – expérience 1.....	32
Figure 9 : Résultats de l'expérience 1 : 20 dB dans le bruit.....	34
Figure 10 : Résultats de l'expérience 1 : 40 dB dans le silence	34
Figure 11 : Résultats de l'expérience 1 : 60 dB dans le silence	34
Figure 12 : Résultats individuels de l'expérience 1 : 20 dB dans le bruit	35
Figure 13 : Résultats individuels de l'expérience 1 : 40 dB dans le silence.....	35
Figure 14 : Résultats individuels de l'expérience 1 : 60 dB dans le silence	36
Figure 15 : Exemples de stimuli – expérience 2.....	38
Figure 16 : Résultats de l'expérience 2 : 8 dB SL.....	40
Figure 17 : Résultats de l'expérience 2 : 18 dB SL.....	40
Figure 18 : Résultats individuels de l'expérience 2.....	41

Schémas

Schéma 1 : Schéma des voies auditives.....	9
--	---

Annexes

Annexe 1 : Lexique

Octave : Une octave est l'intervalle entre deux notes opposées de la gamme, dont les fréquences ont un rapport de 2 : la fréquence de la note la plus aiguë est le double de celle de la plus grave. Perceptivement, deux notes séparées d'une octave présentent une certaine similarité, c'est-à-dire qu'il existe une « affinité » particulière entre les deux notes sur le plan perceptif.

Décibels : L'unité de mesure de l'intensité sonore est appelée *Décibel* (dB). Il existe plusieurs échelles différentes adaptées au type de mesure réalisé. Les plus utilisées dans le domaine de la psychophysique et de l'ORL sont le dB HL (Hearing Level), le dB SPL (Sound Pressure Level) et le dB SL (Sensation Level).

Le dB HL est défini par rapport à l'audition humaine : chez un sujet normo-entendant, la plus faible intensité audible est arbitrairement définie comme 0 dB HL (Danziger & Alamowitch, 2016). Effectuer un audiogramme permet d'obtenir les seuils du sujet : l'intensité sonore minimale (en dB HL) pour chaque fréquence, nécessaire pour que le son pur soit perçu par l'auditeur.

Le dB SPL correspond à l'intensité physique du son, le 0 étant défini par une certaine valeur de pression acoustique. Pour l'oreille humaine, il existe une grande variation de l'intensité perçue en fonction de la fréquence entre des sons ayant une même intensité en dB SPL. Par exemple, un son de 100 Hz à 40 dB SPL n'est pas audible, alors qu'un son de 1000 Hz à 40 dB SPL est d'intensité confortable pour une oreille humaine.

Le dB SL est utilisé en acoustique. Cette dimension correspond aux capacités auditives individuelles (et non à une moyenne considérée comme normale chez l'homme). Quand on propose un stimulus à 20 dB SL par exemple, il s'agit de 20 dB au-dessus du seuil de détection individuel du son.

Bruit : Contrairement au spectre d'un son périodique qui est composé d'une série limitée de fréquences, le spectre d'un bruit contient toutes les fréquences, avec une amplitude plus ou moins importante. L'onde sonore ne présente aucune organisation particulière et la variation de pression est totalement aléatoire au cours du temps. De manière générale, un bruit masquant est utilisé dans des expériences psycho-acoustiques pour empêcher le sujet de détecter certaines composantes du son par exemple.

-Bruit blanc Bruit qui contient toutes les fréquences audibles, chacune ayant la même intensité sonore.

-Bruit rose Une octave grave (100-200 Hz) couvre moins de fréquences qu'une octave aiguë (5000-10000 Hz). Dans le bruit rose, l'intensité des fréquences les plus graves est augmentée pour compenser ce déséquilibre. Quand la fréquence est multipliée par deux, le niveau sonore diminue de 3 dB (Fischetti, 2004). L'intérêt d'utiliser un bruit rose dans une tâche de psychoacoustique est de ne pas favoriser les fréquences graves ou les aiguës.

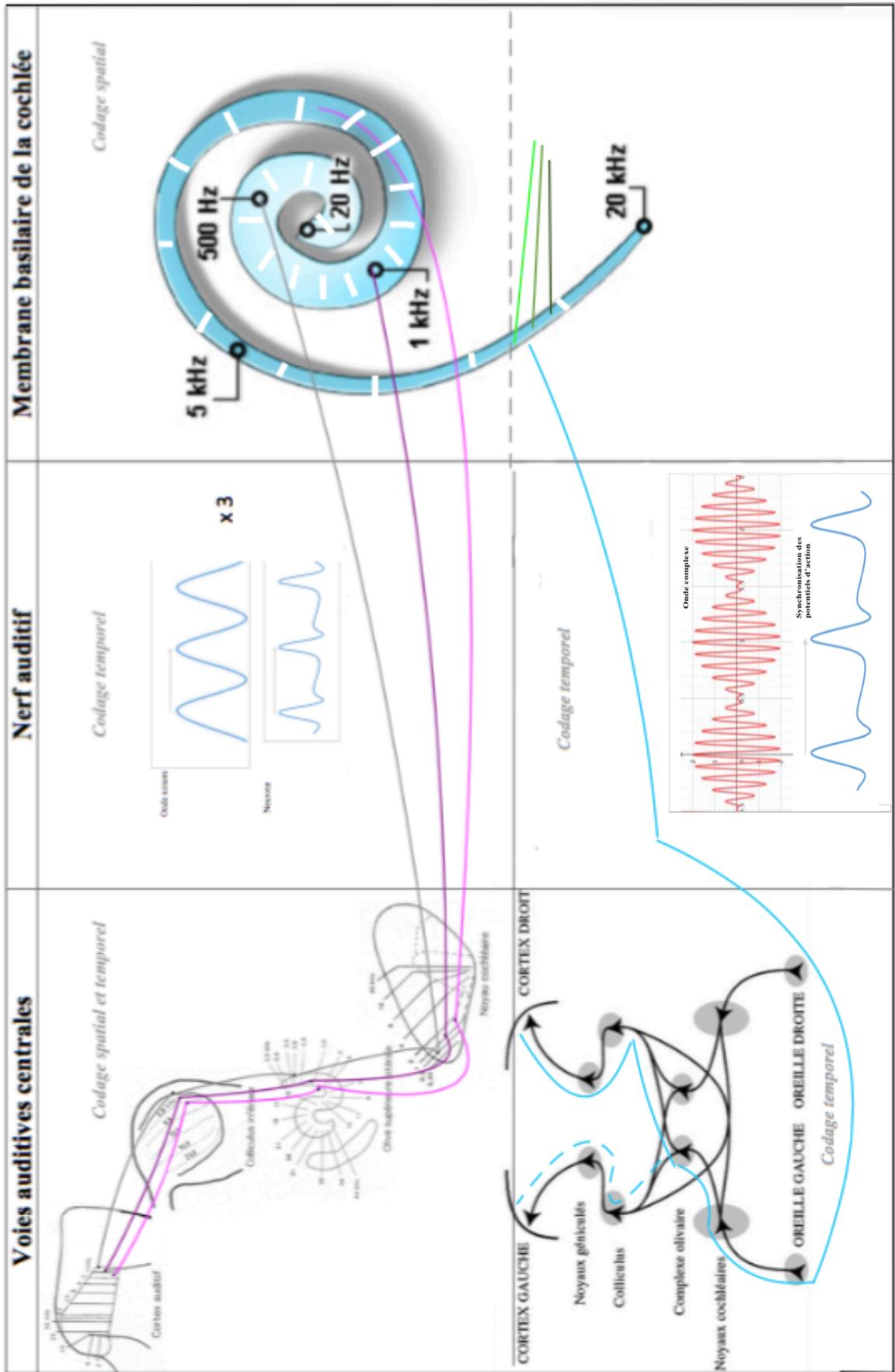
-Threshold Equalizing Noise (TEN) ou « bruit égalisateur de seuil » Bruit qui a un pouvoir uniformément masquant sur toutes les fréquences : la puissance acoustique délivrée à chaque filtre auditif cochléaire est la même. Un exemple d'utilisation du TEN (Moore et al., 2000) est d'éviter la détection d'un produit de distorsion. Ce phénomène, produit par le système auditif périphérique, aboutit à la production, par le système auditif lui-même, de certaines fréquences non présentes initialement dans le stimulus, et détectables par la cochlée. Il est alors pertinent d'utiliser un bruit masquant (le TEN), étudié spécifiquement pour donner la même intensité de masque sur toutes les fréquences. Ce son masquant permet au testeur d'être certain d'obtenir une réponse due aux fréquences réellement envoyées.

Annexe 2 : Conversions

- 1 cent = 1/100 demi-ton
- 100 cents = 1/2 ton = 1/12 octave
- 150 cents = 1,5 demi-ton = de Do à Do#-Ré
- 300 cents = 3 demi-tons = de Do à Ré#
- 600 cents = 6 demi-tons = de Do à Fa#
- 1200 cents = 12 demi-tons = 1 octave (rapport de fréquences égal à 2)

Annexe 3 :

Codage de la hauteur d'un son complexe par le système auditif.



H a r m o R é s.

H a r m o I r r é s.

Nous sommes capables de distinguer deux sons complexes (une voyelle ou une note de musique) produits en même temps, grâce à un phénomène acoustique particulier : la fusion harmonique. En d'autres termes, l'auditeur est capable d'extraire une hauteur unique d'un ensemble de composantes harmoniques d'une même source sonore.

L'objectif de ce mémoire est de comprendre l'origine du codage de la hauteur des sons complexes dans le système auditif en étudiant la détection d'inharmonicité et donc le phénomène de fusion harmonique, à travers deux expériences. Deux théories sont classiquement proposées pour expliquer le codage de la hauteur ; l'une repose sur des mécanismes spatiaux, l'autre sur des mécanismes temporels. Nous étudions notamment l'hypothèse d'un mécanisme temporel à l'origine dans ce codage.

Les résultats de l'expérience 1 indiquent que lorsque le système auditif est contraint d'utiliser une information temporelle, on retrouve une asymétrie dans une tâche de détection d'inharmonicité. Néanmoins, ces résultats pourraient aussi être expliqués par l'intervention d'un mécanisme spatial acquis avant la naissance jusqu'aux premiers mois de vie. Dans l'expérience 2, l'asymétrie est retrouvée dans toutes les conditions sauf dans celle où le niveau sonore est bas et les sons sont antiphases. Cela indique que le système binaural à lui seul est capable de détecter une inharmonicité. Ceci constitue une donnée nouvelle pour la compréhension de la fusion harmonique. Notre étude démontre que les mécanismes temporels jouent un rôle central dans les phénomènes de détection d'inharmonicité, sans exclure d'autres types de mécanismes associés.

Ces résultats ouvrent de nouvelles perspectives pour la recherche fondamentale et pour la pratique clinique en orthophonie. Il est notamment possible d'envisager une rééducation auditive basée sur un entraînement à la détection d'inharmonicité chez les patients atteints de surdité ou porteurs d'implants cochléaires, dans l'optique d'améliorer leurs capacités de fusion harmonique et ainsi leur compréhension de la parole dans un environnement bruyant.

Mots-clés : Harmonicité, Fusion harmonique, Hauteur, Codage temporel.

We are able to distinguish between two complex sounds (a vowel or a music note) produced at the same time, thanks to a particular acoustic phenomenon : the harmonic fusion. In other words, the listener is able to extract a single pitch from a set of harmonic components from the same sound source.

The aim of this thesis is to understand the origin of pitch coding of complex sounds in the auditory system based on the detection of inharmonicity, and therefore the phenomenon of harmonic fusion, through two experiments. Two theories are commonly proposed to explain this coding ; one based on spatial clues and the other on temporal clues. We hypothesized that a temporal mechanism is involved in this coding.

Results of experiment 1 indicates that when the auditory system is forced to use temporal information, we find an asymmetry in an inharmonicity detection task. On the other hand, these results could also be explained by the intervention of a spatial mechanism acquired before the birth until the first months of life. In experiment 2, the asymmetry of experiment 1 is found under all conditions except when the sound level is low and the sounds are antiphase. This means that binaural system alone is able to detect an inharmonicity. This is new data, which adds to our understanding of harmonic fusion. Our study demonstrates that temporal mechanisms play the main role in inharmonicity detection phenomenon, without excluding other kinds of associated mechanisms.

With our results in mind, fundamental research will continue to investigate this topic. Our research also has implications for the clinical practice of speech therapy. For example, it is possible to imagine auditory rehabilitation based on inharmonicity detection training in patients with hearing loss or cochlear implants, in order to improve their harmonic fusion capabilities and thus their speech understanding in a noisy environment.

Mots-clés : Harmonicity, Harmonic fusion, Pitch, Temporal theory.