La perception de l'harmonicité des sons en présence ou en l'absence de surdité neurosensorielle

Damien Bonnard

▶ **To cite this version:**

Damien Bonnard. La perception de l'harmonicité des sons en présence ou en l'absence de surdité neurosensorielle. Médecine humaine et pathologie. 2013. dumas-00960057

HAL Id: dumas-00960057

https://dumas.ccsd.cnrs.fr/dumas-00960057

Submitted on 19 Mar 2014

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.

UNIVERSITÉ BORDEAUX SEGALEN
UFR DES SCIENCES MÉDICALES

Année 2013 Thèse n° 3078

Thèse pour l'obtention du

DIPLÔME d'ÉTAT de DOCTEUR EN MÉDECINE

Présentée et soutenue publiquement
Le 23/10/2013

Par Damien BONNARD

Né le 01/04/1982 à Romans-sur-Isère

La perception de l'harmonicité des sons en présence ou en l'absence de surdité neurosensorielle

Directeur de Thèse :

Monsieur le **Pr. René Dauman**

Jury :

Monsieur le **Pr. Jean-Pierre Bébéar**.....
Monsieur le **Pr. Vincent Darrouzet**.....
Monsieur le **Dr. Laurent Demany**.....
Madame le **Pr. Valérie Franco-Vidal**.....
Monsieur le **Pr. Christian Lorenzi** ..... *Rapporteur*
Monsieur le **Pr. Roger Marthan**..... *Président*

A mon fils, Gabin

Remerciements

A Monsieur le **Pr. Roger Marthan**. Votre présence nous honore, et je souhaite vous témoigner toute ma gratitude pour avoir accepté la présidence de ce Jury.

A Monsieur le **Pr. Christian Lorenzi**. Je vous suis particulièrement reconnaissant d'avoir accepté de juger ce travail. Votre expertise et votre regard scientifique aiguisé garantissent à cette thèse une évaluation de grande qualité.

A Monsieur le **Pr. René Dauman**. Veuillez accepter mes remerciements les plus sincères pour avoir dirigé ce travail, mais aussi pour m'avoir toujours épaulé et guidé dans cette voie depuis plusieurs années. Vos conseils, votre expérience et votre soutien sont pour moi des atouts irremplaçables dont je vous suis infiniment reconnaissant.

A Monsieur le **Pr. Vincent Darrouzet**. Travailler à vos côtés est une chance dont j'ai pleinement conscience. Votre disponibilité, votre extrême compétence et l'attention que vous portez à la formation de vos élèves vous honorent. Soyez assuré de mon plus respectueux dévouement pour les années à venir.

A Monsieur le **Pr. Jean-Pierre Bébéar**. Je vous exprime ici ma plus chaleureuse et sincère reconnaissance. Vous m'avez offert un véritable compagnonnage depuis le début de mon internat et j'ai énormément appris grâce à vous ; sur les implants cochléaires et sur notre belle spécialité naturellement, mais aussi plus largement sur les relations humaines. Votre passion inébranlable et communicative pour notre discipline est un exemple pour moi. Merci du fond du cœur.

A Madame le **Pr. Valérie Franco-Vidal**. Tu arrives à combiner avec talent une disponibilité hors pair, une rigueur de tous les instants et un naturel rafraichissant dans les rapports humains. J'admire ta conception des relations avec les patients et ton dévouement à l'Hôpital Public. C'est avec plaisir et respect que je continuerai à me former sous ta houlette.

A Monsieur le **Dr. Laurent Demany**. J'ai pu bénéficier pendant ces deux ans au « labo » de ton amour de la belle science, de ta précision et de ton honnêteté dans le

travail. La réalisation de cette thèse n'aurait pas été possible sans ton active participation et tes précieux conseils au quotidien. Merci pour tes patientes relectures, pour ta disponibilité, et pour ces discussions qui m'ont guidé tout au long de mon travail.

A toute **l'équipe du B1**, infirmières, aides soignantes et ASH pour tous ces bons moments durant mon internat. Je suis convaincu que cette immense ruche qu'est le CHU fonctionne avant tout grâce au dévouement de ses abeilles ouvrières. J'adresse des remerciements tout particuliers à **Mme Maria Da Cunha** pour la sympathie et le professionnalisme dont elle a toujours fait preuve.

A **l'équipe d'instrumentistes du bloc ORL**, pour votre efficacité qui n'entache en rien votre bonne humeur… à très bientôt !

A **l'équipe des implants cochléaires**, merci de m'avoir accueilli parmi vous, merci d'entretenir ce niveau d'excellence, cette convivialité, et ce profond respect des patients.

Aux **infirmières audiométristes**, merci pour le travail que vous faites, et merci à celle qui a directement participé à cette thèse en se portant volontaire pour l'expérience.

A l'équipe du labo : **Mme Catherine Semal, Mayalen** et **Sam**, pour le très bon souvenir que je garderai de ces deux ans passés ensemble.

A mes chefs de clinique : **Olivier**, **Antoine**, **Violaine**, **Michel**.

A mes co-internes : **Magali, Vincent, Pierre-Louis, Amandine, Chloé, Isabelle, Dorothée, Bertrand, David, Alix, Sylvestre, Julien, Sylvain, Matthieu, Eve, Laure, Thomas, Florence, Solange.**

A mes parents, mes frères, ma grand-mère, qui m'ont toujours accompagné.

A Stooph (qui se reconnaîtra) et à mes amis du lycée (qui se reconnaîtront).

Et enfin à Séverine qui m'apporte chaque jour son amour et qui m'a soutenu pendant ces longs mois de travail. Je n'imaginerai plus vivre sans toi.

Résumé

Le système auditif central d'un sujet normo-entendant fusionne perceptivement les composantes spectrales des sons complexes harmoniques, bien que ces composantes soient initialement séparées par la cochlée. Cette *fusion harmonique* joue un rôle important dans l'analyse des scènes acoustiques multi-sources, situations dans lesquelles la gêne auditive des patients atteints de surdité neurosensorielle est généralement exacerbée. L'objectif de cette thèse était d'une part de clarifier les mécanismes de la fusion harmonique chez les auditeurs sains, et d'autre part d'étudier l'impact des surdités neurosensorielles sur la sensibilité à l'harmonicité.

Une première expérience, menée sur des auditeurs sains, a porté sur la reconnaissance auditive de l'octave harmonique, formée par des sons purs simultanés dont le rapport fréquentiel est égal à 2. L'expérience montre que le mécanisme permettant cette reconnaissance est insensible à la direction d'un écart par rapport à l'octave. Il apparaît par contre que, pour des sons purs présentés *consécutivement*, l'octave ne représente plus une singularité du point de vue de la discrimination perceptive.

Une seconde expérience, incluant des sujets atteints de surdité neurosensorielle légère à moyenne, montre que l'effet de ces surdités sur la sensibilité *absolue* à l'harmonicité est très variable d'un individu à l'autre, et semble refléter avant tout le niveau d'altération de la discrimination fréquentielle. Cependant, l'expérience fait apparaître chez tous les sujets testés la conservation d'une sensibilité *relative* à l'harmonicité : le seuil de détection d'une inharmonicité par rapport à l'octave s'avère toujours meilleur que le seuil de détection d'une différence entre des rapports de fréquences inharmoniques, même en présence de pertes tonales moyennes supérieures à 45 dB entre 500 et 4000 Hz.

SOMMAIRE

1 AVANT-PROPOS.....8

2 PERCEPTION DE LA HAUTEUR ET FUSION HARMONIQUE CHEZ LES SUJETS NORMO-ENTENDANTS 10

2.1 Perception de la hauteur des sons purs.....10

2.2 Perception de la hauteur des sons complexes et fusion harmonique 14

2.2.1 Traitement d'un son complexe par la cochlée 16

2.2.2 Le concept de fusion harmonique 19

2.2.3 Données psychoacoustiques 21

(a) Conséquences perceptives de la suppression de composantes d'un son complexe.....21

(b) Conséquences perceptives du déplacement d'une ou plusieurs composantes d'un son complexe.....23

2.2.4 Les grands modèles de perception de la hauteur et la fusion harmonique..... 26

(a) Les indices potentiels du codage de la hauteur.....26

(b) Les modèles de reconnaissance de pattern 28

(c) Les modèles d'autocorrélation 30

2.2.5 Données neurophysiologiques 35

(a) Nerf auditif, noyau cochléaire et colliculus inférieur.....36

(b) Cortex 38

(c) Synthèse des données neurophysiologiques actuelles.....41

3 LA DISCRIMINATION AUDITIVE DE RAPPORTS DE FREQUENCES 44

3.1 Introduction 44

3.2 Article "Auditory Discrimination of Frequency Ratios: The Octave Singularity" 46

4 PHYSIOPATHOLOGIE ET CONSEQUENCES DES SURDITES NEUROSENSORIELLES . 59

4.1 Les différents types de surdité 59

4.2 La dégradation des mécanismes actifs.....63

4.3 Conséquences des lésions cochléaires sur la sélectivité fréquentielle.....66

4.4 Conséquences des lésions cochléaires sur la sensibilité à la structure temporelle fine 69

5 PERCEPTION DE LA HAUTEUR EN PRESENCE D'UNE SURDITE NEUROSENSORIELLE 72

5.1 Perception de la hauteur des sons purs et zones mortes 72

5.2 Perception de la hauteur des sons complexes en présence de lésions cochléaires 74

6 CONSEQUENCES DES SURDITES NEUROSENSORIELLES SUR LA SENSIBILITE A L'HARMONICITE 78

6.1 Introduction 78

6.2 Matériel et méthodes 82

6.2.1 Participants et critères d'inclusion 82

6.2.2 Stimuli, tâche et conditions 83

6.2.3 Choix du niveau d'intensité 86

6.2.4 Procédure 87

6.3 Résultats 89

6.3.1 Sensibilité à l'harmonicité 89

6.3.2 Effet de la variation aléatoire de F 92

6.4 Discussion 97

7 CONCLUSION GENERALE 104

Bibliographie 106

1 AVANT-PROPOS

Il existe une relation étroite et pourtant méconnue par les oto-rhino-laryngologistes entre la psychoacoustique, science consacrée à l'étude et à la mesure de la perception auditive humaine, et l'activité clinique dans le domaine de l'otologie et de l'audiologie. Un exemple concret de cette relation est représenté par l'audiométrie subjective pratiquée au quotidien dans ces services, et qui reste dans la majorité des cas le moyen le plus fiable d'évaluer cliniquement une surdité. L'audiométrie subjective n'est autre qu'une mesure psychoacoustique du seuil auditif absolu d'un patient. Dès lors, cette discipline apparait comme le prolongement naturel de l'audiologie clinique vers la recherche fondamentale sur le système auditif et sur les différentes conséquences des surdités.

Ce travail m'a permis de matérialiser un peu plus cette relation privilégiée qui peut exister entre un service universitaire d'oto-rhino-laryngologie et un laboratoire de psychoacoustique. Il a été mené au sein de l'équipe « Perception Auditive et Développement » dirigée par le Dr. Laurent Demany (Institut de Neurosciences Cognitives et Intégratives d'Aquitaine, UMR CNRS 5287). Ce laboratoire, situé sur le site du Centre Hospitalier Universitaire Bordeaux Pellegrin, fait partie du campus universitaire de Bordeaux Segalen. Cette proximité géographique avec le service d'oto-rhino-laryngologie du Pr. Vincent Darrouzet, et l'unité d'audiologie du Pr. René Dauman, est un atout indiscutable pour les travaux sur les sujets sourds.

Notre thématique de recherche découle de l'une des plaintes les plus fréquemment rapportées par les patients atteints de surdité neurosensorielle : la gêne auditive dans « le bruit », plus précisément dans un environnement acoustique

comportant plusieurs sources sonores actives simultanément. Bien souvent, ces troubles de l'intelligibilité de la parole qui se manifestent dans des situations de la vie quotidienne telles qu'une réunion ou un repas de famille, persistent –au moins partiellement– malgré le port d'aides auditives. Patient et praticien se retrouvent donc fréquemment démunis face à cette gêne socialement invalidante. Ces difficultés contrastent avec la remarquable faculté des sujets normo-entendants à séparer sans effort des sources sonores concomitantes. Celle-ci repose notamment sur la capacité du système auditif à « fusionner » perceptivement les différentes composantes fréquentielles d'une source sonore, et en particulier les composantes dont les fréquences forment des rapports simples (dits harmoniques). Les travaux présentés ci-après avaient deux objectifs principaux. Premièrement, nous avons cherché à préciser les connaissances sur ce mécanisme de *fusion harmonique* chez les sujets normo-entendants ; même si plusieurs hypothèses concernant le substrat neuronal de la fusion harmonique ont été proposées, cette question reste toujours largement débattue. L'étude psychoacoustique que nous avons réalisée a donné lieu à la publication d'un article intégralement présenté dans cette thèse, après une synthèse des connaissances indispensables à sa compréhension. Le deuxième objectif était d'évaluer les conséquences des surdités neurosensorielles sur la fusion harmonique. Aucune étude n'a été publiée à ce jour sur cette question, et il est très difficile de préjuger de l'impact des lésions cochléaires sur la sensibilité à l'harmonicité à partir des connaissances actuelles. Les résultats préliminaires de cette seconde étude, pour laquelle les inclusions sont toujours en cours, sont assez inattendus ; ils sont présentés et discutés dans la dernière partie de cette thèse.

2 PERCEPTION DE LA HAUTEUR ET FUSION HARMONIQUE CHEZ LES SUJETS NORMO-ENTENDANTS

2.1 Perception de la hauteur des sons purs

Les mécanismes qui sous-tendent la perception de la hauteur sont l'un des phénomènes les plus étudiés par la psychoacoustique et restent pourtant partiellement incompris. La hauteur est classiquement définie comme « l'attribut de la sensation auditive grâce auquel les sons peuvent être ordonnés sur une échelle s'étendant des graves aux aigus » (ANSI, 1994). Certains auteurs préfèrent la caractériser sous un angle plus musical, comme « l'aspect de la sensation auditive dont la variation détermine une mélodie » (Plack, 2005; Burns & Viemeister, 1976). La hauteur d'un son n'est pas une propriété physique directement mesurable mais elle correspond au corrélat perceptif de la fréquence de répétition de l'onde sonore. Le cas le plus simple est celui des sons purs qui correspondent à une onde sinusoïdale. La relation entre la hauteur perçue ou *tonie* (propriété subjective), exprimée en *Mel*, et la fréquence des sons purs (propriété objective) est monotone et suit une fonction logarithmique (Stevens, Volkmann, & Newman, 1937). Les sons de basse fréquence sont perçus comme graves, les sons de haute fréquence comme aigus.

Deux grands types de modèles ont été établis pour rendre compte des mécanismes utilisés par le système auditif pour coder la hauteur d'un son pur : les modèles spatiaux et les modèles temporels. Les modèles spatiaux reposent sur la tonotopie cochléaire, selon laquelle la position du pic maximum d'excitation sur la

membrane basilaire dépend de la fréquence de stimulation (les basses fréquences stimulant l'apex de la cochlée et les hautes fréquences la base). Ainsi, chacun des neurones auditifs constituant le nerf cochléaire répond préférentiellement à une fréquence donnée (appelée *fréquence caractéristique*), qui est déterminée par sa position d'origine le long de la cochlée. Dans les modèles spatiaux, on considère que la « hauteur perçue correspond à la position du maximum d'excitation » (Moore B. C., 2007), ce qui revient à dire que le système auditif détecterait un pic dans le profil d'excitation neuronale déterminant la hauteur perçue. Il est démontré que l'organisation tonotopique, et donc potentiellement le codage spatial de la hauteur des sons purs, se retrouve à différents niveaux des voies auditives, depuis le nerf cochléaire jusqu'au cortex auditif (Humphries, Liebenthal, & Binder, 2010).

Les modèles temporels quant à eux reposent sur le phénomène de verrouillage de phase (ou *phase locking*), c'est-à-dire la synchronisation entre les potentiels d'action dans le nerf auditif et les mouvements d'oscillation de la membrane basilaire stimulée par une onde sonore. Les potentiels d'actions ne se déclenchent pas obligatoirement à chaque cycle d'oscillation, mais ils se produisent approximativement toujours à la même phase de l'onde. En réponse à un son pur de période T, la majorité des intervalles entre potentiels d'action dans le nerf auditif aura une durée voisine de T ou de multiples entiers de T (Rose, Brugge, Anderson, & Hind, 1968). La limite supérieure du verrouillage de phase reste malheureusement inconnue chez l'homme, une mesure électrophysiologique directe étant éthiquement impossible. Les études disponibles sur les mammifères indiquent que ce mécanisme se dégrade pour des fréquences supérieures à 4 à 5 kHz (Palmer & Russell, 1986), même si des données plus récentes suggèrent qu'il pourrait persister au moins partiellement jusqu'à 10 kHz (Heinz, Colburn, & Carney, 2001). La limite neurophysiologique de 4 à 5 kHz traditionnellement admise est à rapprocher des données psychoacoustiques qui indiquent qu'une séquence de sons purs

de fréquences supérieures à 5 kHz n'évoque pas clairement une mélodie (Attneave & Olson, 1971; Semal & Demany, 1990; Oxenham, Micheyl, Keebler, Loper, & Santurette, 2011). Si les intervalles musicaux ne sont perçus correctement qu'en dessous de cette limite, il est néanmoins possible de détecter des variations de hauteur sur toute la gamme des fréquences audibles (Moore B. C., 2007). Les modèles temporels peuvent difficilement rendre compte d'une sensibilité à la hauteur pour des sons purs dont la fréquence est supérieure à 4 à 5 kHz, contrairement aux modèles spatiaux qui sont compatibles avec la perception de la hauteur sur l'ensemble du spectre auditif humain.

Néanmoins, de nombreuses données expérimentales vont à l'encontre des modèles de type spatial, ou du moins indiquent que les informations temporelles jouent un rôle important dans la perception de la hauteur (Moore B. C., 2008). On observe chez un auditeur sain des variations dans la perception de la hauteur d'un son pur (de fréquence constante) en fonction du niveau d'intensité. Ces variations ont été mesurées expérimentalement de 1 à 5% selon le domaine fréquentiel considéré (Verschuure & van Meeteren, 1975). Le décalage spatial des patterns d'excitation vers la base de la membrane basilaire lors d'une augmentation du niveau d'intensité prévoirait une modification de l'ordre d'une demi-octave, donc bien supérieure aux données psychoacoustiques. Le modèle spatial tel que nous l'avons présenté ne peut pas rendre compte des ces observations. De même, un modèle spatial tel que celui proposé par Zwicker (1970) ne permet pas d'expliquer les variations du seuil de discrimination de fréquences, c'est-à-dire la plus petite variation de fréquence détectable (souvent notée DLF pour *difference limen for frequency*) en fonction du domaine de fréquence considéré. Dans un modèle de type spatial, deux fréquences proches sont différenciées sur la base d'une comparaison de la position de leur pattern d'excitation respectif (Figure 1).

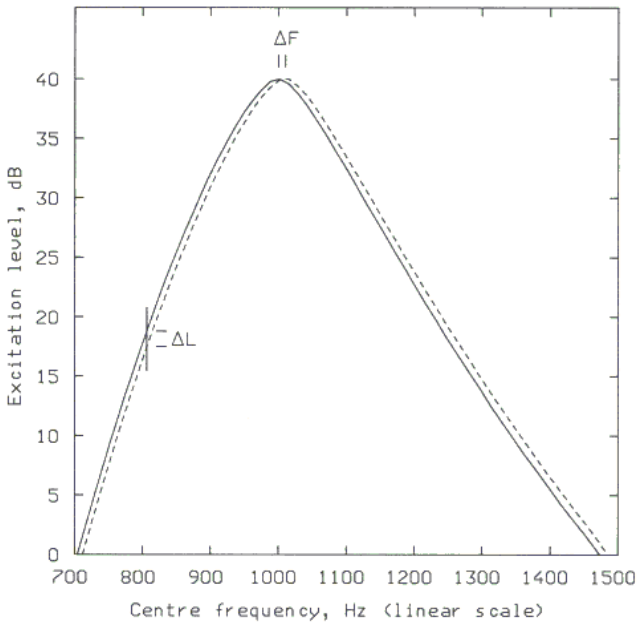


Figure 1. Les patterns d'excitation de deux sons purs de fréquences respectives 1000 Hz (trait plein) et 1010 Hz (trait pointillé) sont représentés. Pour détecter une différence de fréquence de ΔF Hz, le système auditif doit être capable de détecter une différence d'amplitude entre les patterns de ΔL dB en un point donné de la membrane basilaire. Compte tenu de l'asymétrie des pentes des patterns d'excitation, la variation d'amplitude est plus facilement détectable sur le versant correspondant aux plus basses fréquences. En considérant comme Zwicker que la valeur minimale détectable de ΔL est voisine de 1 dB, la valeur de ΔF correspondante (pour deux fréquences proches de 1000 Hz) serait de l'ordre de 7 Hz, ce qui est bien supérieur aux données psychoacoustiques (entre 1 et 2 Hz). D'après Moore (2007).

Le système auditif doit détecter une différence d'amplitude suffisante entre les deux patterns en un point donné de la membrane basilaire (de l'ordre de 1 dB selon Zwicker) pour discriminer les deux fréquences. Pour les basses fréquences (inférieures à 2 kHz), les performances de discrimination de fréquences mesurées expérimentalement qui sont de l'ordre de 0.2 à 0.4% de la fréquence centrale concernée (Sek & Moore, 1995), sont deux à trois fois plus élevées que ne le prédit ce modèle. Autrement dit, un modèle de type spatial a du mal à rendre compte de la précision extrême du codage de la hauteur dans les basses fréquences. Les DLFs augmentent de manière monotone avec la fréquence, et cette augmentation est plus marquée à partir de 4 à 6 kHz (Micheyl, Xiao, & Oxenham, 2012; Sek & Moore, 1995). Cela suggère que le codage de la hauteur des sons purs est dégradé et imprécis pour les hautes fréquences (à partir de 4 kHz), même si des différences de hauteurs sont toujours détectables.

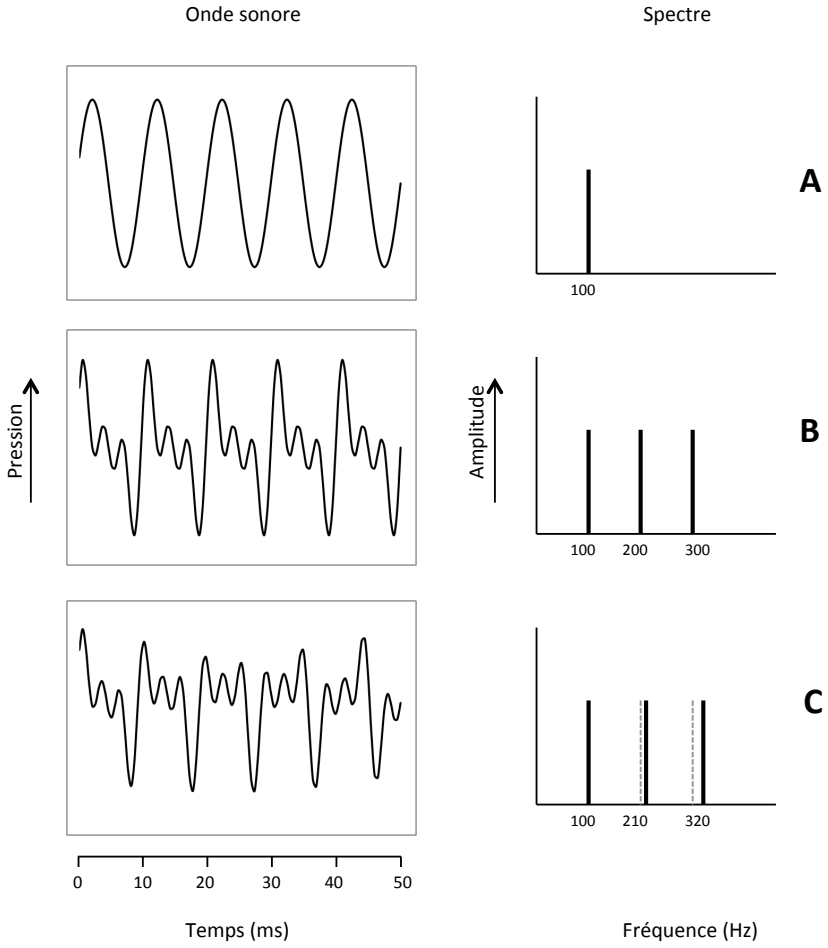
Les deux types d'informations, spatiales et temporelles, sont présentes dans le nerf auditif (cf. section 2.2.5) ; le problème qui reste incomplètement résolu est de savoir

dans quelle mesure ces informations sont utilisées par le système auditif central pour le codage de la hauteur. Un modèle de type temporel rendrait compte des bonnes performances dans les basses fréquences par un codage précis de la hauteur jusqu'aux limites traditionnellement admises du verrouillage de phase (4 à 5 kHz). Pour les fréquences supérieures, l'information temporelle n'étant vraisemblablement plus disponible, un modèle de type spatial expliquerait la conservation d'une sensibilité à la hauteur, même si celle-ci devient moins précise. Ainsi le type de tâche utilisée ou le domaine fréquentiel concerné aboutissent à considérer ces deux modèles non pas comme concurrentiels, mais au contraire complémentaires pour expliquer les mécanismes de perception de la hauteur des sons purs. Sur la plan physique, il y a un lien direct entre la forme de l'onde sonore (sa structure temporelle) et le spectre (sur lequel reposent les modèles spatiaux) : ces deux représentations sont liées par les lois mathématiques de la transformée de Fourier. Comme le note Yost (2009), la difficulté à distinguer les aspects spatiaux et temporels dans les mécanismes de perception de la hauteur est probablement le reflet de cette intrication entre les deux concepts.

2.2 Perception de la hauteur des sons complexes et fusion harmonique

Les sons purs sont très utiles dans la recherche psychoacoustique car ils sont simples et permettent d'étudier des propriétés fondamentales de l'audition. En revanche ils sont quasiment inexistants dans notre environnement sonore naturel. La plupart des sons produits par les animaux, les instruments de musique ou la voix humaine (et notamment les voyelles) sont de type complexes périodiques. Les sons complexes périodiques correspondent à une onde qui n'est pas une simple sinusoïde, mais qui présente tout de même une fréquence de répétition, dite fréquence fondamentale. Un son complexe périodique peut être construit selon le principe mathématique d'une

transformée de Fourier par sommation de plusieurs sons purs sinusoïdaux, appelés harmoniques. Ces harmoniques ont une fréquence égale à la fréquence fondamentale (première harmonique, de fréquence f0) ou à des multiples entiers de la fréquence fondamentale (2*f0, 3*f0, etc.). Par exemple, un son complexe comprenant les composantes fréquentielles 100, 200, et 300 Hz respecte cette définition, il est dit harmonique (Figure 2). Si les composantes d'un son complexe ne sont pas des multiples entiers de f0, le son complexe n'est pas périodique ; il est dit inharmonique.



01

02

03

Figure 2. Représentation de l'onde sonore (partie gauche de la figure) et du spectre (partie droite) de trois sons A, B et C. **(A)** Son pur de fréquence 100 Hz. Le spectre montre une unique composante fréquentielle égale à 100 Hz. **(B)** Son complexe périodique comprenant trois composantes fréquentielles : 100, 200 et 300 Hz. La hauteur perçue pour ce type de son harmonique est celle de la fréquence fondamentale (100 Hz), qui correspond à la fréquence de répétition de l'onde complexe. Les sons A et B ont donc la même hauteur, même si les spectres sont très différents. **(C)** Son complexe comprenant trois composantes fréquentielles : 100, 210 et 320 Hz. Ce type de son est perçu comme inharmonique par le système auditif, les composantes spectrales de rangs 2 et 3 n'étant pas à leur place « correcte » (indiquée par la ligne pointillée grise). Un auditeur perçoit généralement dans ce cas une sensation dissonante de plusieurs sons simultanés de hauteurs différentes.

Il faut cependant nuancer cette définition formelle qui n'est pas perceptivement valable pour des f0s très faibles et/ou pour des séries harmoniques très incomplètes. Par exemple, un son contenant les composantes 200, 310, et 320 Hz (Figure 2C) est perçu comme inharmonique, car on peut grossièrement considérer que le système auditif détecte que les composantes de rangs 2 et 3 ne sont pas des multiples entiers de 100 Hz. Cependant, on pourrait concevoir la série 100, 210 et 320 Hz comme les composantes de rangs 10, 21 et 32 d'une f0 de 10 Hz. Cette série resterait harmonique selon la définition donnée ci-dessus. Mais dans ce cas, d'une part la f0 (10 Hz) est en dehors du champ auditif humain, et d'autre part le nombre d'harmoniques manquantes est très élevé.

Les différentes composantes fréquentielles d'un son complexe constituent son spectre. D'une manière générale, la hauteur d'un son complexe périodique correspond à la fréquence de répétition de l'onde sonore (f0). Un son complexe harmonique de fréquence fondamentale f0 a une hauteur proche d'un son pur de fréquence f0, bien que leurs spectres soient très différents (Figure 2A, B). L'amplitude relative des différentes harmoniques du son complexe détermine le timbre.

2.2.1 Traitement d'un son complexe par la cochlée

Pour comprendre comment la hauteur des sons complexes peut être déterminée, il est primordial de représenter les premières étapes de leur traitement par le système auditif périphérique. Prenons l'exemple d'un son complexe de f0 égale à 440 Hz et comprenant les 18 premières harmoniques. Son spectre contient les composantes fréquentielles suivantes : 440 Hz, 880 Hz, 1320 Hz, …, 7480 Hz, 7920 Hz. Ce son complexe constitue le signal d'entrée pour le système auditif périphérique. On peut le représenter de deux façons : (1) Comme une variation de pression au cours du temps, ce qui équivaut à une représentation temporelle de l'onde sonore et met en évidence la

période égale à 1/f0 soit environ 2.3 ms (Figure 3A). (2) Par l'ensemble de ses composantes fréquentielles, ce qui équivaut à une représentation spectrale indiquant le niveau d'énergie en fonction de la zone fréquentielle considérée (Figure 3B). Cette dernière représentation est particulièrement adaptée au système auditif périphérique dont l'une des propriétés majeures est la sélectivité fréquentielle, c'est-à-dire la capacité à séparer les différentes composantes spectrales d'un son complexe. La sélectivité fréquentielle repose sur l'organisation tonotopique de la cochlée qui permet une représentation spatiale des différentes harmoniques, chacune entrainant la vibration d'une zone spécifique de la membrane basilaire. Ce système passif de vibration est complété par un mécanisme actif dépendant des cellules ciliées externes qui améliore considérablement la sélectivité fréquentielle (Liberman, Dodds, & Learson, 1986). La cochlée peut être assimilée à une juxtaposition de filtres passe-bande, dont la largeur absolue (en Hz) augmente avec la fréquence centrale (Figure 3C). Ainsi les harmoniques de faible rang sont dites résolues, c'est-à-dire individualisées par le système auditif, puisque deux harmoniques successives seront traitées par deux filtres différents. Au contraire, les harmoniques de rang élevé (c'est-à-dire approximativement au-delà de la 8ème harmonique) sont non résolues car plusieurs harmoniques successives sont traitées par un unique filtre (Moore B. C., 2007). On peut analyser le signal obtenu à la sortie des filtres sous la forme d'un pattern d'excitation (Figure 3D). Il correspond à une représentation du niveau d'excitation mécanique ou neuronale en fonction de la position –et donc de la fréquence caractéristique– sur la membrane basilaire. Les harmoniques résolues apparaissent nettement individualisées sous la forme de pics saillants, contrairement aux harmoniques non résolues. Le caractère résolu ou non résolu des harmoniques a une conséquence directe sur la forme de l'onde qui stimule la membrane basilaire dans la zone considérée.

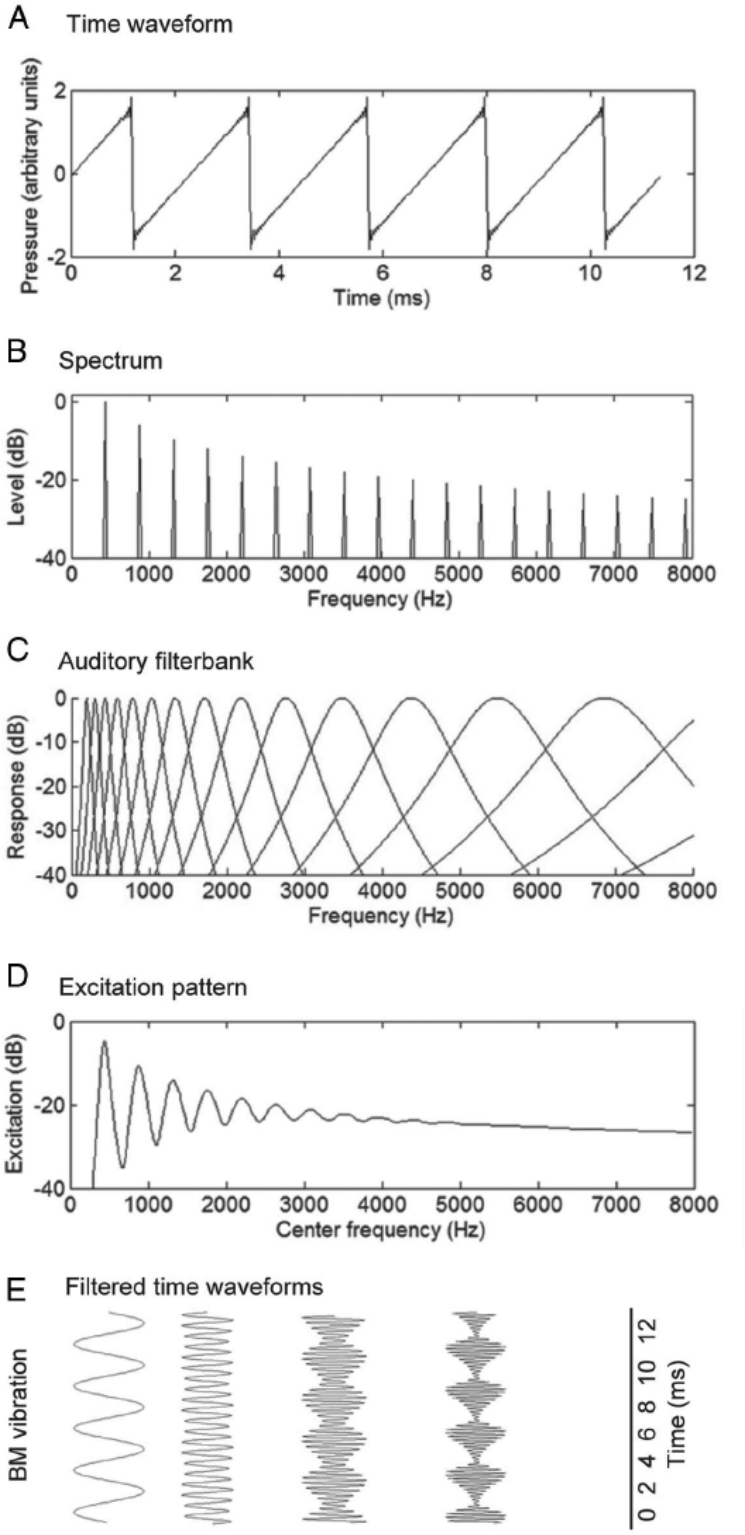


Figure 3. Représentations d'un son complexe harmonique (de f0 égale à 440 Hz et composé des 18 premières harmoniques) et de son traitement par le système auditif périphérique. **(A)** Représentation temporelle de l'onde sonore, faisant apparaître la période fondamentale de 2.3 ms (réciproque de 440 Hz). **(B)** Représentation spectrale du stimulus faisant apparaître les 18 harmoniques s'échelonnant de 440 Hz à 7920 Hz. **(C)** Banque de filtres passe-bande illustrant le mécanisme de filtrage cochléaire. **(D)** Pattern d'excitation exprimant le niveau moyen d'excitation en sortie des filtres auditifs en fonction de la fréquence caractéristique. Les harmoniques de faible rang sont résolues car clairement identifiables par un pic saillant dans le pattern d'excitation ; au contraire les harmoniques de rang élevées non résolues ne sont plus individualisables. **(E)** Représentation de la vibration de la membrane basilaire en sortie de quatre filtres auditifs. Les deux courbes de gauche correspondent aux filtres centrés sur les harmoniques résolues de rangs 1 et 4; la membrane basilaire est stimulée par une onde sinusoïdale. Les deux courbes de droite correspondent aux filtres centrés respectivement sur la 8ème et la 12ème harmonique, non résolues. Dans ce cas, la membrane basilaire est stimulée par une onde complexe issue de la combinaison de plusieurs harmoniques successives interagissant dans un même filtre auditif. D'après Oxenham (2005).

La Figure 3E représente la vibration de la membrane basilaire au cours du temps en quatre emplacements différents. Les deux courbes de gauche correspondent à des harmoniques parfaitement résolues (harmoniques de rangs 1 et 4 respectivement) ; on peut considérer dans ce cas que la portion de la membrane basilaire centrée sur la fréquence caractéristique 440 Hz pour l'harmonique 1 (ou 1760 Hz pour l'harmonique 4), est stimulée par une onde sinusoïdale de cette fréquence, les composantes adjacentes ne passant pas dans le même filtre. En revanche, quand les harmoniques ne sont pas résolues comme dans les deux courbes de droite, la zone correspondante de la membrane basilaire est stimulée par une onde complexe qui résulte de la sommation des différentes composantes fréquentielles adjacentes passant dans le filtre. Ces composantes étant des multiples entiers de la fréquence fondamentale, l'onde complexe présente une périodicité identique à celle de la fondamentale ; cela se traduit par une fluctuation d'amplitude de son enveloppe à la fréquence f0.

2.2.2 Le concept de fusion harmonique

Les premières étapes du traitement du signal par la cochlée aboutissent à une décomposition spectrale des sons complexes. Pourtant, un son harmonique est perçu par un auditeur sain comme une entité cohérente d'une seule hauteur tonale (f0), émise par une seule source (Bregman & Doehring, 1984), et non comme la juxtaposition de plusieurs sons purs. Le système auditif semble avoir la capacité d'identifier que des composantes fréquentielles ont des relations harmoniques entre elles et de les fusionner pour former un son unique (Kalluri, Depireux, & Shamma, 2008). Ce concept est classiquement dénommé *fusion harmonique*. Ce mécanisme joue probablement un rôle très important pour ségréguer les différentes sources sonores en milieu bruyant ou multi-locuteurs (Carlyon & Gockel, 2008). Imaginons une situation simplifiée de conversation multi-locuteurs où, à un instant donné, deux voyelles (v1 et v2) sont émises

simultanément par deux locuteurs distincts (Figure 4) ; chacune donnera au niveau de la cochlée de l'auditeur son propre pattern d'excitation en rapport avec sa propre f0 et la série d'harmoniques qui en découle.

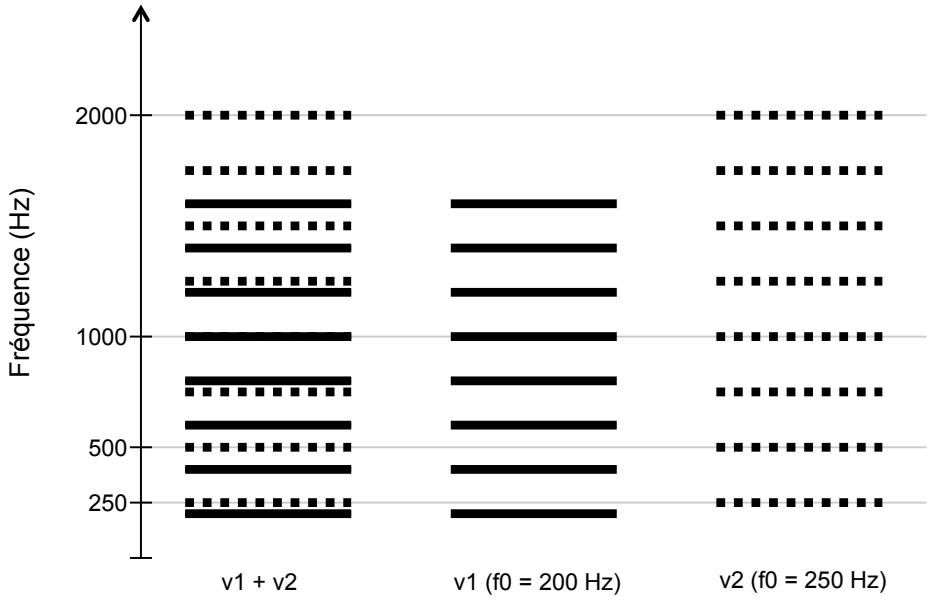


Figure 4. Les huit premières harmoniques de deux voyelles v1 (traits pleins) et v2 (traits pointillés) de f0 respectivement égales à 200 et 250 Hz sont représentées au centre et à droite de la figure. Si les deux voyelles v1 et v2 sont émises simultanément par deux locuteurs distincts (à gauche de la figure), leurs composantes spectrales se retrouvent d'une certaine façon entremêlées le long de la membrane basilaire de l'auditeur. Un des indices utilisable par le système auditif pour réattribuer correctement les composantes correspondant à chacune des sources est l'harmonicité. En effet, d'une manière générale, les composantes fréquentielles issues d'une même source sonore ont des relations harmoniques entre elles.

Les composantes spectrales des deux sources sonores seront représentées simultanément (Figure 4, à gauche) et d'une certaine manière entremêlées le long de la membrane basilaire. Un des mécanismes utilisable par le système auditif pour ségréguer les deux sources sonores en réattribuant à chacune d'elles ses composantes est la fusion harmonique. En effet, les composantes issues d'une même source ont généralement des relations harmoniques entre elles (Figure 4, au centre et à droite). Il faut noter qu'en réalité le mécanisme est plus complexe que ne le suggère la Figure 4, car si les harmoniques représentées son effectivement résolues lorsque chacune des voyelles est présentée isolément, elles ne le sont plus obligatoirement si on associe simultanément

les deux voyelles (des harmoniques issues de voyelles différentes pouvant se retrouver spectralement très proches).

Dans le cas des sons complexes périodiques, les phénomènes de fusion harmonique et d'extraction de la hauteur sont en fait intimement liés. La détection de rapports de fréquences harmoniques peut être considérée comme l'étape commune à deux phénomènes combinés : le regroupement perceptif en une seule unité auditive (*i.e.* la fusion harmonique proprement dite), et l'attribution d'une hauteur unique à cet ensemble de composantes fréquentielles (*i.e.* l'extraction de la hauteur). La psychoacoustique s'est principalement intéressée à cette question sous l'angle de la perception de la hauteur, mais ces données sont donc d'un grand intérêt pour étudier indirectement la fusion harmonique.

2.2.3 Données psychoacoustiques

Des renseignements intéressants sur les mécanismes qui sous-tendent la fusion harmonique peuvent être apportés par l'étude de stimuli dont la structure est modifiée par rapport à une stricte série harmonique. On peut considérer deux grands types de modifications spectrales et analyser leurs conséquences dans la perception du son complexe : (1) la suppression d'une ou plusieurs composantes du complexe, (2) le déplacement (c'est-à-dire la modification fréquentielle) d'une ou plusieurs composantes du complexe.

(a) Conséquences perceptives de la suppression de composantes d'un son complexe

Une propriété importante de la perception de la hauteur de sons complexes est connue sous le nom de *phénomène de la fondamentale absente*. Lorsque l'on supprime la composante de plus basse fréquence (f0) d'une série harmonique, la hauteur perçue par un auditeur sain reste f0 (Ohm, 1843; Schouten, 1940). Cela est vrai même en présence

d'un masquage rendant impossible la détection d'un éventuel produit de distorsion – composante produite par le système auditif périphérique lui-même– de fréquence f0 (Licklider, 1954; Patterson, 1969). Ces données indiquent que la perception de la hauteur des sons harmoniques ne repose pas sur la présence d'une excitation dans la zone de la membrane basilaire tonotopiquement dédiée à la fréquence fondamentale.

En fait, on peut supprimer plusieurs harmoniques de faible rang et plusieurs harmoniques de rang élevé sans modifier la perception globale de la hauteur. Cela a amené de nombreux auteurs à parler de *hauteur virtuelle* pour les sons complexes, illustrant le fait que la hauteur perçue ne correspondait pas nécessairement à une composante fréquentielle physiquement présente, par opposition à la *hauteur spectrale* des sons purs. Il suffit de conserver quelques harmoniques résolues pour que la hauteur du son complexe soit clairement perçue et reste égale à f0 ; l'auditeur percevra cette modification spectrale par un changement dans le timbre. Les six premières harmoniques, et particulièrement celles comprises entre le $2^{ème}$ et le $5^{ème}$ rangs semblent avoir un rôle prépondérant dans la détermination de la hauteur du complexe, et sont considérées comme la *région dominante* du complexe harmonique (Ritsma, 1962; Plomp, 1967; Moore, Glasberg, & Peters, 1985). Plus récemment, les travaux de Dai (2000) ont modifié cette notion en indiquant que les composantes voisines de 600 Hz étaient dominantes dans la détermination de la hauteur quel que soit leur rang dans le complexe. Par exemple, pour un son complexe de f0 égale à 100 Hz, la sixième harmonique était dominante, alors que pour un son complexe de f0 égale à 200 Hz, la troisième harmonique était dominante. Quoi qu'il en soit, il est clair que les harmoniques résolues sont prédominantes dans la détermination de la hauteur des sons complexes. Dans le cas où seules les harmoniques non résolues sont présentes, la hauteur perçue est plus ambigüe, moins saillante, témoignant du manque de précision dans le codage fréquentiel des composantes (Moore B. C., 2007).

(b) Conséquences perceptives du déplacement d'une ou plusieurs composantes d'un son complexe

Des composantes séparées par des rapports de fréquences parfaitement harmoniques fusionnent entre elles ; mais comment sera perçu un son complexe dont l'une des composantes est légèrement déplacée, c'est-à-dire dont la fréquence est un peu supérieure ou inférieure à sa valeur harmonique « correcte »? Moore, Peters & Glasberg (1985) ont mesuré la valeur minimale de déplacement d'une composante d'un son complexe nécessaire pour qu'un auditeur détecte une inharmonicité. Pour les harmoniques de faibles rangs (jusqu'au rangs 4 à 6), cette valeur est en moyenne égale à 1.1% de la fréquence de l'harmonique concernée. Les sujets rapportent que la composante déplacée semble être « exclue » du reste du complexe fusionné et forme un objet auditif distinct. Pour les harmoniques de rang plus élevé qui sont moins bien résolues, 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. Ce phénomène connu sous le nom de *battement de consonance imparfaite* (Plomp, 1967; Viemeister, Rickert, & Stellmack, 2001) constitue un indice différent qui empêche toute interprétation sur le mécanisme de fusion harmonique en tant que tel ; il s'agit là d'un point crucial dans toute étude portant sur la perception de l'harmonicité. Un des mécanismes possibles expliquant ces battements est présenté par la Figure 5. Considérons deux composantes sinusoïdales de fréquences respectives f1 et f2, séparées par un intervalle proche d'une octave. Même si ces composantes sont normalement correctement résolues par la cochlée, on peut concevoir qu'il existe au niveau de la membrane basilaire une zone où les deux patterns d'excitation (respectivement PE1 et PE2) se chevauchent, autrement dit une zone d'interaction (ZI) entre patterns adjacents, qui répond simultanément à deux sons purs. Quand l'intervalle est harmonique (dans notre exemple exactement égal à une octave), l'onde résultant de la combinaison des deux sons purs est une onde complexe périodique (et de période

1/f1), qui ne produit pas de sensation de battements (Figure 5A). En revanche, quand l'intervalle entre f1 et f2 est inharmonique (dans notre exemple légèrement supérieur à une octave), la zone de chevauchement est stimulée par une onde complexe dont la période est beaucoup plus longue (Figure 5B). Il apparaît une fluctuation dans la forme de l'onde qui est liée à un décalage de phase continu entre les deux composantes, et qui peut donner naissance à la perception de battements de consonance imparfaite.

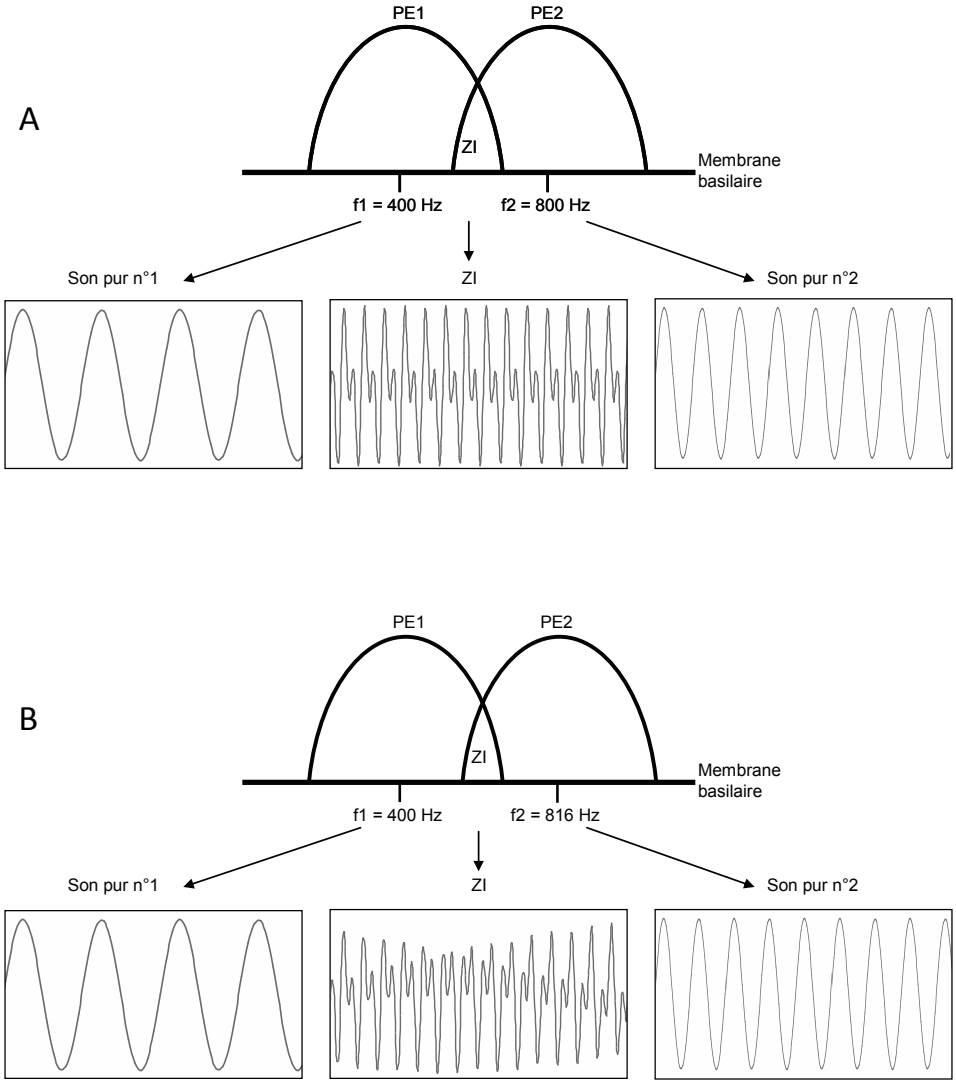


Figure 5. Représentation d'un mécanisme possible du phénomène de battements de consonance imparfaite. Le schéma représente les patterns d'excitation (PE1 et PE2) de deux sons purs de fréquences respectives f1 et f2, ainsi que la zone d'interaction (ZI) entre les deux patterns au niveau de la membrane basilaire. Les trois cadres contiennent une représentation de l'onde stimulant chacune des trois portions de la membrane basilaire correspondante (l'échelle de temps est différente pour le cadre ZI). **(A)** f2 = 2*f1. La ZI entre les deux patterns d'excitation PE1 et PE2 reçoit une onde complexe périodique issue de la combinaison des deux sons purs f1 et f2. Cette onde périodique n'entraîne pas de battements. **(B)** f2 = 2.04*f1. La ZI reçoit une onde complexe de période beaucoup plus longue car le rapport entre f1 et f2 n'est pas entier. Il existe un décalage continu de phase entre les deux composantes, responsable d'une fluctuation dans la forme de l'onde à l'origine de la sensation de battement.

Le déplacement d'une composante a également un impact sur la perception de la hauteur du son complexe dans son ensemble. Pour les harmoniques de faible rang, la hauteur du complexe varie dans le même sens et proportionnellement au déplacement d'une des composantes, et ce pour un déplacement allant jusqu'à 2 à 3% (Moore, Glasberg, & Peters, 1985) ; au-delà la composante n'a plus d'impact significatif sur la hauteur du complexe. Ceci semble indiquer qu'une composante peut encore contribuer à la hauteur d'un son complexe même si elle est considérée comme inharmonique par le système auditif (le seuil de détection d'inharmonicité étant globalement égal à 1.1%). Autrement dit, fusion harmonique et extraction de la hauteur correspondraient bien à deux mécanismes perceptifs distincts, même s'ils sont intriqués ; celui qui sous-tend la fusion harmonique semble être moins tolérant à des écarts minimes à l'harmonicité que celui qui sous-tend la perception de la hauteur.

Prolongeant cette idée d'une dissociation des deux mécanismes, Roberts & Brunstrom (2000) ont avancé l'hypothèse que le mécanisme de fusion serait principalement basé sur la détection d'une régularité dans le spectre, et serait possible même en l'absence de relations harmoniques entre les composantes. A l'inverse, l'extraction de la hauteur reposerait avant tout sur l'harmonicité, tout en tolérant l'écart de certaines composantes par rapport à leur valeur parfaitement harmonique. Partant du principe que plus un son complexe est perceptivement fusionné, plus il est facile d'en détecter une composante déplacée, Roberts & Brunstrom (2001) ont évalué le degré de fusion de deux types de sons inharmoniques présentant une certaine régularité spectrale. Les premiers correspondent à une série harmonique dont le spectre a été globalement « translaté » (par exemple une série harmonique de f0 égale à 200 Hz devient après translation de 25% : 250, 450, 650, 850 Hz, etc.). Les seconds correspondent à une série harmonique dont le spectre a été « étiré », c'est-à-dire que l'écart séparant deux composantes consécutives augmente progressivement avec le rang des composantes

(par exemple une série harmonique de f0 égale à 100 Hz devient après étirement de 3% : 100, 203, 309, 418 Hz, etc.). Leurs données indiquent qu'un son dont le spectre est translaté reste étonnamment fusionné (à un degré moindre cependant qu'un son harmonique), même pour des valeurs de translation atteignant 50% de f0. De même, un étirement modéré du spectre (inférieur à 3%) n'entraine pas de séparation perceptive des composantes. Ces résultats sont compatibles avec l'idée que le système auditif est sensible à une certaine régularité spectrale entre composantes adjacentes, et que l'intensité globale de la fusion d'un son complexe résulte de la concaténation d'analyses de portions successives le long du spectre. Roberts & Brunstrom (2003) ont néanmoins précisé dans une étude plus récente que ce mécanisme de regroupement basé sur la régularité spectrale n'était pas valable pour le premier composant (le composant de plus basse fréquence) d'un son complexe spectralement translaté. Dans ce cas, la fusion du composant 1 au reste du complexe semble très dépendante d'une relation parfaitement harmonique avec le composant 2.

2.2.4 Les grands modèles de perception de la hauteur et la fusion harmonique

(a) Les indices potentiels du codage de la hauteur

Les deux types de représentations (représentation spatiale reposant sur la détection d'un maximum d'excitation le long de la membrane basilaire et représentation temporelle liée au verrouillage de phase) proposées pour le codage de la hauteur des sons purs peuvent être naturellement appliquées aux composantes résolues d'un son complexe. Néanmoins, la présence de composantes fréquentielles simultanées qui interagissent au sein d'un même filtre auditif lorsqu'elles ne sont pas résolues, fait apparaître d'autres indices utilisables par le système auditif pour extraire la hauteur des sons complexes par rapport au cas plus simple des sons purs. L'interaction entre les

harmoniques non résolues (c'est-à-dire approximativement au-delà de la 8ème harmonique) entraine en sortie du filtre auditif une fluctuation d'enveloppe qui suit la fréquence fondamentale (cf. Figure 3E). Dans les fibres du nerf auditif dont les fréquences caractéristiques correspondent à ces composantes, les potentiels d'actions sont déclenchés préférentiellement par les pics les plus saillants de l'onde complexe. En d'autres termes, les intervalles entre potentiels d'actions seront le plus souvent égaux à la période de fluctuation de l'enveloppe (1/f0) ou à des multiples entiers de cette période. Le système auditif peut donc disposer d'une information temporelle sur la fréquence fondamentale par le biais du codage des harmoniques non résolues. Moore et al. (2003; 2006) ont montré que selon le rang des harmoniques non résolues, les auditeurs sains pouvaient probablement utiliser deux types d'informations différentes. Pour les harmoniques comprises entre le 8ème et le 12ème rang, les sujets semblent garder une sensibilité à la structure temporelle fine de l'onde complexe, c'est-à-dire aux fluctuations rapides de l'onde. Cela se traduit par une représentation relativement précise de la période de répétition de l'onde, et donc de la fréquence fondamentale du complexe. Au-delà du 12ème rang, la sensibilité à la structure fine disparait et seules les variations d'amplitude de l'enveloppe peuvent être exploitées, ce qui entraine une imprécision dans l'évaluation de l'intervalle séparant deux maxima locaux successifs, et donc une ambigüité dans la détermination de la fréquence fondamentale.

Schouten et al. (1962) ont proposé un modèle purement temporel du codage de la hauteur reposant sur ces informations issues des harmoniques non résolues. Même si ce modèle peut rendre compte de la perception de la hauteur d'un son complexe dont la fréquence fondamentale est absente, il n'est pas cohérent avec le fait que les harmoniques résolues jouent un rôle prépondérant dans la perception de la hauteur (Moore, Glasberg, & Peters, 1985). En effet, un son complexe ne contenant que des harmoniques résolues évoque une hauteur plus saillante et plus précise qu'un son

complexe ne contenant que des harmoniques non résolues. Il était nécessaire de proposer d'autres modèles prenant également en compte les informations issues des harmoniques résolues. Dans un souci de clarification, on peut schématiquement les regrouper en deux grandes catégories : les modèles de reconnaissance de pattern (où prédomine l'information spectrale) et les modèles d'autocorrélation (où prédomine l'information temporelle).

(b) Les modèles de reconnaissance de pattern

Les modèles utilisant un mécanisme de reconnaissance de pattern ont été proposés notamment par Goldstein (1973), Terhardt (1974), ou plus récemment Shamma & Klein (2000). Ils considèrent que l'extraction de la hauteur repose sur l'information portée par les harmoniques résolues du son complexe. Dans une première étape, le système auditif périphérique réalise une analyse spectrale individualisant les différentes harmoniques de bas rangs contenues dans le complexe. Notons que ce type de modèle admet tout aussi bien que les composantes fréquentielles soient individualisées sur la base d'un codage spatial du maximum d'excitation, que d'un codage temporel reposant sur le verrouillage de phase. Dans une seconde étape, le modèle suppose l'existence d'un mécanisme central capable d'analyser la répartition de ces harmoniques le long du spectre fréquentiel pour déterminer la valeur de la fréquence fondamentale. Ce mécanisme est souvent illustré par l'image du gabarit : pour reconnaitre une série harmonique et en extraire la hauteur, le système auditif central choisirait parmi une banque de gabarits internes celui dont les « repères » (régulièrement espacés sur une échelle linéaire de la fréquence) s'ajustent le mieux aux composantes fréquentielles individualisées. La fréquence fondamentale serait alors déterminée comme égale à la fréquence la plus basse du gabarit (ou égale à l'espace séparant les repères du gabarit), et cela même en l'absence de certaines composantes du complexe. La Figure 6

en donne une illustration schématique : le spectre d'un son complexe composé des harmoniques de rangs 3, 4 et 5 d'une fréquence fondamentale de 100 Hz absente est représenté en bas de la figure. Ces harmoniques correspondent parfaitement avec un

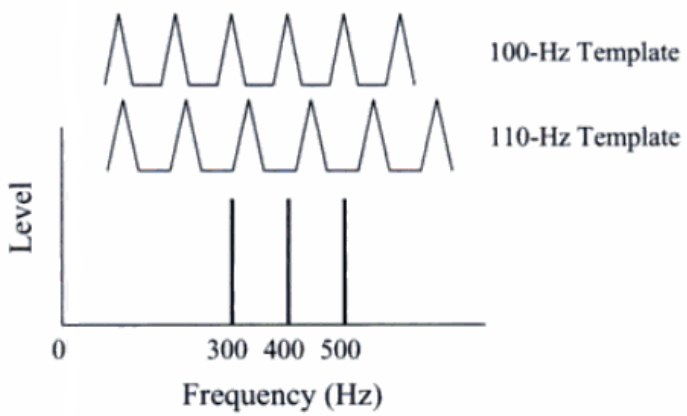


Figure 6. Représentation schématique d'un modèle spectral de reconnaissance de pattern. Le spectre d'un stimulus composé des trois harmoniques de rangs 3, 4 et 5 d'une fréquence fondamentale absente de 100 Hz est représenté dans la partie basse de la figure. Le modèle suppose l'existence de gabarits harmoniques au niveau du système auditif central ; les deux gabarits correspondant à des f0 respectives de 100 Hz et 110 Hz sont symbolisés par des lignes brisées en haut de la figure. Le gabarit « 100 Hz » s'ajuste parfaitement aux composantes du stimulus, contrairement au gabarit « 110 Hz ». Selon ce modèle, les composantes respectant les repères du gabarit seraient fusionnées entre elles en un son unique dont la hauteur perçue est celle de la plus basse fréquence acceptée par le gabarit choisi (ou, de manière équivalente, égale à l'écart entre les repères). D'après Plack (2005).

gabarit de 100 Hz tel qu'il est admis au niveau du système auditif central, alors que la concordance est moins bonne avec un gabarit différent, par exemple 110 Hz. Pour être fusionnée au reste du complexe, une composante fréquentielle donnée doit s'ajuster précisément au gabarit choisi, dont la tolérance théorique est de l'ordre de 1 % selon les données de Moore et al. (1985).

Plusieurs critiques ont néanmoins été opposées à ces modèles de reconnaissance de pattern. Tout d'abord, il n'a jamais été possible de mettre en évidence un support neuronal précis à ce type de gabarits. Ensuite, la manière dont ils sont générés au niveau du système auditif central fait également débat. L'hypothèse la plus communément admise est celle d'un apprentissage progressif par le système auditif sous l'influence de notre environnement acoustique naturel riche en sons complexes harmoniques (Terhardt, 1974). Shamma et Klein (2000) ont même développé un modèle utilisant une

banque de filtres cochléaires et un détecteur de coïncidence, dans lequel tout stimulus large bande (y compris du bruit) serait suffisant pour expliquer la formation de gabarits harmoniques. Cette hypothèse est cependant remise en question par des données suggérant que l'on peut mettre en évidence chez de très jeunes enfants (dès l'âge de trois mois) une sensibilité « innée » aux relations harmoniques entre composantes résolues (Demany & Armand, 1984; Clarkson & Rogers, 1995), même si un apprentissage au cours de la période fœtale et des tous premiers mois de vie ne peut pas être exclu. Un autre argument est avancé à l'encontre de ces modèles, qui ne considèrent que les harmoniques résolues dans la détermination de la hauteur. Même si, comme nous l'avons vu, les harmoniques non résolues produisent une sensation de hauteur moins saillante et plus ambiguë, elle n'en reste pas moins présente et permet par exemple la reconnaissance de mélodies simples (Houtsma & Smurzynski, 1990). Les modèles de reconnaissance de pattern ne peuvent pas rendre compte de ce type de données concernant l'extraction de la hauteur, et impliquent qu'il existe un mécanisme différent pour les harmoniques non résolues.

(c) Les modèles d'autocorrélation

Les modèles d'autocorrélation dans lesquels l'information temporelle joue un rôle prépondérant, présentent l'avantage de tenir compte à la fois des informations issues des harmoniques résolues (comme les modèles de reconnaissance de pattern) mais également des harmoniques non résolues (comme les modèles exclusivement temporels) en les intégrant dans un mécanisme unique d'extraction de la hauteur. Le premier modèle de ce type a été proposé par Licklider (1951), et a été enrichi par la suite notamment par Meddis & Hewitt (1991), Meddis et O'Mard (1997) ou Bernstein & Oxenham (2005).

Les Figures 7 et 8 illustrent les grands principes d'un modèle d'autocorrélation. Le stimulus entrant, dans cet exemple un son complexe harmonique de f0 égale à 200 Hz, est spectralement analysé par la cochlée assimilable à un ensemble de filtres passe-bande (Figure 7, à gauche et au centre). La réponse en sortie de filtre est représentée pour les quatre canaux centrés par les fréquences caractéristiques 200, 400, 800 et 1200 Hz (soient les harmoniques résolues de rangs 1, 2, 4 et 6), ainsi que pour le canal centré par la fréquence caractéristique 3200 Hz répondant à l'interaction de plusieurs harmoniques non résolues (Figure 7, à droite). Rappelons que les neurones dont la fréquence caractéristique se situe dans une zone spectrale résolue se verrouillent en

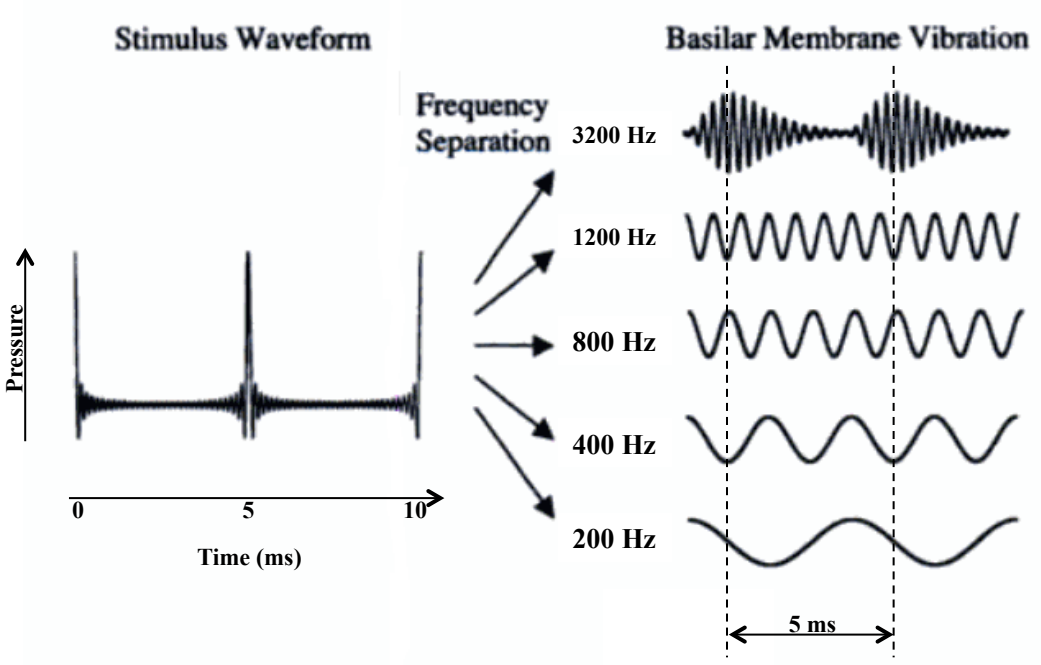


Figure 7. L'onde sonore d'un son complexe harmonique de f0 égale à 200 Hz est représentée dans la partie gauche de la figure. L'analyse spectrale réalisée par la cochlée est illustrée pour cinq filtres cochléaires, ou canaux fréquentiels, centrés par différentes fréquences caractéristiques (au centre de la figure) : quatre correspondent aux harmoniques résolues de rangs 1, 2, 4 et 6 (respectivement 200, 400, 800 et 1200 Hz), le cinquième canal centré par la fréquence 3200 Hz répond quant à lui à plusieurs harmoniques non résolues. La réponse de chaque canal est représentée sur la partie droite de la figure. L'intervalle de temps 5 ms (réciproque de la fréquence fondamentale 200 Hz) est indiqué par les lignes pointillées. Cet intervalle possède la particularité d'être représenté dans les réponses de tous les canaux. Il est égal à la période de l'onde sinusoïdale (ou à un multiple entier de cette période) dans les canaux répondant aux harmoniques résolues, et il est égal à la période de fluctuation d'amplitude de l'enveloppe dans les canaux répondant aux harmoniques non résolues. Les modèles d'autocorrélation reposent sur l'utilisation de cette information temporelle inter-canaux. D'après Plack (2005).

phase avec l'harmonique correspondante, alors que dans les zones non résolues ils se verrouillent en phase avec les fluctuations de l'enveloppe synchrones de f0. De ce fait, quelque soit le canal fréquentiel considéré, les différents patterns de réponse des neurones ont un point commun ; ils contiennent tous des intervalles entre potentiels d'action égaux à 5ms (réciproque de la fréquence fondamentale 200 Hz), même si d'autres intervalles sont présents (Figure 7, à droite). Le mécanisme d'autocorrélation fournit une solution pour expliquer comment le système auditif pourrait extraire et utiliser cette information. L'autocorrélation repose sur la comparaison d'un signal avec des copies progressivement retardées du signal lui-même. Lorsque ce retard est égal à un intervalle très représenté dans le signal, la corrélation entre le signal et les copies est forte. Appliquée à la réponse d'un groupe de neurones au sein d'un canal fréquentiel donné, l'autocorrélation permet d'obtenir un pattern d'activité (ou autocorrélogramme, Figure 8A) dans lequel les pics successifs codent directement la période de l'onde traitée

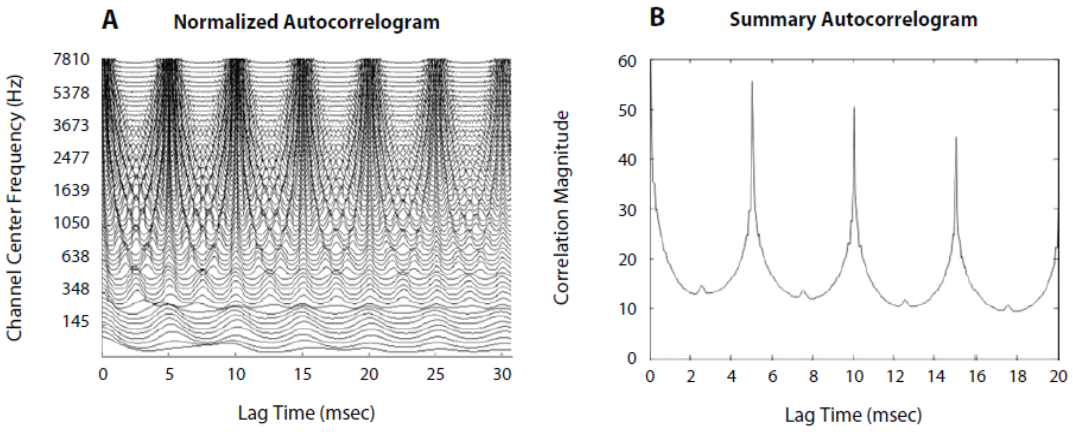


Figure 8. (A) Autocorrélogramme d'un son complexe harmonique de fréquence fondamentale égale à 200 Hz. Chaque ligne horizontale représente la fonction d'autocorrélation obtenue à partir du pattern de réponse neuronale dans un canal fréquentiel donné. On remarque graphiquement que pour les délais multiples de 5 ms (réciproque d'une fréquence de 200 Hz), il existe une forte corrélation entre les réponses de tous les canaux. **(B)** Fonction globale d'autocorrélation (SACF) obtenue par sommation des fonctions des différents canaux fréquentiels. Les pics saillants indiquent les délais où la corrélation inter-canaux est forte, toutes les 5 ms. Le système auditif pourrait utiliser cette information temporelle pour extraire la hauteur correspondante du complexe (200 Hz), à partir des harmoniques résolues et non résolues. D'après Yost (2009).

par ce canal. En sommant simplement les autocorrélogrammes des différents canaux fréquentiels, on obtient une fonction d'autocorrélation « globale » ou *summary autocorrelation function* (SACF) qui rend compte des intervalles communs issus des harmoniques résolues et non résolues (Figure 8B). Le premier pic visible sur cette fonction (qui est également le plus ample) correspond au délai de corrélation maximale entre tous les canaux (ici, 5 ms), et les pics suivants de forte corrélation sont répartis toutes les 5 ms. La SACF donne une représentation de l'indice qui pourrait servir de support à l'extraction de la hauteur des sons complexes.

Roberts & Brunstrom (2001) se sont appuyés sur ce type de modèle pour proposer un mécanisme possible de détection d'inharmonicité (Figure 9). Le stimulus considéré ici est un son complexe comprenant les 12 premières harmoniques d'une

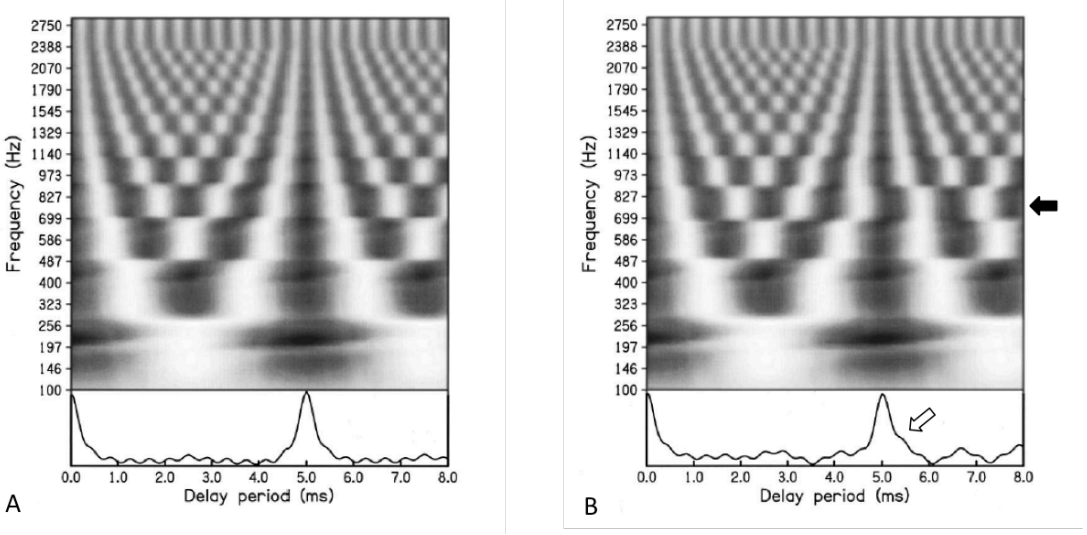


Figure 9. Chacun des deux panneaux comporte un autocorrélogramme (partie supérieure du panneau) et la SACF correspondante (partie inférieure). Sur l'autocorrélogramme, le degré de corrélation en fonction du délai est représenté par une échelle de gris (les zones les plus foncées correspondant à une forte corrélation). **(A)** Le stimulus analysé est un son complexe parfaitement harmonique comprenant les 12 premières composantes d'une fréquence fondamentale de 200 Hz. L'autocorrélogramme fait apparaître une « colonne » verticale correspondant à une forte corrélation inter-canaux pour un délai de 5 ms, retrouvée sous la forme d'un pic dans la SACF. **(B)** Stimulus identique au précédent à l'exception de la quatrième composante spectrale rendue inharmonique par un décalage fréquentiel de -4%. Cette inharmonicité fait apparaître une « encoche » dans la colonne verticale de l'autocorrélogramme (flèche noire), ainsi qu'une ondulation supplémentaire dans le pic de la SACF (flèche évidée), qui pourraient témoigner d'indices utilisés par le système auditif pour détecter l'inharmonicité de la composante et l'exclure perceptivement du complexe. D'après Roberts & Brunstrom (2001).

fréquence fondamentale de 200 Hz. L'autocorrélogramme et la SACF de ce stimulus sont représentés sur la Figure 9A. Il apparaît nettement sur l'autocorrélogramme ce que les auteurs décrivent comme une « colonne » verticale pour le délai de 5 ms (correspondant à la corrélation maximale), associée à un pic sur la SACF. Les conséquences du décalage fréquentiel de -4% de la quatrième composante fréquentielle (768 Hz au lieu de 800 Hz) sont représentées dans la Figure 9B. Deux modifications peuvent être notées : d'une part l'apparition d'une « encoche » dans la colonne verticale observée sur l'autocorrélogramme en regard du canal répondant à la composante déplacée, et d'autre part une modification de la forme du pic de la SACF (apparition d'une ondulation supplémentaire sur le versant descendant du pic). La traduction neuronale de ces indices n'est pas connue, néanmoins ils pourraient servir de support à la détection d'inharmonicité et à la ségrégation perceptive de la composante par rapport au reste du complexe.

Les modèles d'autocorrélation présentent l'avantage de rendre compte par un unique mécanisme de nombreux phénomènes observés dans les études psychoacoustiques, comme la perception de la fondamentale absente, ou l'extraction de la hauteur sur la seule base d'harmoniques non résolues. Néanmoins, certaines lacunes peuvent leur être opposées. Premièrement, comme dans le cas des modèles de reconnaissance de pattern, les structures neuronales supportant un mécanisme d'autocorrélation n'ont jamais été formellement identifiées. Deuxièmement, les modèles d'autocorrélation ont du mal à expliquer pourquoi les harmoniques résolues sont associées à une sensation de hauteur beaucoup plus précise et plus saillante que les harmoniques non résolues, ce qui se traduit par exemple par des meilleures performances de discrimination de fréquences fondamentales (Moore, Glasberg, & Shailer, 1984; Carlyon & Shackleton, 1994), même si certaines explications dont nous ne rendrons pas compte ici ont été avancées (de Cheveigné & Pressnitzer, 2006). Il faut

noter que d'autres types de modèles ne faisant pas intervenir d'autocorrélation, mais privilégiant également l'information temporelle ont été proposés (Moore B. C., 1982; de Cheveigné, 1993; Patterson, Allerhand, & Giguère, 1995).

Le débat passionné entre les modèles privilégiant l'information spectrale et ceux reposant sur le codage temporel est donc toujours d'actualité. Même si durant ces vingt dernières années, ces derniers ont eut les faveurs de nombreux auteurs, certaines données récentes remettent en avant l'importance de l'information spatiale dans le codage de la hauteur (Oxenham, Bernstein, & Penagos, 2004). Les hypothèses conciliant les deux théories en supposant que les harmoniques résolues et non résolues soient traitées par des mécanismes distincts reste également envisageable (Carlyon, 1998).

2.2.5 Données neurophysiologiques

Les études neurophysiologiques portant sur les sons complexes se sont souvent intéressées en priorité au substrat neuronal de la perception de la hauteur ou à la ségrégation des sources sonores. Néanmoins, certaines données nous renseignent sur les mécanismes potentiellement impliqués dans la fusion harmonique, notamment celles portant sur le codage de sons complexes présentés simultanément ou sur le codage d'une inharmonicité résultant du déplacement de composantes. Les approches multiples (enregistrement unitaires et multi-unitaires chez l'animal, potentiels évoqués auditifs, magnétoencéphalographie, imagerie fonctionnelle) apportent des informations sur différents niveaux du système auditif, depuis le nerf cochléaire jusqu'au cortex, et permettent d'évaluer la vraisemblance neurophysiologique des modèles établis à partir des données psychoacoustiques.

(a) Nerf auditif, noyau cochléaire et colliculus inférieur

Larsen, Cedolin & Delgutte (2008) ont comparé la précision de deux méthodes d'analyse des réponses du nerf auditif (chez le chat) dans l'estimation des f0s de deux sons complexes harmoniques présentés simultanément. La première méthode utilisait des informations de nature spectrale, basée sur le taux de décharge des fibres du nerf auditif selon leur fréquence caractéristique. La seconde méthode reposait sur des informations temporelles, à savoir la distribution des durées séparant les potentiels d'action. Les auteurs ont montré que les deux méthodes pouvaient fournir des estimations précises des f0s de sons harmoniques simultanés, même si la précision de la méthode temporelle était moins bonne pour les fréquences caractéristiques supérieures à 2-4 kHz, reflétant probablement la dégradation du verrouillage de phase dans les hautes fréquences. Ces données indiquent que les informations véhiculées par le nerf auditif sont suffisantes, aussi bien dans le domaine spectral que temporel, pour permettre au système auditif de ségréguer deux sons complexes harmoniques simultanés.

Des travaux plus anciens avaient porté sur le codage temporel au niveau du nerf auditif de cochons d'Inde (Palmer, 1988). Les stimuli utilisés étaient notamment des paires de voyelles de f0 proches (100 et 125 Hz) présentées simultanément. Les résultats montraient que selon la fréquence caractéristique des fibres, le pattern de réponse pouvait refléter soit l'une ou l'autre des f0s, soit des composantes harmoniques individualisées. Palmer avait estimé que l'information temporelle était suffisamment précise pour permettre une identification des deux f0s sans ambigüité, et servir de support à la ségrégation des deux voyelles. Keilson et al. (1997) avaient obtenu des résultats comparables au niveau du noyau cochléaire ventral du chat ; les décharges neuronales étaient synchronisées avec l'une ou l'autre des f0s de deux voyelles présentées simultanément, esquissant une base neurophysiologique à un modèle de type autocorrélation pour ce type de tâche.

Tramo et al. (2001) ont étudié la représentation temporelle de paires de sons complexes harmoniques simultanés au niveau du nerf auditif chez le chat. L'intervalle séparant les f0s des deux sons complexes était soit consonant, soit dissonant. Comme cela avait été observé avec les voyelles, leurs résultats indiquaient que pour les intervalles consonants, le pattern temporel pouvait se synchroniser avec l'une ou l'autre des f0s. En revanche, dans le cas des intervalles dissonants, la réponse du nerf suivait les modulations lentes de la forme de l'enveloppe observables à la sortie des filtres auditifs. Ces modulations, liées à l'interaction de composantes *inharmoniques* adjacentes au sein d'un même filtre auditif, se traduisent perceptivement par les battements de consonance imparfaite. Il faut les distinguer des modulations d'enveloppe rapides (synchrones de f0) observées pour les composantes non résolues des sons complexes *harmoniques*.

Les travaux de Sinex (2008) sur le codage de sons complexes harmoniques simultanés chez le chinchilla ont montré que le noyau cochléaire semblait comporter deux populations neuronales fonctionnellement distinctes. La première présentait des réponses synchronisées avec des composantes individuelles du son complexe, comme Sinex et al. (2003) l'avaient retrouvé au niveau du nerf auditif. La seconde au contraire répondait préférentiellement aux fluctuations lentes de l'enveloppe présentes en cas d'interaction cochléaire entre des composantes séparées par un intervalle dissonant. Sinex & Li (2007) avaient déjà mis en évidence ce type de fluctuations mais à des niveaux supérieurs du système auditif, notamment dans le colliculus inférieur. Cela a amené les auteurs à proposer l'hypothèse selon laquelle le noyau cochléaire était un relais important dans la différentiation des réponses entre stimuli harmoniques et inharmoniques. En amont de ce relais, la réponse neuronale suit principalement les composantes individuelles du complexe, que le stimulus soit harmonique ou non (même si, nous l'avons vu, des réponses synchrones des fluctuations lentes d'enveloppe ont également été observées au niveau du nerf auditif par d'autres auteurs). En aval, la

réponse reflète préférentiellement les fluctuations lentes d'enveloppe liées aux sons inharmoniques, entrainant une distinction qualitative marquée entre stimuli harmoniques et inharmoniques.

(b) Cortex

Pantev et al. (1996) ont recueilli l'activité magnétoencéphalographique en réponse à des sons complexes composés des harmoniques de rangs 4 à 11 d'une série de f0 égale à 250 Hz (cette fondamentale étant absente). Ces huit composantes étaient présentées soit monauralement, soit dichotiquement auquel cas l'oreille droite recevait un son complexe composé des quatre harmoniques impaires (1250, 1750, 2250 et 2750 Hz), l'oreille gauche des quatre harmoniques paires (1000, 1500, 2000 et 2500 Hz). Les auteurs ont vérifié que la hauteur perçue lorsque l'on présentait isolément l'un de ces sons complexes à quatre composantes était voisine de 500 Hz. Leurs données montraient que la même région du cortex auditif était activée en réponse aux stimuli monoraux (de f0 = 250 Hz) et aux stimuli dichotiques décrits ci-dessus. Cette région corticale était celle qui répondait également préférentiellement à un son pur de 250 Hz. Ces résultats semblaient apporter la preuve que (1) la même région corticale était activée en réponse à des stimuli de hauteur spectrale ou de hauteur virtuelle égale à 250 Hz, et que (2) la fusion harmonique binaurale mise en évidence depuis longtemps par des expériences psychoacoustiques (Cramer & Huggins, 1958) avait un substrat neurophysiologique.

Plus récemment, Bendor & Wang (2005) ont mis en évidence par des enregistrements multi-unitaires chez le singe une région corticale contenant des neurones qui semblaient coder spécifiquement la hauteur. Ils se sont intéressés à une population de neurones situés à proximité du bord antérolatéral du cortex auditif primaire, dans une région tonotopiquement dédiée aux basses fréquences. Ces neurones

répondaient aussi bien à des sons purs qu'à des sons complexes harmoniques de même hauteur, même en l'absence de la fréquence fondamentale. Pourtant, ces mêmes neurones ne répondaient à aucune des composantes fréquentielles du son complexe si elles étaient présentées isolément. De plus, la région du cortex mise en évidence ici chez le singe correspondait anatomiquement chez l'homme à la portion latérale du gyrus de Heschl qui apparaissait spécifique du codage de la hauteur selon certaines études d'imagerie fonctionnelle humaine (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Penagos, Melcher, & Oxenham, 2004). Ces travaux laissaient entrevoir une avancée déterminante dans nos connaissances sur la perception auditive, à savoir la découverte d'une zone corticale précise spécifiquement dédiée au codage de la hauteur en tant que telle, indépendamment des caractéristiques spectrales des stimuli.

Ils ont néanmoins étaient remis en question, notamment par Hall & Plack (2009). Ces auteurs ont avancé l'hypothèse que l'activité mise en évidence chez l'homme au niveau de la portion latérale du gyrus de Heschl était liée avant tout à la nature des stimuli utilisés dans la majorité des études d'imagerie fonctionnelle (l'I.R.N. pour *iterated-ripple noise*), et ne témoignait pas directement du codage de la hauteur du stimulus. Ils ont étudié par IRM fonctionnelle la réponse à différent type de stimuli évoquant perceptivement une hauteur de 200 Hz (dont l'I.R.N. mais également des sons purs ou des sons complexes avec harmoniques résolues et non résolues). Leurs résultats démontraient que cette zone supposée « spécifique de la hauteur » semblait d'une part beaucoup plus étendue et plus postérieure que la portion latérale du gyrus de Heschl, et que d'autre part la variabilité interindividuelle était très élevée. Chez certains sujets elle était même située dans des zones corticales qui ne sont pas classiquement considérées comme faisant partie du cortex auditif.

D'autres études se sont intéressées non pas au traitement cortical de la hauteur, mais à celui de la ségrégation perceptive d'un stimulus en plusieurs « objets » auditifs.

Alain et al. (2002) ont recueilli des potentiels évoqués auditifs en réponse à des sons complexes harmoniques dont l'une des composantes était spectralement déplacée. Leurs résultats montraient que lorsque l'inharmonicité était suffisante pour entrainer l'exclusion perceptive de la composante par rapport au reste du complexe, une onde négative tardive (dénommée ORN pour *object-related negativity*) apparaissait sur les potentiels évoqués auditifs. Cette onde, retrouvée aussi bien dans des conditions d'écoute active que passive, serait pour les auteurs le marqueur d'un mécanisme automatique de détection d'une inharmonicité par le système auditif central en cas d'inadéquation entre le stimulus perçu et les gabarits internes.

La détection d'inharmonicité liée au déplacement d'une des composantes d'une série harmonique a également été étudiée par des enregistrements multi-unitaires du cortex auditif primaire chez le singe par Fishman & Steinschneider (2010). Leurs résultats mettaient en évidence deux mécanismes neuronaux complémentaires pouvant témoigner de « l'exclusion » perceptive (ou *pop out*) de la composante inharmonique par rapport au reste du complexe. Premièrement, ils ont observé une augmentation locale (dans la zone corticale tonotopiquement dédiée à la fréquence de la composante déplacée) du taux de décharge neuronale en réponse à des stimuli inharmoniques par rapport aux stimuli harmoniques. Celle-ci était d'autant plus importante que la composante était déplacée, ce qui était cohérent avec les données psychoacoustiques humaines sur le *pop out* perceptif. Deuxièmement, il existait une modification du pattern temporel de la réponse des neurones codant préférentiellement la fréquence de cette composante ; la réponse corticale se synchronisait aux fluctuations lentes de l'enveloppe –ou battements– présentes en cas de stimulus inharmonique, comme cela a été observé au niveau du nerf auditif, du noyau cochléaire et du colliculus inférieur (voir ci-dessus). Néanmoins, ces résultats pourraient refléter avant tout la saillance perceptive ou *pop out* de la composante déplacée, plus que le mécanisme de détection d'inharmonicité lui-

même. Les auteurs ont d'ailleurs souligné qu'avec des stimuli inharmoniques dont le spectre était globalement étiré –similaires à ceux utilisés par Roberts & Brunstrom (2001)– et qui ne sont pas associés à un *pop out* perceptif, l'augmentation du taux de décharge neuronale n'était plus observée. Complétant ces données, Kalluri et al. (2008) ont étudié plus directement les mécanismes neuronaux sous-tendant la détection d'inharmonicité (en l'absence de *pop out*) par des enregistrements unitaires de neurones du cortex auditif primaire et de l'aire auditive antérieure chez le furet. Les stimuli utilisés étaient des sons complexes harmoniques (les rapports successifs entres composantes étant par définition 2, 3/2, 4/3, 5/4, etc.) ou inharmoniques (et dans ce cas les même rapports étaient utilisés mais dans un ordre aléatoire). Les sons étaient construits de manière à contenir une composante commune, de fréquence égale à la fréquence caractéristique du neurone étudié ; les différences dans le pattern de réponse du neurone pouvaient donc être attribuées au « contexte » harmonique ou inharmonique entourant cette composante. Malheureusement, cette étude n'a pas permis de mettre en évidence un marqueur neurophysiologique robuste de la détection d'inharmonicité. Elle retrouvait néanmoins que 10 à 15% des neurones étudiés répondaient différemment selon la nature harmonique ou inharmonique du stimulus, et pouvaient donc servir de support au codage de la fusion harmonique.

(c) Synthèse des données neurophysiologiques actuelles

- Les données neurophysiologiques n'ont pas permis pour l'instant de déterminer lequel des deux grands types de modèle (reconnaissance de pattern ou autocorrélation) était prédominant dans le codage de la hauteur et la fusion harmonique. Au contraire, au niveau du nerf auditif en tout cas, des informations spectrales et temporelles précises sont disponibles et peuvent servir de support neurophysiologique crédible aux deux types de modèles. Des réseaux neuronaux supportant un modèle

reposant sur des « gabarits » sont envisageables, par exemple en combinant l'activité de plusieurs neurones dont les fréquences caractéristiques sont harmoniquement liées (Plack, 2005). Les modèles de type autocorrélation quant à eux peuvent théoriquement être supportés par des réseaux de neurones comportant des lignes parallèles dont les délais de traitement sont différents. Des structures réalisant une analyse temporelle comparable, mais dédiées à la comparaison du délai inter-aural dans des tâches de localisation spatiale ont été décrites chez les mammifères au niveau du colliculus inférieur (McAlpine, Jiang, & Palmer, 2001). En revanche, dans le domaine de la perception de la hauteur, de tels réseaux de neurones n'ont pas pu être clairement identifiés pour l'instant.

- Il n'y a pas à ce jour de certitude sur la localisation précise au sein du système nerveux de la structure mise en jeu dans le codage de la hauteur. On sait depuis longtemps qu'elle se situe à un niveau capable de traiter des informations provenant des deux oreilles, puisqu'un sujet peut extraire la hauteur à partir de deux harmoniques ne comprenant pas la fréquence fondamentale, chacune étant présentée à une oreille (Houtsma & Goldstein, 1972). La mise en évidence d'une zone corticale dédiée à la perception de la hauteur fait encore débat.

- Au niveau du nerf et des premiers relais du système auditif, d'une manière générale, les stimuli dissonants se traduisent par une activité neuronale plus soutenue, des patterns de réponse plus complexes et qui présentent moins de cohérence entre les différents canaux auditifs que les stimuli consonants (Micheyl & Oxenham, 2010).

- Concernant la fusion harmonique proprement dite, il n'existe pas non plus de données suffisantes pour établir par quels mécanismes neuronaux et à quel niveau du système auditif elle s'établit. Deux types de contraintes peuvent expliquer les limites actuelles des connaissances de ce phénomène : (1) Dès que l'on utilise des stimuli harmoniques, l'intrication profonde entre fusion harmonique et extraction de la hauteur

complique l'interprétation des résultats, et il est en pratique difficile d'étudier le phénomène de fusion isolément. (2) Dans le cas de stimuli inharmoniques, les interactions périphériques entre composantes génèrent des fluctuations lentes de l'enveloppe, perçues comme des battements. On retrouve des réponses neuronales synchrones à ces battements à de multiples niveaux des voies auditives (nerf auditif, noyau cochléaire, colliculus inférieur, cortex), qui peuvent potentiellement renseigner le système auditif sur la présence d'une inharmonicité. La question qui reste en suspend est de savoir si cette information est réellement utilisée comme indice pour fusionner les composantes qui n'entrainent pas de tels battements (Micheyl & Oxenham, 2010), ou si elle ne représente qu'un « artefact » neuronal lié aux propriétés physiques des sons inharmoniques. Ces battements représentent un facteur de confusion puisqu'ils constituent un indice de discrimination entre consonance et dissonance différent du mécanisme de fusion harmonique proprement dit.

3 LA DISCRIMINATION AUDITIVE DE RAPPORTS DE FREQUENCES

3.1 Introduction

Si un auditeur sain est capable de détecter que les composantes fréquentielles d'un son complexe sont harmoniquement liées pour les fusionner entre elles, c'est que le système auditif est sensible aux rapports de fréquences entre ces composantes. Plus précisément, il est sensible à la *simplicité* des rapports de fréquences (un rapport étant dit *simple* s'il peut s'écrire comme le quotient de deux petits nombres entiers). A l'extrême, un rapport de 2/1 définit l'octave, deux sons purs étant séparés par une octave quand la fréquence de l'un est le double de celle de l'autre. La sensibilité à l'harmonicité dépend directement de la capacité à discriminer des rapports de fréquences.

Nous l'avons vu, les mécanismes de la fusion harmonique restent en grande partie inconnus, alors même que cet aspect de la perception auditive joue probablement un rôle majeur dans l'analyse de notre environnement sonore quotidien. En particulier, la nature de l'information –spectrale ou temporelle– utilisée préférentiellement par le système auditif pour fusionner les harmoniques n'a pas pu être formellement identifiée. Un des modèles spectraux de référence a été proposé par Terhardt (1974). Il indique que la fusion harmonique s'effectue selon deux étapes successives. Premièrement, le système auditif extrait la hauteur « spectrale » de chacune des composantes du son complexe, comme s'il s'agissait de sons purs présentés isolément. Deuxièmement, les intervalles séparant ces hauteurs sont comparées à des gabarits internes (acquis par l'exposition sonore naturelle riche en sons complexes harmoniques) ; toutes les

composantes qui s'ajustent correctement au gabarit choisi sont fusionnées entre elles en un percept unique de hauteur « virtuelle » f0. Selon Terhardt, ces mêmes gabarits seraient utilisés pour analyser les relations *mélodiques* entre sons purs successifs. Cela expliquerait par exemple pourquoi deux sons purs consécutifs séparés par une octave présentent une affinité perceptive particulière. De manière parfaitement superposable à ce que suppose le modèle dans le cas d'un son complexe, le système auditif pourrait extraire la hauteur de chacun des sons purs successifs, puis évaluer leur correspondance avec les gabarits harmoniques à sa disposition.

L'expérience présentée ci-après cherchait à clarifier les mécanismes qui sous-tendent la fusion harmonique par le biais d'une tâche de discrimination de rapports de fréquences. Les stimuli utilisés étaient formés de quatre sons purs, équidistants en termes de rapport de fréquence. La tâche des sujets consistait à discriminer un rapport de fréquence étalon (égal à 0.97 octave), d'un rapport cible légèrement plus grand (compris entre 0.98 et 1.04 octave). Les sons purs pouvaient être présentés simultanément (condition *Simul*) ou consécutivement (condition *Consec*). Les résultats montrent que dans la condition *Simul*, la performance des sujets ne suit pas une fonction monotone du rapport cible : si l'on augmente graduellement le rapport cible de 0.98 à 1.04 octave, la performance augmente initialement (jusqu'au rapport 1.0 octave), puis diminue pour des rapports légèrement supérieurs à 1.0 octave, et augmente de nouveau. A l'inverse, dans la condition *Consec* la performance augmente quasi linéairement avec le rapport cible sans effet particulier de la simplicité du rapport, ce qui est cohérent avec certaines données déjà rapportées dans la littérature. Par exemple Dobbins & Cuddy (1982), dans une étude portant sur la perception de l'octave mélodique, ont pu montrer que le rapport de fréquence physique et son corrélat perceptif étaient liés par une fonction monotone : à partir du moment où un sujet détecte une différence entre deux rapports de fréquences, il est capable de les ordonner sur une échelle de grandeur. Cela

ne semble pas être le cas dans la condition *Simul* où nos résultats indiquent que l'octave est associée à un statut perceptif particulier, et reconnue par un mécanisme différent capable de discriminer un très faible écart à l'octave, sans pour autant préciser le sens de cet écart. En d'autres termes, l'auditeur détecte qu'un rapport de fréquences n'est pas parfaitement harmonique, mais il ne peut pas dire s'il est trop grand ou trop petit. Nos résultats, comme ceux de Demany & Semal (1990), vont à l'encontre des propriétés théoriques des gabarits harmoniques tels qu'ils sont proposés par le modèle de Terhardt ; l'utilisation des mêmes gabarits pour reconnaitre l'octave harmonique (condition *Simul*) et l'octave mélodique (condition *Consec*) aurait du aboutir à des résultats équivalents dans les deux conditions. Ils semblent en revanche compatibles avec un mécanisme détectant simplement la présence (ou l'absence) d'une concordance temporelle parfaite entre les différents canaux fréquentiels contenant des sons purs distants d'une octave physiquement exacte.

3.2 Article "Auditory Discrimination of Frequency Ratios: The Octave Singularity"

Cette section présente l'article issu de cette expérience, associée à une expérience contrôle visant à écarter l'hypothèse que ces résultats puissent s'expliquer simplement par la présence de battements de consonance imparfaite. Nous avons choisi de respecter l'intégralité du contenu et de la mise en forme de l'article, tel qu'il a été publié en langue anglaise dans Journal of *Experimental Psychology: Human Perception and Performance* (Bonnard, Micheyl, Semal, Dauman, & Demany, 2013), à l'exception de l'appendice concernant la modélisation mathématique réalisé par le deuxième auteur (Micheyl, C.) et qui n'est pas le résultat direct de mon travail.

Auditory Discrimination of Frequency Ratios: The Octave Singularity

Damien Bonnard
Université de Bordeaux and CNRS

Christophe Micheyl
University of Minnesota

Catherine Semal, René Dauman, and Laurent Demany
Université de Bordeaux and CNRS

Sensitivity to frequency ratios is essential for the perceptual processing of complex sounds and the appreciation of music. This study assessed the effect of ratio simplicity on ratio discrimination for pure tones presented either simultaneously or sequentially. Each stimulus consisted of four 100-ms pure tones, equally spaced in terms of frequency ratio and presented at a low intensity to limit interactions in the auditory periphery. Listeners had to discriminate between a reference frequency ratio of 0.97 octave (about 1.96:1) and target frequency ratios, which were larger than the reference. In the simultaneous condition, the obtained psychometric functions were nonmonotonic: as the target frequency ratio increased from 0.98 octave to 1.04 octaves, discrimination performance initially increased, then decreased, and then increased again; performance was better when the target was exactly one octave (2:1) than when the target was slightly larger. In the sequential condition, by contrast, the psychometric functions were monotonic and there was no effect of frequency ratio simplicity. A control experiment verified that the nonmonotonicity observed in the simultaneous condition did not originate from peripheral interactions between the tones. Our results indicate that simultaneous octaves are recognized as "special" frequency intervals by a mechanism that is insensitive to the sign (positive or negative) of deviations from the octave, whereas this is apparently not the case for sequential octaves.

Keywords: spectral fusion, musical intervals, octave, harmony, melody

Human listeners are, to some extent, sensitive to the frequency ratios formed by pure tones. For pure tones presented *sequentially*, this is shown by the fact that a melody retains its perceptual identity when it is transposed in the frequency domain, that is, when the frequencies of the tones are multiplied by a common factor (Attneave & Olson, 1971). Moreover, a familiar melody is easier to recognize when it is presented with the correct frequency ratios than when the ratios are substantially distorted, even if, in the latter case, the global melodic contour is preserved (Dowling & Fujitani, 1971). For pure tones presented *simultaneously*, a well-known perceptual effect which seems to indicate that listeners are also sensitive to frequency ratios is the phenomenon of "harmonic fusion." A sum of simultaneous pure tones with frequencies forming a harmonic series (f, $2f$, $3f$, $4f$, etc.) is perceived as a single sound, with a pitch corresponding to f, rather than as a mixture of tones with different pitches. If one component of a harmonic series is mistuned by a few percent from the "correct" frequency, then this component pops out perceptually, while the other components remain fused (Moore, Peters, & Glasberg, 1986; Hartmann, McAdams & Smith, 1990; Lin & Hartmann, 1998; Roberts & Brunstrom, 1998). Because the components of a harmonic series are related by *simple* frequency ratios (i.e., ratios of small whole numbers), the phenomenon of harmonic fusion reveals a sensitivity to the simplicity of frequency ratios of simultaneous tones. However, in the case of sequentially presented tones, there is also evidence that people are, to some extent, sensitive to frequency-ratio simplicity. In particular, several studies have suggested that successive pure tones that are one octave apart and thus form a frequency ratio of 2:1 are perceptually similar (see, e.g., Deutsch, 1973; Idson & Massaro, 1978; Demany & Armand, 1984).

The present study is an investigation of human listeners' ability to detect slight differences between the frequency ratios formed by pure tones presented either simultaneously or sequentially. We wished to clarify the mechanisms underlying ratio discrimination in both cases. More specifically, we were especially interested in the influence of ratio simplicity on ratio discrimination. With regard to this influence, three hypotheses are a priori possible; they are hereafter referred to as H0, H1, and H2. According to H0, even though listeners are sensitive to frequency-ratio simplicity, as indicated by the evidence mentioned above, the discriminability of frequency ratios does not depend on their simplicity. Under this hypothesis, the discriminability of two ratios is expected to be a monotonic function of their difference or their ratio (Levelt, van de Geer & Plomp, 1966; Plomp, Wagenaar & Mimpen, 1973). Contrary to H0, H1 posits that frequency-ratio discriminability is

This article was published Online First October 22, 2012.

Damien Bonnard, Catherine Semal, René Dauman, and Laurent Demany, INCIA, Université de Bordeaux and CNRS, Bordeaux, France; Christophe Micheyl, Department of Psychology, University of Minnesota.

Damien Bonnard was supported by a grant from the Agence Régionale de Santé d'Aquitaine. Christophe Micheyl was supported by National Institutes of Health grant R01 DC05216.

Correspondence concerning this article should be addressed to Laurent Demany, INCIA, Université de Bordeaux and CNRS, BP 63, 146 rue Leo-Saignat, F-33076 Bordeaux Cedex, France. E-mail: laurent.demany@u-bordeaux2.fr

influenced by ratio simplicity. As a result, discrimination thresholds may vary nonmonotically as a function of ratio magnitude. However, both H0 and H1 posit that ratio discrimination takes place along a single, unidimensional perceptual continuum on which ratio magnitude is represented in an orderly fashion, with "small" ratios (i.e., ratios close to 1) at one end and large ratios at the other end. This means that for any set of three ratios, *r1*, *r2*, and *r3*, such that *r1* *r2* *r3*, it will always be easier to discriminate *r1* from *r3* than to discriminate *r1* from *r2* or *r2* from *r3*. According to H2, on the other hand, ratio simplicity has a deeper impact on ratio discrimination because differences in ratio simplicity can be by themselves effective discrimination cues, independently of—and in addition to—differences in ratio magnitude. Thus, H2 posits that ratio discrimination takes place along two distinct perceptual continua representing simplicity and magnitude, rather than along a single continuum representing magnitude. As explained below, if this hypothesis is correct, then one might find that for three ratios, *r1*, *r2*, and *r3*, such that *r1* *r2* *r3*, discriminating *r1* from *r3* is more *difficult* than discriminating *r1* from *r2*. In other words, H2 implies that comparative judgments of frequency ratios may not always obey the transitivity principle.

The effect of ratio simplicity on ratio discrimination for *simultaneously* presented pure tones has been investigated in a few previous studies (Viemeister & Fantini, 1987; Demany & Semal, 1988, 1992; Schellenberg & Trehub, 1996b; Schellenberg, 2002; Trainor, 1997; Stoelinga & Lutfi, 2010). All of these studies indicate that, for simultaneous tones, discriminating two frequency ratios is often much easier when the standard ratio is simple (so that the stimulus is harmonic) than when this is not the case. The results are therefore globally inconsistent with hypothesis H0; they are consistent with both H1 and H2. However, the simplicity effects observed in most of these studies can in fact be interpreted in a rather trivial way. Because of the limited frequency resolution of the cochlea (see, e.g., Moore, 2003, chap. 3), two simultaneous pure tones that are presented to the same ear and that are not both soft and well-separated in frequency will interact in the cochlea. As a result, some auditory-nerve fibers will respond to both tones and will encode properties of the global waveform. If the frequency ratio of the two tones is simple, for example, 2:1 or 3:2, then the global waveform has a short periodicity, in the pitch range. If, on the other hand, the frequency ratio deviates slightly from a simple value such as 2:1 or 3:2, then the global waveform has a much longer periodicity, which may be audible as a periodic beat (Plomp, 1967, 1976; Viemeister, Rickert & Stellmack, 2001).[1] This provides a cue for the discrimination between simple ratios and ratios close to a simple value.

The cue in question was probably available in most of the studies quoted above. However, Demany and Semal (1988, 1992) endeavored to eliminate it by presenting the two components of each stimulus dichotically and at a low SPL, namely 45 dB. The present research was again specifically concerned with the auditory processing of frequency relations *across* peripheral frequency channels, rather than with the temporal analysis of information available *within* peripheral frequency channels. Therefore, care was taken to minimize peripheral interactions between the tones and the audibility of beats resulting from such interactions. In addition, a control experiment was performed to check that beats of that kind were not audible.

For pure tones presented *sequentially*, previous studies have yielded conflicting results concerning the effect of frequency-ratio simplicity on ratio discrimination. Houtsma (1968) measured discrimination thresholds for ratios in the vicinity of one octave. One of the three listeners whom he tested had extensive musical training and all listeners received extensive training in the discrimination task before data collection. The results of Houtsma's study showed no significant variation in thresholds (expressed on a logarithmic scale, i.e., in octaves or in semitones; 1 octave 12 semitones) across ratios ranging from 1.9 (about 11 semitones) to 2.1 (about 13 semitones). Burns and Ward (1978) performed similar experiments, but they used a different range of frequency ratios (2.5–5.5 semitones) and they compared the thresholds of musicians and nonmusicians. In the nonmusical group, they obtained results akin to those of Houtsma, in that thresholds were approximately independent of the standard ratio; ratio simplicity had no effect. In the musical group, on the other hand, performance was markedly better than for nonmusicians overall and it was systematically *worse* for ratios very close to simple values (4:3, 5:4, or 6:5) than for ratios roughly halfway between these simple values. The latter finding was interpreted as evidence for categorical perception of musical intervals based on musical acculturation and practice. The authors concluded that "there is no evidence for the existence of natural perceptual boundaries for melodic musical intervals" (Burns & Ward, 1978, p. 466). In the same vein and in a broader perspective, McDermott, Keebler, Micheyl, and Oxenham (2010) have claimed that, contrary to a common belief, melodic intervals are not perceived more precisely than, and differently from, relations between successive sounds differing in loudness or brightness of timbre. However, the findings of Burns and Ward (1978) and Houtsma (1968) are at odds with results reported more recently by Schellenberg and Trehub (1994, 1996a, 1996b) and Schellenberg (2002). In these four studies, it was found that discriminating a melodic interval corresponding to a simple frequency ratio (3:2 or 4:3) from a somewhat smaller or larger interval (corresponding to a complex ratio) is easier when the simple ratio is presented before the complex ratio than when the two ratios are presented in the reverse order. This asymmetry was observed not only in musically educated adults (Schellenberg, 2002) but also in 6-year-old children (Schellenberg & Trehub, 1996a) and 6-month-old infants (Schellenberg & Trehub, 1996b). Even younger infants appear to perceive two melodies as similar when they differ from each other by octave transpositions of some of their components (Demany & Armand, 1984). The results of Schellenberg and Trehub (1996b) and Demany and Armand (1984) clearly suggest that some melodic intervals are more "natural" than others, contrary to the conclusion drawn by Burns and Ward (1978).

In sum, none of the three hypotheses labeled above as H0, H1, and H2 seems to be definitely ruled out by the current experimental evidence relating to the discrimination of melodic intervals. Nevertheless, the idea that, in the melodic domain, frequency-ratio discriminability could depend on ratio simplicity per se is chal-

[1] Such beats are called "beats of mistuned consonances." They differ from the better-known beats produced by pairs of pure tones with small frequency differences. However, for the sake of concision, they will be most often referred to simply as "beats" in the following.

lenged by the fact that the simplest frequency ratio formed by nonidentical successive tones, that is, 2:1, can be perceived as mistuned: when listeners are required to adjust two successive pure tones one octave apart, the obtained frequency ratio is often slightly but significantly larger than 2:1, especially for high-frequency tones (Ward, 1954; Walliser, 1969: Sundberg & Lindqvist, 1973; Demany & Semal, 1990; Hartmann, 1993). The neural basis of this "octave enlargement" effect is unclear, although tentative explanations have been proposed (Terhardt, 1974; Ohgushi, 1983; Hartmann, 1993; McKinney & Delgutte, 1999).

The current study stemmed from introspective observations that we made about the perception of stimuli consisting of four synchronous pure tones presented at a low SPL and regularly spaced on a logarithmic frequency scale. Each stimulus was defined by two parameters: *Fmin*, the frequency of the lowest tone, and *Oct*, the frequency spacing of the tones expressed in octaves. In our stimulus set, *Oct* varied in the range 0.9–1.1. We focused our attention on the perceptual correlate of changes in *Oct* while *Fmin* was varied concomitantly and randomly, in a wide range. Because of the randomness of the variations in *Fmin*, there was essentially no correlation between *Oct* and the global pitch of the stimuli or the frequency of individual tones. We first observed that the stimuli obtained for two values of *Oct* equidistant from 1, and not too close from each other (e.g., 0.9 and 1.1), were discriminable on the basis of a spectral "compactness" cue. This perceptual variable was a monotonic function of *Oct*: a stimulus with a relatively small value of *Oct*, for example, 0.9, sounded more compact than a stimulus with a relatively large value of *Oct*, for example, 1.1. It was also apparent, however, that a different perceptual variable, namely spectral fusion, was more efficient for the detection of changes from *Oct* 1 to *Oct* 1. For *Oct* 1, the four tones were perceptually fused into a single sound; this fusion decreased as *Oct* deviated from 1. Because fusion decreased for both positive and negative deviations from 1, this cue was not a monotonic function of *Oct*, in contrast to the compactness cue. Overall, therefore, our informal observations suggested that within a small range of *Oct* values centered on 1, changes in *Oct* were detectable on the basis of two different perceptual variables, each of them being more efficient than the other in certain conditions. This fitted with the hypothesis identified above as H2.

We reasoned that the just-described stimuli could be used to test objectively H2 as follows. Let $Oct_{\text{reference}}$ be an *Oct* value slightly smaller than 1 and moderately discriminable from 1. Suppose that, as suggested by our introspective observations, (1) the discrimination between $Oct_{\text{reference}}$ and 1 essentially rests on the fusion cue, and (2) fusion decreases as soon as *Oct* exceeds 1. It can then be expected that $Oct_{\text{reference}}$ will be easier to discriminate from 1 than from *Oct* values slightly larger than 1. However, even larger *Oct* values will be discriminable from $Oct_{\text{reference}}$ on the basis of the compactness cue. So, the psychometric function obtained for the discrimination between $Oct_{\text{reference}}$ and progressively larger values of *Oct* might show an initial rise followed by a fall and then a second rise. Such a nonmonotonic psychometric function would provide strong evidence for H2.

This was the rationale underlying the first half of Experiment 1, our main experiment. The second half of this experiment was identical to the first, except that the pure tones forming the stimuli were now presented successively rather than simultaneously. This allowed us to compare the mechanisms subtending frequency-ratio discrimination for simultaneous versus successive tones. Experiment 2 was a control experiment intended to check, as mentioned above, that all the data collected in the main experiment reflected frequency comparisons *across* peripheral frequency channels.

Experiment 1

Method

Participants. The tested listeners were six students in their twenties (three male, three female). They all had normal hearing. Some of them were amateur musicians, but none had received a thorough musical education. All listeners were paid for their participation, except listener L5, who was the first author.

Stimuli, conditions, and task. Each stimulus consisted of four diotically presented pure tones. In the "simultaneous" condition, these four tones were synchronous. In the "sequential" condition, they were presented consecutively, in ascending frequency order, with 100-ms silent interstimulus intervals (ISIs) between them. In both conditions, each tone had a total duration of 100 ms and was gated on and off with 20-ms raised cosine functions. The tones had a nominal SPL of 31 dB for listeners L1 and L2, and 46 dB for the four remaining listeners.[2] The four components of each stimulus were equally spaced on a logarithmic frequency scale. Their spacing, *Oct*, was expressed in octaves. On each stimulus presentation, the frequency of the lowest tone (*Fmin*) was selected randomly between 125 and 250 Hz; the corresponding probability distribution was rectangular, frequency being scaled logarithmically. The starting phases of the tones were random variables and did not depend on frequency.

On every trial, four successive stimuli were presented, as illustrated in Figure 1 for the "simultaneous" condition. There was a silent ISI of 300 ms between the first two stimuli (S1 and S2) and between the last two stimuli (S3 and S4). A longer ISI—600 ms—separated S3 from S2, thus segmenting the whole sequence into two pairs of stimuli. In three of the four stimuli, *Oct* had a fixed reference value ($Oct_{\text{reference}}$) of 0.97. In the remaining stimulus, which was either S2 or S4, at random, *Oct* was larger; this target value of *Oct* (Oct_{target}) could be equal to 0.98, 0.99, 1.00, 1.01, 1.02, 1.03, or 1.04. The listener's task was to indicate whether the odd man out was S2 or S4, by making a mouse click on one of two virtual buttons on a computer screen. Immediately after this response, the button was colored in white if the response was correct, or in black otherwise. Response time was not limited. Within each block of trials, a response triggered the next trial after a delay of about 1 s.

The stimuli were generated at a 44.1-kHz sampling rate, using 24-bit digital-to-analog converters (RME). They were presented via headphones (Sennheiser HD 650). Harmonic distortion at the output of the headphones was assessed using an artificial ear (Bruel & Kjaer 4153 equipped with a flat-plate coupler; microphone model: 4134) and a spectrum analyzer (Stanford Research, SR 780). For nominally pure tones ranging in frequency from 125

[2] We had initially planned to use 46-dB tones for each listener. However, during the practice sessions, it appeared that ceiling effects were likely to be obtained in listeners L1 and L2 if they were tested with tones at this level. It was thus decided to test these listeners using softer tones, making the task more difficult.

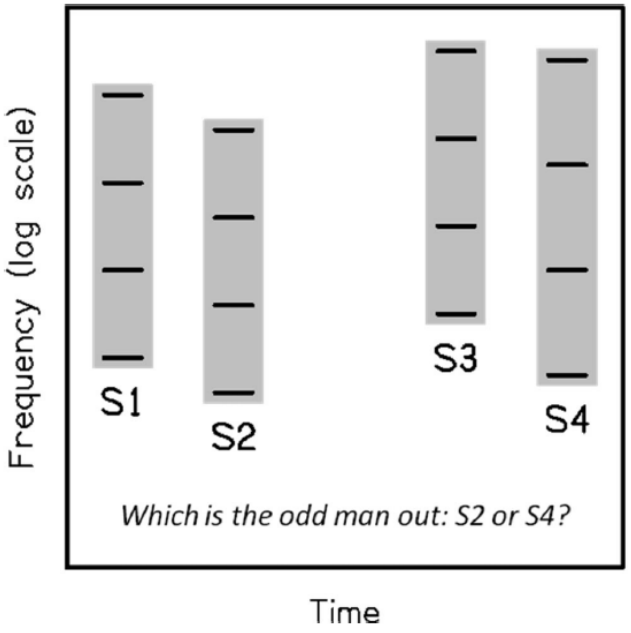


Figure 1. Schematic representation of listeners' task in the "simultaneous" condition of Experiment 1. On each trial, four stimuli (S1, S2, S3, S4) were successively presented (their relative timing is not faithfully depicted in this Figure). Each stimulus consisted of four synchronous pure tones, equally spaced on a logarithmic frequency scale. The spacing of the tones was the same (namely, 0.97 octave) for three stimuli, and was larger for the remaining stimulus, which was either S2 or S4. Listeners had to identify the odd man out as S2 or S4. The frequency of the lowest component of each stimulus was randomly chosen between 125 and 250 Hz. The "sequential" condition of Experiment 1 was similar, except that in this case the four components of each stimulus were presented consecutively.

to 1000 Hz and produced at 80–90 dB SPL, it was found that each component of the harmonic distortion was at least 50 dB below the fundamental.

Procedure. Listeners were tested individually in a triple-walled sound-attenuating booth (Gisol, Bordeaux). Trials were run in blocks of 50, during which Oct_{target} had a fixed value, known by the listener. Each experimental session consisted of seven blocks of trials, one block for each of the seven possible values of Oct_{target}. These seven blocks were randomly ordered.

For each listener, the experiment proper was preceded by at least six practice sessions in the "simultaneous" condition. Then, 10 formal sessions were run in that condition. They were followed by one or two practice sessions in the "sequential" condition, and then 10 formal sessions in that condition. Overall, therefore, the formal data resulted from 500 trials per listener for each combination of condition and Oct_{target} value.

Results

Before considering the results obtained in each individual listener, we shall focus on the mean data. Figure 2 shows the overall proportion of correct responses obtained in each condition ("simultaneous": black disks; "sequential": white disks) for each Oct_{target} value. Performance was not markedly better in one condition than the other, but the effect of Oct_{target} on performance clearly differed across the two conditions. In the "sequential" condition, the mean psychometric function obtained was approximately a straight line, consistent with the hypothesis identified above as H0 (no effect of frequency-ratio simplicity on ratio discrimination). In the "simultaneous" condition, by contrast, the mean psychometric function obtained was clearly nonmonotonic; it showed a peak for Oct_{target} 1, followed by a dip for Oct_{target} 1.02; this outcome is consistent with hypothesis H2.

The data were submitted to a repeated-measures ANOVA, Listener Condition Oct_{target}, which revealed a significant main effect of Oct_{target}, $F(6, 30)$ 14.80, p .001, and a significant interaction between Condition and Oct_{target}, $F(6, 30)$ 4.81, p .0015. There was no significant main effect of Condition, $F(1, 5)$ 0.03, p .88. Given the existence of a reliable interaction between Condition and Oct_{target}, we then submitted the data obtained in each condition to a separate repeated-measures ANOVA, Listener Oct_{target}. In each case, not surprisingly, the effect of Oct_{target} appeared to be significant, $F(6, 30)$ 8.45, p .001. Our main goal was to determine whether there were significant cubic trends in the effect of Oct_{target}, as predicted by H2. There was indeed a significant trend of this type for the "simultaneous" condition, $t(30)$ 5.11, p .001, but not for the "sequential" condition, $t(30)$ 0.27, p .79.

The data provided by each listener are plotted in Figure 3, where the curves represent best-fitting psychometric functions obtained from two quantitative models of perceptual processing, one model for each condition. These models are described in detail in the Appendix to this manuscript. It is especially noticeable in Figure 3 that there was essentially no correlation, across listeners, between the overall level of performance in the two conditions. For in-

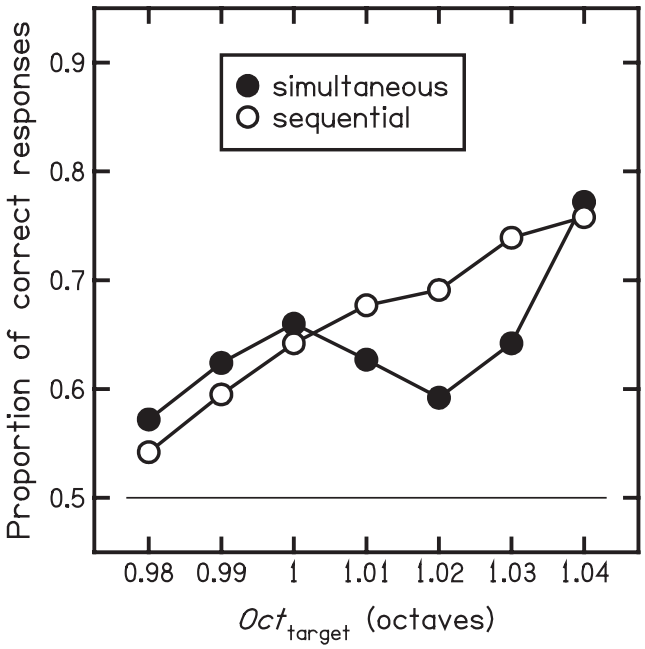


Figure 2. Results obtained in the "simultaneous" and "sequential" conditions of Experiment 1. Each data point represents the mean performance of six listeners. The thin horizontal line indicates the chance level of performance (50% of correct responses).

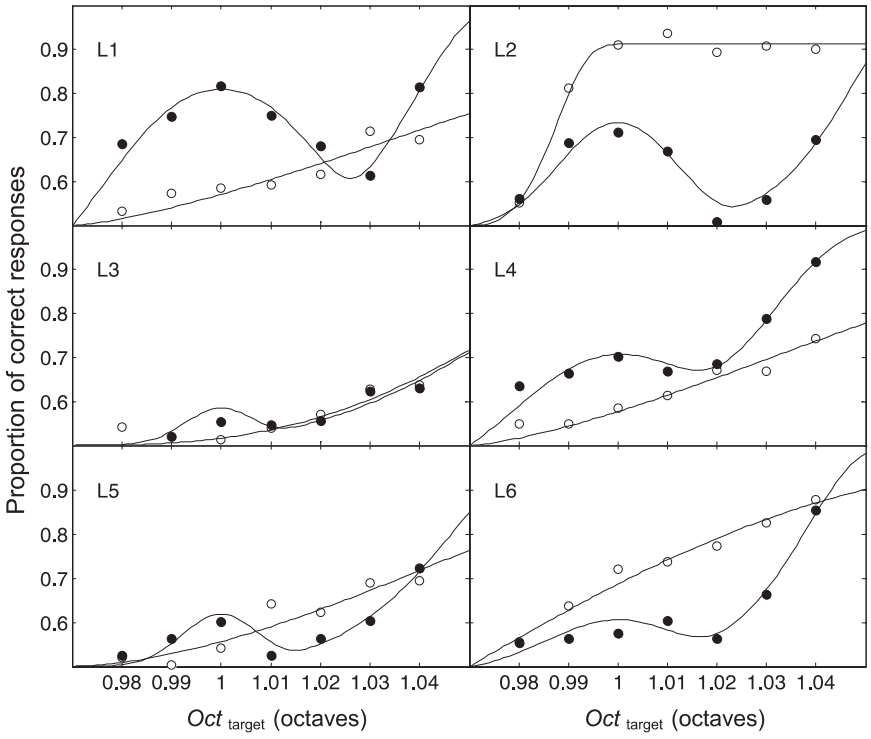


Figure 3. Each panel represents the results obtained from a given listener in Experiment 1. As in Fig. 2, black disks represent the "simultaneous" condition and white disks represent the "sequential" condition. Each data point is the outcome of 500 trials. The curves represent best-fitting psychometric functions derived from two models described in the Appendix. The model used for the "sequential" condition assumes that listeners used a single perceptual cue, which was a monotonic function of Oct_{target}. The model used for the "simultaneous" condition assumes that, in addition to a cue that was a monotonic function of Oct_{target}, listeners used a cue determined by the absolute value of the distance of Oct_{target} (and $Oct_{reference}$) from 1.

stance, while listener L1 performed markedly better in the "simultaneous" condition than in the "sequential" condition, the reverse was found in listener L2.

Discussion

Can we be sure that performance in this experiment always reflected listeners' ability to discriminate between frequency ratios, rather than between frequencies? To minimize the latter possibility, we had randomized the frequency of the lowest component of each stimulus, as pointed out above and illustrated in Figure 1. However, was the randomization range (1 octave) large enough? To answer that question, we determined by means of Monte Carlo simulations the performance expected from a virtual listener insensitive to changes in frequency ratio but producing optimal responses on the basis of frequency comparisons between stimuli. This virtual listener votes for stimulus S2 if the frequency of the highest component of S2 is higher than the frequency of the highest component of S4, and votes for S4 otherwise. We found that the proportion of correct responses provided by the virtual listener is approximately a linear function of $Oct_{target} - Oct_{reference}$. This proportion is 0.664 for Oct_{target} 1.03 and 0.688 for Oct_{target} 1.04. In the "sequential" condition, for these two values of Oct_{target}, five of the six listeners tested in the experiment (all participants except L3) performed better than the virtual listener; therefore, those five listeners must have based their judgments on comparisons between frequency ratios, as intended, rather than on comparisons between frequencies; listener L3, on the other hand, may have been unable to perform the task as intended. In the "simultaneous" condition, for all listeners except again L3, the nonmonotonic effect of Oct_{target} on performance is of course at odds with the idea that performance was merely determined by frequency comparisons.

Our main finding is that frequency–ratio discrimination was generally a nonmonotonic function of Oct_{target} in the "simultaneous" condition but not in the "sequential" condition. One conceivable explanation of this difference is that in the "simultaneous" condition, the pure tones forming the stimuli interacted in the cochlea, whereas this was precluded in the "sequential" condition. As pointed out above, cochlear interactions of simultaneous tones would have allowed listeners to discriminate between frequency ratios on the basis of beat cues. Beats were liable to be elicited by all of the stimuli used in the "simultaneous" condition, except those composed of tones exactly one octave apart. This can potentially account for the fact that, in the "simultaneous" condition, a local peak in discriminability was observed for Oct_{target} 1. To limit cochlear interactions, we presented the tones at a low SPL, namely 46 dB for four listeners and 31 dB for the remaining two listeners. To further investigate the extent of interactions between

the tones, we computed "excitation patterns" for the stimuli of the "simultaneous" condition, using the model of Moore, Glasberg and Baer (1997) revised by Glasberg and Moore (2006). Figure 4 shows the simulated excitation pattern obtained for the harmonic stimulus formed of tones at 125, 250, 500, and 1000 Hz when the SPL per tone was 46 dB (upper solid curve) and 31 dB (lower solid curve). Note that 125 Hz was the lowest possible tone frequency in the experiment, and that in this case cochlear interactions between tones were maximized (because the relative bandwidth of the peripheral auditory filters tends to increase at low frequencies). The dashed curve in Figure 4 shows the excitation corresponding to the absolute threshold of hearing (in normal young listeners). This curve thus represents a detection floor for the excitation patterns: portions of the patterns that fall below it are, in principle, undetectable by a normal-hearing listener. It can be seen that in the case of the lower-level excitation pattern, any detectable excitation originates from a single tone. At the higher level, however, listeners could in theory detect excitation produced by two simultaneous tones.

As a matter of fact, investigations on the audibility of beats of mistuned consonances in monaural dyads of pure tones have suggested that such beats are not audible when each of the tones is below 50 dB SPL (Plomp, 1967, 1976; Viemeister et al., 2001). Nevertheless, it was important to check that beats resulting from cochlear interactions were indeed never detectable in Experiment 1. This was the main purpose of Experiment 2.

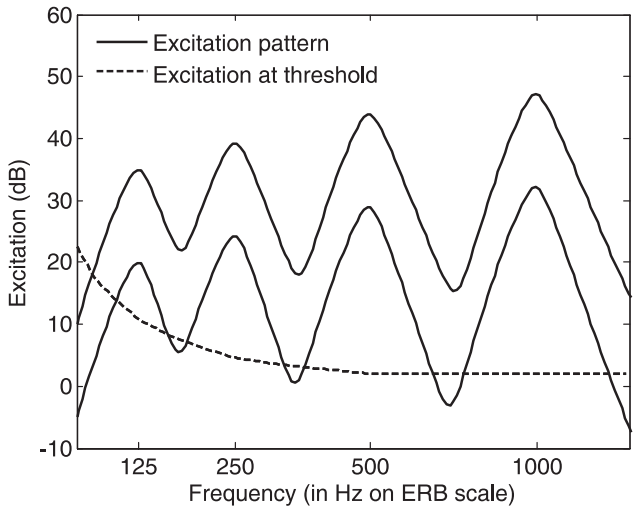


Figure 4. The two solid curves represent the cochlear excitation pattern produced by a sum of four pure tones at 125, 250, 500, and 1000 Hz when the SPL of each tone is 46 dB (upper curve) and 31 dB (lower curve). The dashed curve represents the amount of excitation produced by a just-detectable pure tone, as a function of its frequency. Frequency, on the abscissa, is scaled in accordance with the frequency dependence of the auditory filters' equivalent rectangular bandwidth (ERB; cf. Moore, 2003, chap. 3). The three curves were calculated on the basis of the model proposed by Moore, Glasberg, and Baer (1997) and revised by Glasberg and Moore (2006).

Experiment 2

Rationale

In a previous study by Demany, Semal, and Carlyon (1991), listeners' ability to detect deviations from the octave interval was investigated using dyads of simultaneous pure tones. On each trial, listeners were presented with two successive dyads, one composed of tones exactly one octave apart and one in which the interval formed by the tones was either slightly larger (positive mistuning) or slightly smaller (negative mistuning); the task was to indicate which of the two dyads was a mistuned octave. The two components of each dyad were presented at a low sensation level in a pink-noise background, so that they were unlikely to interact in the cochlea. Unexpectedly, the experimental results showed that negative mistunings were easier to detect than positive mistunings of the same size in terms of relative frequency deviation.[3] Although the origin of this perceptual asymmetry was not elucidated, one could infer from it that the listeners did not detect mistuning by detecting beats resulting from cochlear interactions: Had this been the case, no significant effect of the sign of mistuning was expected because the waveform obtained by adding two tones with frequencies f and $(2f \quad \varepsilon)$ Hz is essentially a mere temporal inversion of the waveform obtained when $2f \quad \varepsilon$ is replaced by $2f - \varepsilon$. We thought that if the same perceptual asymmetry was observable for stimuli similar to those used in Experiment 1, this would provide evidence that, in the latter experiment as well as in the previous study, mistuned octaves did not elicit significant cochlear beats. That was the rationale underlying Experiment 2.

Method

As in the "simultaneous" condition of Experiment 1, the stimuli consisted of four synchronous pure tones equally spaced on a logarithmic frequency scale, gated on and off with 20-ms raised cosine functions, and presented diotically at the same nominal SPL. Once more, the frequency of the lowest component of each stimulus was selected randomly between 125 and 250 Hz, and each component had a random initial phase. However, the stimuli now had a total duration of 500 ms instead of 100 ms; this increase in duration facilitated beat detection (Hartmann, 1988). As in Experiment 1, four successive stimuli (S1, S2, S3, S4) were presented on each trial and listeners were requested to identify the odd man out, which was equiprobably either S2 or S4. Again, the odd man out differed from the other three stimuli with respect to the frequency ratio of neighboring tones (*Oct*). The silent ISIs separating the four stimuli were the same as in Experiment 1, and correct-answer feedback was again provided visually following each response.

The main novelty of Experiment 2 was that *Oct* now had a reference value ($Oct_{\text{reference}}$) of exactly 1 octave. On each trial, therefore, the odd man out was the only stimulus for which *Oct* differed from 1 octave. In one condition, named the "stretching" condition, the value of *Oct* in the odd man out (Oct_{target}) was always larger than 1 octave. In another condition, named the "compression" condition, Oct_{target} was always smaller than 1 oc-

[3] For mistunings matched in terms of absolute, rather than relative, frequency deviation, the observed perceptual advantage of negative mistunings would have been very slightly (and negligibly) *larger*.

tave. In both conditions, Oct_{target} was varied adaptively within blocks of trials, in order to measure a perceptual threshold corresponding to the mistuning value, Δ_{oct} $|Oct_{target} - Oct_{reference}|$, for which the probability of a correct response was 0.75. This was done using the "weighted up-down" paradigm described by Kaernbach (1991). At the onset of each block of trials, Δ_{oct} was large, well above the expected threshold. Then, Δ_{oct} was decreased after each correct response and increased after each incorrect response. A block ended after the 14th reversal in the variation of Δ_{oct}. Until the 4th reversal, Δ_{oct} was multiplied by 2.25 when it was increased, and divided by the cube root of the same factor when it was decreased. After the 4th reversal, Δ_{oct} was either multiplied by 1.5 or divided by the cube root of this factor. The threshold measure obtained in a block of trials was computed as the geometric mean of all the Δ_{oct} values used from the 5th reversal on. Within each experimental session, thresholds were measured alternately in the "stretching" condition and the "compression" condition; the switch occurred after each threshold measurement.

The component tones of the stimuli were varied in SPL from 46 dB (the highest SPL of the tones in Experiment 1) to 76 dB, or vice versa, in 10-dB steps. The SPL was fixed within sessions. At each SPL and for each listener, 10 threshold measurements were made in the "stretching" condition and in the "compression" condition.

The apparatus used in this experiment was the same as in Experiment 1. Seven listeners (four male, three female) were tested. Three of them were listeners L4, L5, and L6 in Experiment 1. The four additional listeners (L7–L10) included three students in their twenties and author LD (aged 58 and designated as L7); these four listeners had normal hearing, at least for the frequency range covered by the stimuli; none of them had received a thorough musical education, but two (L7 and L10) had substantial previous experience with psychoacoustics. For each of the seven listeners, the experiment proper was preceded by a single practice session.

Results and Discussion

Figure 5 displays the results. Seven panels show, as a function of SPL, the mean threshold measured in each listener for the two conditions ("stretching" vs. "compression"). The individual results are averaged in the rightmost panel. Note first that SPL had a strong effect on thresholds: as the stimuli increased in SPL, thresholds systematically decreased (except in the case of L5 for the highest SPL). This trend was predictable because an increase in SPL increased the potential influence of beat cues originating from cochlear interactions between the tones. In line with the hypothesis that the task was performed using beat cues when the SPL was high, there was no significant difference between the thresholds obtained in the "stretching" and "compression" conditions at 66 dB, $t(6)$ 1, as well as at 76 dB, $t(6)$ 1.25, p .26. By contrast, there was a marginally significant effect of condition at 56 dB, $t(6)$ 2.02, p .06, and the effect of condition was definitely significant at 46 dB, $t(6)$ 3.97, p .007. At these two lower SPLs, thresholds were better in the "compression" condition than in the "stretching" condition, as expected from the study by Demany et al. (1991).

For 46-dB tones, we obtained a mean threshold of 0.0223 octave in the "stretching" condition and 0.0157 octave in the "compression" condition. It must be underlined that, in comparison with the reference value of *Oct* (1 octave), these thresholds represent quite small frequency deviations, although listeners' performance was even better at higher SPLs. If the thresholds obtained at 46 dB had been markedly higher, then their dependence on the sign of mistuning might have been accounted for under the hypothesis that, even for this low SPL, mistuning detection rested on beat detection. It would have been so because a compressive mistuning, by decreasing the frequency distance between the tones, increased the possibility of cochlear interactions, whereas the opposite occurred

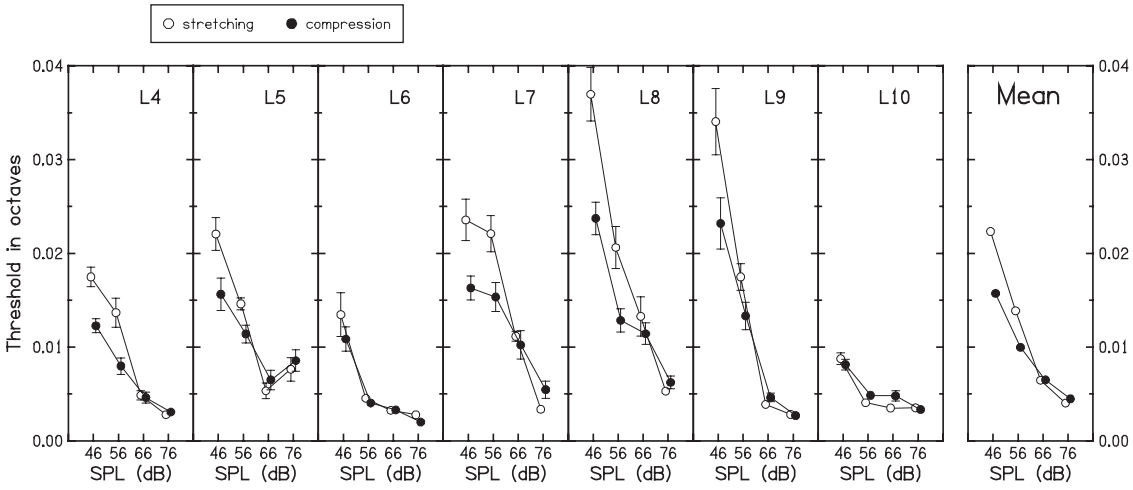


Figure 5. Results of Experiment 2: mistuning thresholds measured in the "stretching" condition and the "compression" condition as a function of the SPL of the tones. The data obtained from each of the seven listeners (L4, L5, . . . L10) are displayed in a separate panel. In these seven panels, each data point is the mean of 10 threshold measurements and the error bars represent 1 standard error of the mean. In the rightmost panel, the results of the seven listeners are averaged. In each panel, for clarity, the data points corresponding to the two mistuning conditions are slightly shifted horizontally (to the left for the "stretching" condition and to the right for the "compression" condition).

in the case of stretching. To quantify the corresponding contrast for 46-dB tones and mistunings such as the thresholds mentioned above, we used again the excitation-pattern model of Moore et al. (1997). We assumed that the magnitude of a cochlear interaction between two tones was directly reflected by the height of the excitation-pattern trough between their frequencies. The height of the troughs increased as *Oct* decreased. According to the model, this effect was strongest for the third trough, representing a potential interaction of the upper two tones; on average, the third trough was about 1.7-dB higher for the *Oct* value corresponding to a just-detectable compression than for the *Oct* value corresponding to a just-detectable stretching. Such a small difference does not seem able to account for the very substantial dependence of thresholds on the sign of mistuning. Moreover, if a 1.7-dB difference in excitation level were responsible for this dependence, then increasing the SPL of the tones by 10 dB should have improved thresholds considerably, whatever the sign of mistuning. According to the model of Moore et al. (1997), the third trough in the excitation pattern was about 12 dB higher for 56-dB tones than for 46-dB tones. In spite of this, the thresholds measured in the "stretching" condition with 56-dB tones did not differ significantly from those measured in the "compression" condition for 46-dB tones, $t(6)$ 1,14, p .30. Clearly, therefore, the perceptual asymmetry observed in the present experiment for 46-dB tones does not seem consistent with the hypothesis that, at this level, deviations from one octave were detected by means of beat cues originating from cochlear interactions. Consequently, this hypothesis also appears inadequate to account for the nonmonotonic psychometric functions found in Experiment 1.

It has been reported that, for some listeners, a mistuned octave composed of simultaneous pure tones can elicit the perception of faint beats when the two tones are presented to opposite ears (Thurlow & Bernstein, 1957; Thurlow & Elfner, 1959; Tobias, 1964; Feeney, 1997); according to Thurlow (Thurlow & Bernstein, 1957; Thurlow & Elfner, 1959), this is possible even for tones at a sensation level as low as 30 dB. The corresponding "dichotic beats of mistuned consonances (DBMCs)" have been interpreted as a *between*-channel phase effect originating from neural processes involved in sound localization by the binaural system (Feeney, 1997). Because in our two experiments the stimuli were presented diotically, it is conceivable that beat detection was facilitated by the binaural mechanism responsible for the audibility of DBMCs. However, the perceptual asymmetry disclosed by Experiment 2 does not seem easier to fit in with this "central beats" hypothesis than with the "peripheral beats" hypothesis considered above.

We still have no explanation for the perceptual asymmetry. As discussed later in this article, it might be expected, in theory, that simultaneous and soft pure tones approximately one octave apart fuse maximally when the octave is slightly stretched rather than perfectly tuned from the physical point of view. This bias could give rise to a perceptual advantage of negative mistunings (compressions) in mistuning detection. However, the results of Experiment 1 fail to provide evidence for the bias in question. Of course, negative mistunings could still be better detected than positive mistunings if maximum fusion was systematically obtained for the physically perfect octave: it may simply be, for some reason, that fusion decreases more rapidly for negative deviations from the perfect octave than for positive deviations.

General Discussion

The underpinnings of auditory sensitivity to frequency relations, and more specifically to the simplicity of frequency ratios, remain a matter of controversy. Although it is now understood why two simultaneous *complex* tones (for instance two vowels or violin sounds) are perceived as more consonant when their fundamental frequencies are in a simple frequency ratio than when this is not the case (Plomp, 1976; McDermott, Lehr, & Oxenham, 2010), the roots of harmonic fusion for simultaneous *pure* tones are still unclear. Moreover, the reason why an affinity is perceived between sequentially presented pure tones one octave apart is also unknown. As discussed below, both of these phenomena might originate from a learning mechanism, or they might stem from innate properties of the auditory system. In the present study, we sought to clarify the processes underlying the perception of frequency-ratio simplicity for pure tones. In particular, we sought to test the idea that the same processes are at work for simultaneous tones and for consecutive tones. We assessed the effect of ratio simplicity on ratio discrimination, for simultaneous and sequentially presented tones. Markedly different results were obtained in these two situations (Experiment 1), in spite of the absence of significant cochlear interactions between the tones when they were simultaneous (Experiment 2). For simultaneous tones, frequency-ratio discrimination appeared to be affected by ratio simplicity. Moreover, we found that ratio simplicity—or, more precisely, distance to the simple ratio 2:1—was in itself the physical correlate of one discrimination cue. This result fitted with a hypothesis called "H2" in the introduction of our article; according to H2, frequency-ratio discrimination takes place along two distinct perceptual continua, representing ratio simplicity and ratio magnitude. For sequentially presented tones, on the other hand, ratio simplicity had no effect.

Can the latter finding, concerning sequential tones, be ascribed to peculiarities of our stimuli? The tones that we presented either simultaneously or sequentially were the same in these two conditions; this was of course mandatory for a fair comparison between the two conditions. One might think, however, that the melodic sequences produced in the "sequential" condition were such that the frequency ratios of the tones could not be perceived optimally. Against this concern, it may be noted first that the speed of the sequences (five tones per second) was far from amusical: this speed is that of a series of eighth notes in "allegro" tempo; such series are very common in music. More crucially, it must be emphasized that, overall, listeners did not perform more poorly in the "sequential" condition than in the "simultaneous" condition, as shown by Figure 2; this is clearly at odds with the idea that frequency-ratio perception was disadvantaged in the "sequential" condition.

The fact that ratio simplicity had no effect in the "sequential" condition is not particularly surprising in the light of previous research. Although, for consecutive tones, Schellenberg and Trehub (1994, 1996a, 1996b) found an advantage of ratio simplicity in ratio discrimination, Houtsma (1968) reported that ratio discrimination thresholds do not vary significantly when the standard ratio varies from 1.9 to 2.1. Moreover, Dobbins and Cuddy (1982), who used stimuli similar to those used by Houtsma (1968), concluded from their frequency-ratio categorization experiment that the function relating physical frequency ratios to their percep-

tual representations is simply logarithmic (for musicians as well as nonmusicians). It may be that, for consecutive tones, ratio simplicity affects ratio discrimination only if special stimuli and/or experimental procedures are used, while simplicity effects are easier to demonstrate with simultaneous tones. We must also point out that although the results obtained in our "sequential" condition are inconsistent with the hypothesis referred to above as H2, they are not clearly inconsistent with H1. H2 posits that frequency-ratio discrimination takes place along two distinct perceptual continua representing simplicity and magnitude. According to H1, by contrast, magnitude is the only discrimination cue; simplicity is not in itself a discrimination cue; however, discriminability is better near simple ratios than near complex ratios. Our measurement of psychometric functions with a reference ratio of 0.97 octave was not well suited for a confrontation of H1 with H0 (according to which ratio simplicity has no influence at all on ratio discrimination).

Given that we obtained no effect of ratio simplicity with sequentially presented tones, and that a simultaneous presentation of the same tones did not produce extra cues resulting from cochlear interactions, it is remarkable that in the latter condition we did obtain an effect of ratio simplicity. This contrast reflects the fact that ratio simplicity produces quite different perceptual effects in the melodic and harmonic domains. Two successive pure tones forming a melodic octave are perceived as having related pitches, but, at least for normal adult listeners, they are easily distinguishable from each other. Indeed, with respect to pitch, their difference is not less obvious than their affinity. By contrast, two simultaneous pure tones forming an octave interval are perceptually fused into a single entity. In the auditory system, this fusion may well take place below the level at which representations of pitch per se are extracted from pure tones (as well as from complex tones). An analogous point was recently made by Borchert, Micheyl, and Oxenham (2011). In their study, listeners had to detect slight differences in pitch (fundamental frequency) between two complex tones which consisted of harmonics filtered into two separate, nonoverlapping spectral regions. When the tones were presented consecutively, performance was poor; listeners found the task difficult because of the large difference in timbre between the tones. However, when the tones were presented synchronously, performance was markedly better and the task was subjectively much easier, because it could then be performed by using a fusion cue rather than by explicit pitch comparisons.

How Can the Octave Be Recognized as a Simple Frequency Ratio?

In a very influential paper, Terhardt (1974) argued that the phenomenon of harmonic fusion originates from a learning process. He suggested that, early in life, humans initially perceive simultaneous pure tones forming a harmonic series as separate auditory entities, but that subsequently such tones are fused because of their frequent co-occurrence in the natural acoustic environment (and especially their systematic co-occurrence in vocal sounds). Shamma and Klein (2000) also took the view that harmonic fusion is learnt, and they put forth a physiologically based model whereby this learning could occur even in the absence of harmonic inputs. Although the spectral structure of the stimuli used here was always dissimilar to that of natural harmonic sounds (because the components of the latter sounds do not form *only* octave intervals), the frequency–ratio simplicity effect that we obtained with simultaneous tones could in principle be based on a learning process.

However, this is not necessarily the case. Interestingly, harmonicity is not a mandatory condition for the perceptual coherence of simultaneous pure tones, as shown by Roberts and his coworkers (Brunstrom & Roberts, 2000; Roberts & Bailey, 1996; Roberts & Brunstrom, 1998, 2001, 2003). When the elements of a harmonic series (e.g., 200, 400, 600, . . . 2000 Hz) are all shifted in frequency by the same amount in hertz (e.g., 50 Hz), their perceptual coherence is somewhat reduced but not dramatically, although they no longer form a harmonic series. Their constant frequency spacing is sufficient to produce perceptual coherence, as indicated by the fact that if a single element of the inharmonic set is slightly shifted in frequency, thus making the frequency spacing locally irregular, the shifted element tends to stand out perceptually. Given the rarity, in our acoustic environment, of inharmonic stimuli such as the one just described, their perceptual coherence is unlikely to originate from a learning process.[4]

As pointed out by Roberts and Brunstrom (2001), the perceptual coherence of these inharmonic stimuli, as well as harmonic fusion, may instead originate from the temporal coding of frequency by the auditory system. Because the successive action potentials elicited by a pure tone in an auditory-nerve fiber tend to occur at a particular phase of the tone, these successive neural spikes are generally separated by time intervals close to the period of the tone and its integer multiples (see, e.g., McKinney & Delgutte, 1999). This implies that simultaneous pure tones forming simple frequency ratios will produce, in separate auditory-nerve fibers, spikes separated by common time intervals. For two pure tones one octave apart, indeed, all the time intervals present in the neural response to the lower-frequency tone should also be present, approximately, in the neural response to the higher-frequency tone. Relations of that kind may be recognized at a higher level of the auditory system, perhaps after an autocorrelation of the peripheral neural responses (Licklider, 1951; Meddis & Hewitt, 1991, 1992; Patterson, Handel, Yost, & Datta, 1996; Yost, 1996; Cariani & Delgutte, 1996; Cariani, 2001; de Cheveigné, 2005). Such a mechanism can in principle explain harmonic fusion, as well as many aspects of pitch perception, including the perceptual affinity of *consecutive* tones one octave apart (Ohgushi, 1983; Hartmann, 1993; McKinney & Delgutte, 1999).

Therefore, the perceptual discrimination of an octave from slightly larger or smaller frequency ratios could rest upon the use of temporal information. Alternatively, this perceptual discrimination might be based on the use of spatial information. In the latter view, consistent with the fact that the auditory system has a tonotopic organization, an octave is perceptually defined as a specific *distance* between neural activations induced by pure tones. For simultaneous tones, our data would then imply, rather paradoxically, that deviations from this specific distance can be detected without being recognized as positive or negative deviations. Indeed, in the "simultaneous" condition of Experiment 1, the fact

[4] One can also discard the idea that their perceptual coherence originates from special intrinsic properties of their spatial (tonotopic) representation in the auditory system: Pure tones equally spaced in Hertz do *not* produce equally spaced neural excitations.

that performance decreased when the target frequency ratio was made larger than one octave implied that the (positive) sign of this mistuning from one octave could not be discriminated from the (negative) sign of the mistuning of the reference frequency ratio (0.97 octave).

Harmonic Versus Melodic Octaves: Is There a Perceptual Link?

As mentioned in the previous section of this Discussion, Terhardt (1974) has argued that harmonic fusion stems from a learning process. In his theory, two simultaneous pure tones one octave apart are fused merely because they co-occur in frequently heard sounds, such as the vowels of speech. Terhardt also suggested that, in consequence of our frequent exposition to vowels and other periodic complex tones, we acquire not only harmonic octave templates but also melodic octave templates. More generally, the learning process responsible for harmonic fusion would also account for the affinity perceived between successive pure tones forming simple frequency ratios, such as the octave. The results reported here do not support the latter hypothesis. If, as supposed by Terhardt, melodic and harmonic octaves were recognized by means of the same internal templates, then these templates should show similar properties in melodic and harmonic conditions. However, our results imply that the harmonic octave is recognized by a mechanism which is insensitive to the sign of deviations from the octave, whereas this is apparently not the case for the melodic octave.

To explain the "octave enlargement" phenomenon described in the introductory section of this article, Terhardt (1970, 1971, 1974; see also Terhardt, Stoll, & Seewann, 1982) argued that the harmonic complex tones typical of our auditory environment are such that their spectral components partially mask each other in the auditory periphery. As a result, the pitches evoked by these spectral components would be slightly different from those evoked by physically identical pure tones presented in isolation or successively. Partial masking, according to Terhardt, produces repulsive pitch shifts; this would account for the fact that, in general, a melodic octave is heard as optimally tuned for a physical frequency ratio slightly larger than 2:1.[5] Here, we used complex tones composed of pure tones which did not mask each other significantly, as shown by Experiment 2. Thus, the pitch-shift effects invoked by Terhardt should not have occurred for these stimuli. In consequence, Terhardt's theory predicted that maximum fusion would be obtained for frequency ratios slightly larger than one octave. Our results do not conform to this prediction: in the "simultaneous" condition of Experiment 1, all listeners but one (L6) performed better when the target frequency ratio was exactly one octave than when it was equal to 1.01 octave, as can be seen in Figure 3. Admittedly, the target frequency ratios may have been too coarsely sampled to reveal an octave-enlargement effect in the "simultaneous" condition: when listeners are required to set two successive tones one octave apart, only very small enlargement effects (1%) are found at low frequencies, in spectral regions covered by the stimuli used here (cf. Figure 1 in McKinney & Delgutte, 1999).

Nevertheless, as pointed out above, our results clearly question the validity of Terhardt's hypothesis on the origin of tonal affinity and the perception of the melodic octave. Two other problems for this hypothesis have been uncovered previously. First, Terhardt's observations on the pitch shifts of components of harmonic complex tones are apparently difficult to replicate; it has been suggested that these pitch shifts do not really exist (Peters, Moore, & Glasberg, 1983; Hartmann & Doty, 1996). Second, the precision with which harmonic octave dyads can be identified becomes quite poor when the frequency of the lower tone exceeds about 1000 Hz, whereas this is not the case for melodic octave dyads (Demany & Semal, 1990; Demany et al., 1991); if sensitivity to melodic "octaveness" were derived from experience with harmonic complex sounds, how could it be more acute than sensitivity to harmonic octaveness? However, the latter objection applies only to the perception of melodic octaves at *high* frequencies. This is also largely true for the former objection, insofar as the octave enlargement effect is quite small and barely detectable at low frequencies. In contrast, the present research suggests that harmonic and melodic octaves are perceptually recognized as "special" frequency ratios by different processes even at low frequencies.

[5] Terhardt's explanation for the octave enlargement effect is based on the assumption that the pitch of a pure tone is coded spatially (tonotopically) in the auditory system. Ohgushi (1983) and McKinney and Delgutte (1999) have proposed an alternative explanation, assuming instead that pure tone pitch depends on the temporal coding of frequency in the auditory nerve.

References

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *The American Journal of Psychology, 84,* 147–166. doi:10.2307/1421351

Borchert, E. M., Micheyl, C., & Oxenham, A. J. (2011). Perceptual grouping affects pitch judgments across time and frequency. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 257–269. doi:10.1037/a0020670

Brunstrom, J. M., & Roberts, B. (2000). Separate mechanisms govern the selection of spectral components for perceptual fusion and for the computation of global pitch. *Journal of the Acoustical Society of America, 107,* 1566–1577. doi:10.1121/1.428441

Burns, E. M., & Ward, W. D. (1978). Categorical perception - phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America, 63,* 456–468. doi:10.1121/1.381737

Cariani, P. (2001). Temporal codes, timing nets, and music perception. *Journal of New Music Research, 30,* 107–135. doi:10.1076/jnmr.30.2.107.7115

Cariani, P. A., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology, 76,* 1698–1716.

Dai, H., & Micheyl, C. (2011). Psychometric functions for pure-tone frequency discrimination. *Journal of the Acoustical Society of America, 130,* 263–272. doi:10.1121/1.3598448

de Cheveigné, A. (2005). Pitch perception models. In C. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 169–233). New York, NY: Springer.

Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America, 76,* 57–66. doi:10.1121/1.391006

Demany, L., & Semal, C. (1988). Dichotic fusion of two tones one octave apart: Evidence for internal octave templates. *Journal of the Acoustical Society of America, 83,* 687–695. doi:10.1121/1.396164

Demany, L., & Semal, C. (1990). Harmonic and melodic octave templates. *Journal of the Acoustical Society of America, 88,* 2126–2135. doi:10.1121/1.400109

Demany, L., & Semal, C. (1992). Detection of inharmonicity in dichotic pure-tone dyads. *Hearing Research, 61,* 161–166. doi:10.1016/0378-5955(92)90047-Q

Demany, L., Semal, C., & Carlyon, R. P. (1991). On the perceptual limits of octave harmony and their origin. *Journal of the Acoustical Society of America, 90,* 3019–3027. doi:10.1121/1.401776

Deutsch, D. (1973). Octave generalization of specific interference effects in memory for tonal pitch. *Perception & Psychophysics, 13,* 271–275. doi:10.3758/BF03214138

Dobbins, P. A., & Cuddy, L. L. (1982). Octave discrimination: An experimental confirmation of the "stretched" subjective octave. *Journal of the Acoustical Society of America, 72,* 411–415. doi:10.1121/1.388093

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49,* 524–531. doi:10.1121/1.1912382

Feeney, M. P. (1997). Dichotic beats of mistuned consonances. *Journal of the Acoustical Society of America, 102*(4), 2333–2342. doi:10.1121/1.419602

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. Boca Raton, FL: Chapman & Hall.

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

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.

Hartmann, W. M. (1988). Pitch perception and the segregation and integration of auditory entities. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function – Neurobiological bases of hearing* (pp. 623–645). New York, NY: Wiley.

Hartmann, W. M. (1993). On the origin of the enlarged melodic octave. *Journal of the Acoustical Society of America, 93*(6), 3400–3409. doi:10.1121/1.405695

Hartmann, W. M., & Doty, S. L. (1996). On the pitches of the components of a complex tone. *Journal of the Acoustical Society of America, 99,* 567–578. doi:10.1121/1.414514

Hartmann, W. M., McAdams, S., & Smith, B. K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *Journal of the Acoustical Society of America, 88,* 1712–1724. doi:10.1121/1.400246

Houtsma, A. J. M. (1968). Discrimination of frequency ratios. *Journal of the Acoustical Society of America, 44,* 383 (abstract). doi:10.1121/1.1970636

Idson, W. L., & Massaro, D. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics, 24,* 551–565. doi:10.3758/BF03198783

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics, 49,* 227–229. doi:10.3758/BF03214307

Levelt, W. J. M., van de Geer, J. P., & Plomp, R. (1966). Triadic comparisons of musical intervals. *British Journal of Mathematical and Statistical Psychology, 19,* 163–179. doi:10.1111/j.2044-8317.1966.tb00366.x

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia, 7,* 128–134. doi:10.1007/BF02156143

Lin, J. Y., & Hartmann, W. M. (1998). The pitch of a mistuned harmonic: Evidence for a template model. *Journal of the Acoustical Society of America, 103,* 2608–2617. doi:10.1121/1.422781

McDermott, J. H., Keebler, M. V., Micheyl, C., & Oxenham, A. J. (2010). Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. *Journal of the Acoustical Society of America, 128,* 1943–1951. doi:10.1121/1.3478785

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology, 20,* 1035–1041. doi:10.1016/j.cub.2010.04.019

McKinney, M. F., & Delgutte, B. (1999). A possible neurophysiological basis of the octave enlargement effect. *Journal of the Acoustical Society of America, 106,* 2679–2692. doi:10.1121/1.428098

Meddis, R., & Hewitt, M. J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America, 89,* 2866–2882. doi:10.1121/1.400725

Meddis, R., & Hewitt, M. J. (1992). Modeling the identification of concurrent vowels with different fundamental frequencies. *Journal of the Acoustical Society of America, 91,* 233–245. doi:10.1121/1.402767

Moore, B. C. J. (2003). *An introduction to the psychology of hearing*. Amsterdam, The Netherlands: Elsevier.

Moore, B. C. J., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society, 45,* 224–240.

Moore, B. C. J., Peters, R. W., & Glasberg, B. R. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America, 80,* 479–483. doi:10.1121/1.394043

Ohgushi, K. (1983). The origin of tonality and a possible explanation of the octave enlargement phenomenon. *Journal of the Acoustical Society of America, 73,* 1694–1700. doi:10.1121/1.389392

Patterson, R. D., Handel, S., Yost, W. A., & Datta, A. J. (1996). The relative strength of the tone and noise components in iterated rippled noise. *Journal of the Acoustical Society of America, 100,* 3286–3294. doi:10.1121/1.417212

Peters, R. W., Moore, B. C. J., & Glasberg, B. R. (1983). Pitch of components of a complex tone. *Journal of the Acoustical Society of America, 73,* 924–929. doi:10.1121/1.389017

Plomp, R. (1967). Beats of mistuned consonances. *Journal of the Acoustical Society of America, 42,* 462–474. doi:10.1121/1.1910602

Plomp, R. (1976). *Aspects of tone sensation*. London, UK: Academic Press.

Plomp, R., Wagenaar, W. A., & Mimpen, A. M. (1973). Musical interval recognition with simultaneous tones. *Acustica, 29,* 101–109.

Roberts, B., & Bailey, P. J. (1996). Spectral regularity as a factor distinct from harmonic relations in auditory grouping. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 604–614. doi:10.1037/0096-1523.22.3.604

Roberts, B., & Brunstrom, J. M. (1998). Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes. *Journal of the Acoustical Society of America, 104(4),* 2326–2338. doi:10.1121/1.423771

Roberts, B., & Brunstrom, J. M. (2001). Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch. *Journal of the Acoustical Society of America, 110,* 2479–2490. doi:10.1121/1.1410965

Roberts, B., & Brunstrom, J. M. (2003). Spectral pattern, harmonic relations, and the perceptual grouping of low-numbered components. *Journal of the Acoustical Society of America, 114,* 2118–2134. doi:10.1121/1.1605411

Schellenberg, E. G. (2001). Asymmetries in the discrimination of musical intervals: Going out of tune is more noticeable than going in tune. *Music Perception, 19,* 223–248. doi:10.1525/mp.2001.19.2.223

Schellenberg, E. G., & Trehub, S. E. (1994). Frequency ratios and the discrimination of pure tone sequences. *Perception & Psychophysics, 56,* 472–478. doi:10.3758/BF03206738

Schellenberg, E. G., & Trehub, S. E. (1996a). Children's discrimination of melodic intervals. *Developmental Psychology, 32*(6), 1039–1050. doi:10.1037/0012-1649.32.6.1039

Schellenberg, E. G., & Trehub, S. E. (1996b). Natural musical intervals: Evidence from infant listeners. *Psychological Science, 7,* 272–277. doi:10.1111/j.1467-9280.1996.tb00373.x

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464. doi:10.1214/aos/1176344136

Shamma, S., & Klein, D. J. (2000). The case of the missing pitch templates: How harmonic templates may form in the early auditory system. *Journal of the Acoustical Society of America, 107,* 2631–2644. doi:10.1121/1.428649

Stoelinga, C., & Lutfi, R. A. (2010). Discrimination of frequency ratios. Poster presented at the 23rd Midwinter Research Meeting of the Association for Research in Otolaryngology (Anaheim, CA). Abstract 333.

Sundberg, J. E. F., & Lindqvist, J. (1973). Musical octaves and pitch. *Journal of the Acoustical Society of America, 54,* 922–929. doi:10.1121/1.1914347

Terhardt, E. (1970). Oktavspreizung und Tonhöhenverschiebung bei Sinustönen. *Acustica, 22,* 345–351.

Terhardt, E. (1971). Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon. In *Proceedings of the 7th International Congress on Acoustics,* Budapest, Hungary, vol. 3, pp. 621–624.

Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America, 55*(5), 1061–1069. doi:10.1121/1.1914648

Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music Perception, 1,* 276–295.

Terhardt, E., Stoll, G., & Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America, 71,* 679–688. doi:10.1121/1.387544

Thurlow, W. R., & Bernstein, S. (1957). Simultaneous two-tone pitch discrimination. *Journal of the Acoustical Society of America, 29,* 515–519. doi:10.1121/1.1908946

Thurlow, W. R., & Elfner, L. F. (1959). Pure-tone cross-ear localization effects. *Journal of the Acoustical Society of America, 31,* 1606–1608. doi:10.1121/1.1907666

Tobias, J. V. (1963). Application of a "relative" procedure to a problem in binaural-beat perception. *Journal of the Acoustical Society of America, 35,* 1442–1447. doi:10.1121/1.1918710

Trainor, L. J. (1997). Effect of frequency ratio on infants' and adults' discrimination of simultaneous intervals. *Journal of Experimental Psychology: Human Perception and Performance, 23*(5), 1427–1438. doi:10.1037/0096-1523.23.5.1427

Viemeister, N. F., & Fantini, D. A. (1987). Discrimination of frequency ratios. In W. A. Yost & C. S. Watson (Eds.), *Auditory Processing of Complex Sounds* (pp. 47–56). Hillsdale, NJ: Lawrence Erlbaum.

Viemeister, N. F., Rickert, M., & Stellmack, M. A. (2001). Beats of mistuned consonances: Implications for auditory coding. In D. J. Breebart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, & R. Schoonhoven (Eds.), *Physiological and psychophysical bases of auditory function* (pp. 113–120). Maastricht, The Netherlands: Shaker.

Walliser, K. (1969). Über die Spreizung von empfundenen Intervallen gegenüber mathematisch harmonischen Intervallen bei Sinustönen. *Frequenz, 23,* 139–143. doi:10.1515/FREQ.1969.23.5.139

Ward, W. D. (1954). Subjective musical pitch. *Journal of the Acoustical Society of America, 26,* 369–380. doi:10.1121/1.1907344

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function. I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63,* 1293–1313. doi:10.3758/BF03194544

Yost, W. A. (1996). Pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America, 100,* 3329–3335. doi:10.1121/1.416973

4 PHYSIOPATHOLOGIE ET CONSEQUENCES DES SURDITES NEUROSENSORIELLES

4.1 Les différents types de surdité

Les surdités peuvent être classées selon deux modalités principales : le degré de sévérité de la déficience auditive et le mécanisme à l'origine de la perte auditive. Le degré de sévérité est mesuré en pratique clinique par une audiométrie tonale liminaire qui estime le seuil de détection de sons purs de différentes fréquences, généralement entre 125 et 8000 Hz. Le seuil absolu d'un sujet est exprimé pour chacune des fréquences étudiées par rapport au niveau théorique considéré comme normal chez l'homme (égal à 0 dB HL pour *Hearing Level*), établi en moyennant les seuils de détection de sujets jeunes sans anomalie auditive. En audiologie, le seuil auditif absolu d'un sujet pour une fréquence donnée s'exprime en dB HL, et correspond à l'écart entre le seuil de détection mesuré cliniquement et ce niveau théorique.

Le Tableau 1 présente la classification audiométrique des déficiences auditives proposée par le BIAP (Bureau International d'Audiophonologie, 1997). Elle repose sur la moyenne des seuils absolus obtenus pour les fréquences 500, 1000, 2000 et 4000 Hz (la somme des seuils en dB étant divisée par quatre puis arrondie à l'unité supérieure). Toute fréquence non perçue est notée à 120 dB. En cas de surdité asymétrique, le niveau moyen de perte en dB est multiplié par 7 pour la meilleure oreille et par 3 pour la plus mauvaise oreille ; la somme est divisée par 10.

Classification audiométrique	*Perte tonale moyenne (dB)*
Audition **normale et subnormale**	≤ 20
Déficience auditive **légère**	21 – 40
Déficience auditive **moyenne 1er degré**	41 – 55
2ème degré	56 – 70
Déficience auditive **sévère 1er degré**	71 – 80
2ème degré	81 – 90
Déficience auditive **profonde 1er degré**	91 – 100
2ème degré	101 – 110
3ème degré	111 – 119
Déficience auditive **totale** (cophose)	120

Tableau 1. Classification audiométrique des déficiences auditives selon le BIAP (1997)

Les surdités sont également traditionnellement classées en deux grandes catégories reposant sur le mécanisme à l'origine de la perte auditive ; on distingue les surdités de transmission des surdités de perception (aussi appelées neurosensorielles). Parfois, les deux peuvent s'associer, définissant une surdité mixte. Les surdités de transmission entrainent une diminution quantitative de l'énergie acoustique mécanique transmise depuis le milieu extérieur jusqu'à l'oreille interne. Les structures anatomiques concernées sont essentiellement le conduit auditif externe et l'oreille moyenne (notamment le tympan et la chaine ossiculaire). La conséquence principale de ce type de surdité est une diminution de l'intensité des sons perçus (et donc une dégradation des seuils absolus), qui peut s'accompagner dans une certaine mesure de modifications tonales si le degré de perte varie selon la fréquence. Néanmoins, contrairement aux surdités de perception, les surdités de transmission ne s'accompagnent pas d'une

altération significative des autres aspects de la perception auditive ; elles n'entrainent pas par exemple de phénomènes de distorsion objectivables par une audiométrie vocale. Quand le traitement parvient à rétablir un niveau suffisant d'énergie acoustique atteignant la cochlée (par restauration de l'effet columellaire ou par amplification), le système auditif dans son ensemble retrouve des performances très proches de la normale.

Le cas des surdités de perception, ou neurosensorielles, est plus complexe. Cette entité regroupe en effet un grand nombre d'anomalies de nature et de localisation anatomique variées, pouvant toucher la cochlée elle-même (origine endocochléaire), mais aussi le nerf auditif ou le système nerveux central jusqu'au cortex auditif (origine rétrocochléaire). Le point commun à toutes les surdités de perception « pures » est que l'information sonore disponible à l'entrée de la cochlée est strictement normale. L'anomalie peut concerner la transduction, la conduction ou le traitement de l'information auditive. Schématiquement, trois facteurs peuvent rendre compte des difficultés à fixer des limites nosologiques précises à ces surdités neurosensorielles.

(1) Une étiologie donnée de surdité peut entrainer des lésions très différentes au sein d'une même structure, et peut parfois concerner plusieurs structures simultanément. La limite conceptuellement claire entre surdité endo et rétrocochléaire est donc en pratique parfois difficile à tracer. Un bon exemple est donnée par la presbyacousie, cause la plus fréquente de surdité de perception de l'adulte, qui peut toucher isolément ou de manière associée les cellules ciliées, la strie vasculaire, la membrane basilaire et les neurones auditifs (Schuknecht, 1993; Nelson & Hinojosa, 2003).

(2) Une lésion périphérique sévère mais isolée et stable (typiquement endocochléaire par exemple) va entrainer une déprivation auditive, qui peut elle-même à long terme aboutir à une réorganisation de structures plus centrales du système auditif.

(3) Les outils diagnostiques disponibles sont souvent pris en défaut pour affirmer l'origine anatomique et physiopathologique précise d'une surdité neurosensorielle chez un sujet donné.

Les surdités neurosensorielles ne constituent donc pas une classe homogène de déficiences auditives sur le plan physiopathologique. Néanmoins, cliniquement, certains signes auditifs sont fréquemment retrouvés. A la dégradation des seuils absolus s'ajoutent des altérations qualitatives supraliminaires de l'information sonore portant notamment sur la hauteur tonale (diplacousie), le timbre ou l'intensité (recrutement). La gêne auditive se manifeste principalement par des troubles de la compréhension de la parole, particulièrement aggravés dans les environnements sonores multi-sources (réunions, repas de groupes, transport en commun, etc.). Les sujets rapportent souvent qu'ils « entendent, mais ne comprennent pas », ce qui témoigne de la distorsion sonore objectivable sur l'audiométrie vocale.

En tenant compte des nuances énoncées plus haut, on peut admettre qu'en dehors des cas de lésions rétrocochléaires focales indiscutables (tumeur du nerf auditif, lésion localisée objectivable d'une structure auditive centrale…), beaucoup de surdités neurosensorielles résultent au moins en partie d'une anomalie d'origine cochléaire. Dans la suite de cette thèse, il est convenu que le terme de *surdité neurosensorielle* (voire simplement de *surdité*) sous-entend la présence d'un dysfonctionnement d'origine cochléaire, même s'il n'est pas isolé. Les troubles perceptifs rapportés par les patients ont d'ailleurs été avant tout explorés par des études portant sur les mécanismes cochléaires eux-mêmes. La section suivante en résume les données essentielles.

4.2 La dégradation des mécanismes actifs

La physiologie d'une cochlée saine repose en grande partie sur la présence d'un mécanisme actif qui peut être mis en évidence chez l'animal en étudiant notamment la « réponse » de la membrane basilaire (c'est-à-dire l'amplitude ou la vitesse de son déplacement) en fonction de la fréquence de stimulation. Sellick, Patuzzi & Johnstone (1982) ont mesuré, en un point donné de la membrane basilaire, le niveau d'intensité sonore minimal requis pour entrainer une vitesse de déplacement constante prédéterminée, en fonction de la fréquence du stimulus. Ces mesures permettent d'établir ce que l'on appelle des *courbes d'accord*, présentées par la Figure 10.

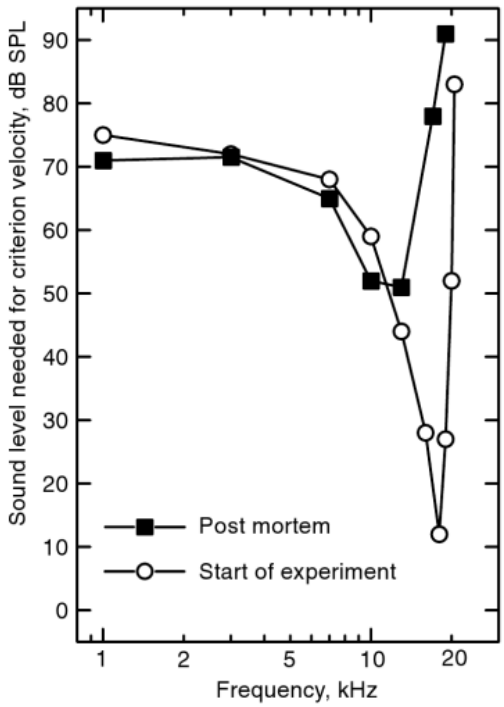


Figure 10. Courbes d'accord indiquant, en un point donné de la membrane basilaire, le niveau d'intensité sonore minimal requis (axe des ordonnées) pour déclencher un mouvement de la membrane d'une vitesse constante prédéterminée (0.04 mm/s), en fonction de la fréquence du stimulus (axe des abscisses). La courbe balisée par les cercles blancs indique les données enregistrées sur un cochon d'inde vivant, dont la cochlée est considérée comme physiologiquement normale. La courbe balisée par les carrés noirs montre les données obtenues après la mort de l'animal, et témoigne donc de la seule présence des mécanismes passifs. Schéma de Moore (2007), d'après les données de Sellick et al. (1982).

Au début de l'expérience (courbe balisée par des cercles blancs), la cochlée peut être considérée comme physiologiquement normale. La courbe d'accord se présente

sous la forme d'un « V » dont la pointe est très effilée, et descend très bas jusqu'à de faibles niveaux de stimulation (environ 12 dB SPL). La seconde courbe d'accord (balisée par des carrés noirs) a été obtenue après la mort de l'animal, ce qui implique que seul le mécanisme passif est préservé. Dans ce cas la courbe d'accord est beaucoup large, la pointe est beaucoup moins effilée et le niveau d'intensité sonore minimal nécessaire à entrainer le déplacement requis est beaucoup plus élevé (environ 50 dB SPL). Ces données montrent que lorsque la cochlée est physiologiquement saine, chaque point de la membrane basilaire est particulièrement sensible à une gamme étroite de fréquences, entrainant une réponse dès les faibles niveaux d'intensité. Si le stimulus est en dehors de cette gamme de fréquences, le niveau d'intensité requis pour déclencher la même réponse est beaucoup plus élevé. La courbe d'accord obtenue dans ce cas reflète la combinaison des mécanismes actif et passif. En revanche, en l'absence de mécanisme actif, la réponse d'un point de la membrane basilaire est déclenchée par des stimulations de plus forte intensité, avec une spécificité fréquentielle beaucoup moins marquée.

De nombreux travaux ont permis de démonter que ce mécanisme actif était lié à l'intégrité des cellules ciliées externes (CCE). Evans & Harrison (1976) ont utilisé un aminoside (la kanamycine), dont la toxicité est principalement dirigée contre les CCE. Ils ont montré que la normalité des seuils et de la spécificité fréquentielle de la réponse d'une fibre du nerf auditif dépendait directement de l'intégrité des CCE. Liberman, Dodds & Learson (1986) ont également utilisé une substance ototoxique, en association avec un traumatisme sonore, pour obtenir plusieurs types de lésions pouvant toucher isolément les CCE ou les cellules ciliées internes (CCI). Ils ont ensuite étudié les courbes d'accord obtenues à partir des réponses d'une fibre du nerf auditif. Une technique de marquage permettait de repérer la CCI avec laquelle le neurone étudié faisait synapse, permettant ainsi de mettre en rapport les résultats électrophysiologiques obtenus avec

l'étude microscopique de la portion exacte de l'organe de Corti correspondante. La Figure 11 résume les principaux résultats.

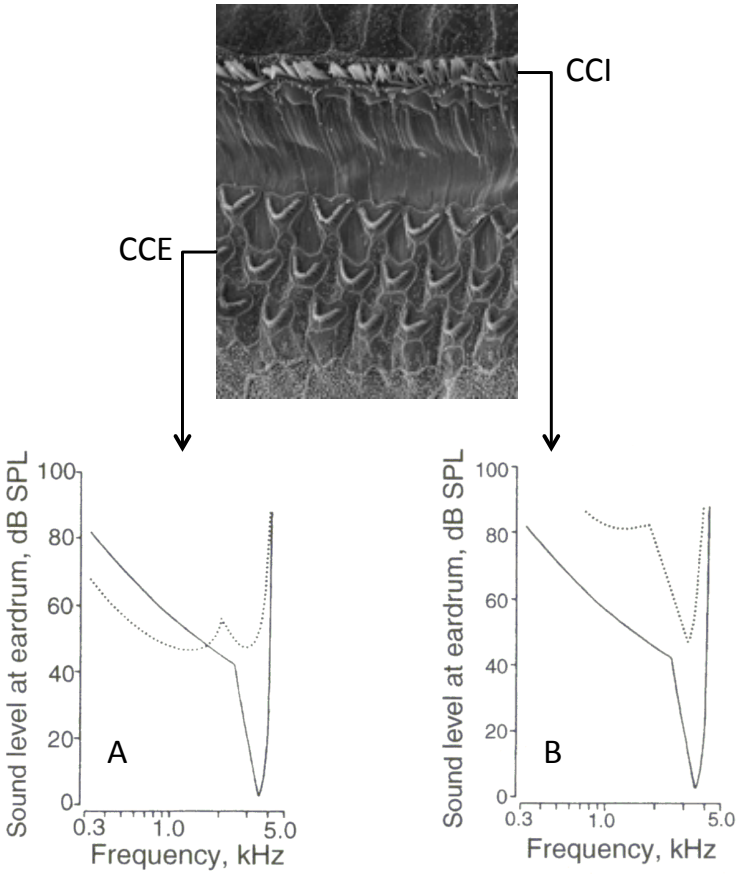


Figure 11. (En haut) Image de microscopie électronique d'une cochlée de mammifère normale, mettant en évidence la rangée de CCI et les trois rangées de CCE isolément touchées lors des manipulations expérimentales. **(A)** Courbe d'accord obtenue en cas de lésion isolée des CCE (traits pointillés), par rapport à une courbe normale en l'absence de lésion cochléaire (trait plein). **(B)** Courbe d'accord obtenue en cas de lésion isolée des CCI (traits pointillés), par rapport à une courbe normale en l'absence de lésion cochléaire (trait plein). Schéma adapté de Moore (2007), d'après les données de Liberman, Dodds & Learson (1986).

En cas de lésion isolée des CCE, la courbe d'accord (Figure 11A, ligne pointillée) est tout à fait superposable aux résultats de Sellick, Patuzzi & Johnstone (1982) obtenus par une mesure directe des déplacements de la membrane basilaire présentés plus haut : le profil en « V » de la courbe d'accord est très aplati, sans pointe saillante individualisée, et le niveau sonore nécessaire au déclenchement d'une réponse est très élevé. En revanche, quand la lésion concerne uniquement les CCI la courbe d'accord obtenue est différente. Il faut noter que cette situation est rare en dehors d'un contexte expérimental, les CCE étant plus fragiles et en général plus précocement et plus

massivement atteintes que les CCI. Dans ce cas (Figure 11B, ligne pointillée), on observe une diminution uniforme de la sensibilité (une élévation des seuils) de l'ordre de 40 à 50 dB, sans modification de la forme globale de la courbe. Ces résultats impliquent que les CCE sont responsables du mécanisme actif, caractérisé par la saillance de la pointe de la courbe d'accord, qui correspond à l'abaissement des seuils dans une zone fréquentielle précise. En revanche, une atteinte isolée des CCI entraine une diminution globale de l'efficacité de la transduction, avec une élévation homogène des seuils et ne dégrade pas le mécanisme actif.

En pratique clinique courante, le recueil des oto-émissions acoustiques provoquées permet d'évaluer dans une certaine mesure l'intégrité des CCE. Une diminution de l'amplitude (voir une abolition) de ces oto-émissions s'associe en général à une altération du mécanisme actif (Kemp, 2002). C'est le cas notamment en cas de surdité d'origine ototoxique ou traumato-sonore, où l'altération transitoire ou définitive des oto-émissions constitue un bon marqueur de l'état physiologique de la cochlée (Halsey, Fegelman, Raphael, Grosh, & Dolan, 2005).

4.3 Conséquences des lésions cochléaires sur la sélectivité fréquentielle

Certains des troubles auditifs fréquemment rapportés par les patients atteints de surdité endocochléaire (et notamment les troubles de la compréhension dans un milieu sonore multi-sources) sont liés, au moins en partie, à la dégradation de la sélectivité fréquentielle. La sélectivité fréquentielle est définie comme la capacité du système auditif à séparer (ou à *résoudre*) les composantes fréquentielles d'un son complexe. En pratique si l'on présente simultanément deux sons purs dont les fréquences sont suffisamment différentes (par exemple 100 Hz et 360 Hz), le sujet percevra deux objets auditifs distincts ayant chacun leur propre hauteur. Chacune des composantes sinusoïdales peut

être codée individuellement dans le nerf auditif à condition que les pics qu'elles entrainent dans le profil d'excitation de la membrane basilaire soient assez distants. Comme cela a été décrit dans la section précédente, une altération des CCE entraine un élargissement de la gamme de fréquences à laquelle répond préférentiellement un point donné de la membrane basilaire (ou une fibre du nerf auditif). Cette observation laisse supposer que deux composantes spectrales résolues par une cochlée physiologiquement saine ne le seront pas forcément en cas de lésion cochléaire ; en d'autres termes, la détérioration du mécanisme actif devrait entrainer une dégradation de la sélectivité fréquentielle.

Une des méthodes utilisées en psychoacoustique pour évaluer la sélectivité fréquentielle d'un sujet consiste à établir le profil de ses filtres auditifs. Pour cela, on peut mesurer le seuil de détection d'un son pur présenté en association avec un masquage spécifique (appelé *notched-noise*, littéralement *bruit encoché*). Ce bruit large bande présente une « encoche » spectrale de part et d'autre de la fréquence du son pur à détecter. Grossièrement, le principe est le suivant : plus les filtres auditifs sont élargis (et donc plus la sélectivité fréquentielle est dégradée), plus la quantité de bruit qui passe dans le filtre centré par la fréquence du son pur à détecter est grande. En sortie de ce filtre, le rapport signal/bruit est d'autant moins favorable qu'il est élargi, et que l'encoche spectrale du bruit est étroite. En faisant varier la largeur de l'encoche spectrale, on peut ainsi évaluer la largeur des filtres auditifs à partir de mesures de seuils de détection du son pur dans ce bruit.

De nombreux auteurs ont mesuré la largeur des filtres auditifs chez des sujets normo-entendants ou atteints de surdité d'origine cochléaire. Les résultats sont globalement homogènes et retrouvent, conformément aux prédictions, un élargissement des filtres auditifs en cas de lésions cochléaires. D'une manière générale, la largeur des filtres auditifs tend à augmenter avec le degré de surdité, même si une grande variabilité

interindividuelle empêche une prédiction fiable de la largeur des filtres à partir des seuils auditifs absolus chez un sujet donné. Glasberg & Moore ont établi par la méthode du *notched-noise*, le profil des filtres auditifs (centrés sur la fréquence 1 kHz) de l'oreille normo-entendante et de l'oreille sourde de cinq sujets atteints de déficit cochléaire unilatéral (Figure 12). Ainsi, les différences entre oreilles saines et atteintes peuvent être attribuées à des facteurs primitivement sensoriels (et non à des troubles liés à l'âge, un déficit cognitif, etc.).

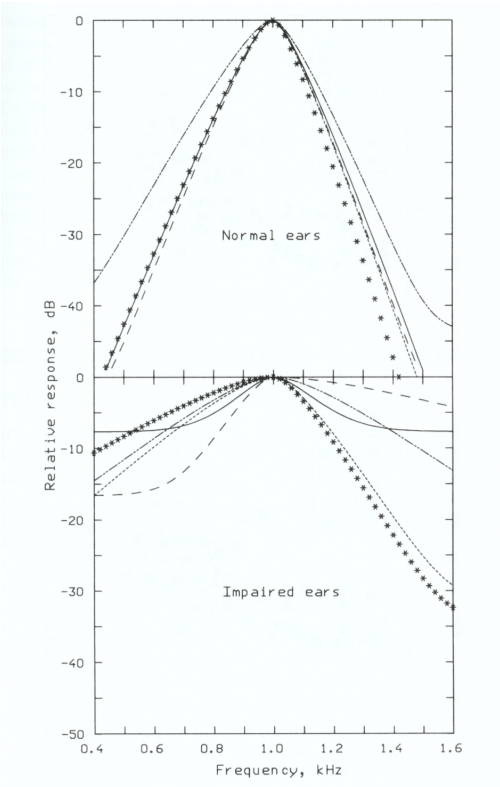


Figure 12. Profils des filtres auditifs centrés sur 1 kHz de cinq sujets présentant une audition normale sur une oreille (en haut de la figure), et une surdité d'origine cochléaire sur l'autre (en bas). La réponse est donnée relativement au sommet des courbes, arbitrairement fixé à 0 dB ; cela équivaut à considérer que l'amplitude d'un son pur d'un 1 kHz est identique en entrée et en sortie du filtre (atténuation de 0 dB). D'après Glasberg & Moore (1986).

Leurs résultats indiquent que les profils des filtres auditifs des oreilles normo-entendantes sont très homogènes entre les sujets, et possèdent une bonne sélectivité fréquentielle (filtres étroits). En revanche, en cas de lésions cochléaires, les profils sont plus variés selon les individus, avec une tendance globale d'élargissement des filtres (prédominant du coté des basses fréquences). Cela implique que les bruits de basses

fréquences seront particulièrement masquants (et donc gênants) pour ces patients. La variabilité interindividuelle est probablement le reflet de la diversité des lésions couvertes par la notion de surdité endocochléaire. Un élargissement majeur des filtres auditifs suppose une atteinte sévère des CCE, mais des lésions combinées de plusieurs structures peuvent cohabiter.

Le principe sur lequel repose le *notched-noise*, à savoir que l'élargissement d'un filtre auditif détériore le rapport signal/bruit dans ce filtre, souligne le lien entre la sélectivité fréquentielle et les troubles de la compréhension des sujets sourds en milieu bruyant. Henry & Heinz (2012) ont confirmé dans une étude électrophysiologique récente qu'en présence de bruit, l'élargissement des filtres auditifs était lié à une dégradation du codage temporel d'un son pur dans une fibre du nerf auditif répondant préférentiellement à la fréquence du son pur. La quantité de bruit passant dans le filtre auditif correspondant est trop élevée, perturbant le verrouillage de phase avec le son pur.

4.4 Conséquences des lésions cochléaires sur la sensibilité à la structure temporelle fine

A la sortie des filtres auditifs, l'onde sonore peut être analysée selon deux caractéristiques : les fluctuations lentes de l'amplitude du signal au cours du temps correspondant à l'*enveloppe*, et les fluctuations rapides correspondant à la *structure temporelle fine* (STF). Une illustration en est donnée par exemple dans la Figure 7 (à droite). Les quatre composantes fréquentielles résolues se présentent sous la forme d'une onde dont la STF fluctue à la fréquence de la composante correspondante (respectivement 200, 400, 800 et 1200 Hz) et dont l'enveloppe est « plate ». A la sortie du filtre centré sur 3200 Hz, l'interaction de plusieurs composantes entraine une fluctuation de l'enveloppe à la fréquence 200 Hz, beaucoup plus lente que la fluctuation de la STF. Le codage de la

structure temporelle fine dans le nerf auditif repose sur le verrouillage de phase qui suit les cycles individuels de fluctuations rapides de l'onde ; il est dégradé dans les hautes fréquences, classiquement au-delà de 4 à 5 kHz (cf. section 2.1). La STF apparait très importante notamment dans la perception de la hauteur des sons complexes (cf. section 2.2.4 a), où elle participe au codage des composantes résolues, mais aussi à celui de certaines composantes non résolues jusqu'au 12ème rang environ (au-delà seule la sensibilité aux variations lentes de l'enveloppe est préservée). La STF joue également un rôle important dans la compréhension de la parole, notamment en présence d'un bruit de fond fluctuant (Moore B. C., 2008).

Plusieurs études psychoacoustiques se sont intéressées à la sensibilité à la structure temporelle fine chez les sujets présentant des lésions cochléaires. Ces travaux reposaient sur des tâches de détection de différences de phase interaurales –qui dépendent de la sensibilité à la STF– (Lacher-Fougère & Demany, 2005), sur la comparaison de stimuli dont les STF différaient mais dont l'enveloppe était identique (Lacher-Fougère & Demany, 1998; Hopkins & Moore, 2007), ou encore sur la compréhension de séquences de parole dont seules les informations d'enveloppe ou de STF étaient préservées (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006). Les résultats indiquent que les surdités d'origine cochléaire s'accompagnent d'une diminution (et parfois d'une disparition) de la sensibilité à STF. De plus, chez certains sujets sourds, cette dégradation peut être retrouvée même dans une zone fréquentielle où le seuil audiométrique absolu est dans les limites de la normale (Lacher-Fougère & Demany, 2005; Lorenzi, Debruille, Garnier, Fleuriot, & Moore, 2009).

L'origine physiopathologique de la diminution de la sensibilité à la STF en cas de lésions cochléaires n'est pas clairement établie. L'hypothèse la plus « naturelle » parait celle d'une dégradation dans la précision du verrouillage de phase lui-même. Néanmoins, les données électrophysiologiques chez l'animal sont contradictoires.

Certaines études ont retrouvé que le verrouillage de phase était particulièrement altéré en cas de lésion cochléaire ototoxique ou traumato-sonore (Woolf, Ryan, & Bone, 1981; Miller, Schilling, Franck, & Young, 1997). D'autres auteurs ont des conclusions diamétralement opposées, leurs résultats ne montrant aucune atteinte particulière du verrouillage de phase dans le nerf auditif secondaire à une lésion cochléaire (Harrison & Evans, 1979; Kale, Micheyl, & Heinz, 2013). Il existe par ailleurs d'autres explications possibles à la dégradation de la sensibilité à la STF, même si le mécanisme de verrouillage de phase n'est pas primitivement dégradé. L'une d'elles, comme le suggèrent Henry & Heinz (2012) ou Moore (2008), repose sur l'idée que l'information disponible en sortie de filtres auditifs anormalement larges est plus complexe et en quelque sorte « parasitée » par le bruit de fond ou par des composantes spectrales proches, ce qui peut rendre plus difficile le décodage de l'information relative à la STF par le système auditif. Cette hypothèse est néanmoins remise en question par les données récentes de Hopkins & Moore (2011) indiquant que la sensibilité à la STF n'est que très faiblement corrélée à l'élargissement des filtres auditifs ; en revanche elle décline avec l'âge, et ce même en l'absence de déficit objectivable à l'audiométrie tonale.

5 PERCEPTION DE LA HAUTEUR EN PRESENCE D'UNE SURDITE NEUROSENSORIELLE

5.1 Perception de la hauteur des sons purs et zones mortes

Chez certains sujets atteints de surdité neurosensorielle, les cellules ciliées internes sont fonctionnellement détruites dans une ou plusieurs régions le long de la membrane basilaire. Ces zones dont le mécanisme de transduction est aboli sont donc électrophysiologiquement « éteintes » et sont classiquement qualifiées de *zones mortes*. Un test psychoacoustique connu sous le nom de TEN test (pour *Threshold-Equalizing Noise Test*) développé par Moore et al. (2000) est disponible pour diagnostiquer en pratique clinique courante la présence de zones mortes cochléaires. La prévalence de ces zones mortes est loin d'être marginale. Sur un large échantillon de 308 sujets atteints de surdité neurosensorielle moyenne à sévère testés par Vinay & Moore (2007), 57.4 % des sujets présentaient au moins une zone morte sur l'une des régions fréquentielles testées.

Imaginons que l'on présente à un sujet un son pur dont la fréquence se trouve dans une zone morte (on suppose par exemple que le sujet présente une zone morte touchant les basses fréquences). Si l'intensité du son pur est suffisante, il sera détecté par une zone de la cochlée tonotopiquement dédiée à des fréquences plus élevées. L'étude de ces zones mortes présente un grand intérêt théorique puisqu'un modèle de type spatial prédit dans ce cas que la hauteur perçue sera plus aigue, contrairement à un modèle purement temporel qui repose sur l'analyse des intervalles entre potentiels d'action dans les fibres du nerf auditif indépendamment de leur position tonotopique. Certains auteurs ont évalué les modifications de perception de la hauteur en présence de

zones mortes, par exemple en demandant à des sujets atteints d'une surdité unilatérale d'ajuster comparativement la hauteur de deux sons purs entre leur oreille saine et leur oreille sourde (Huss & Moore, 2005). Les résultats montrent que la hauteur d'un son pur tombant dans une zone morte peut effectivement être modifiée (elle est généralement perçue comme plus aigue en cas de zone morte sur les basses fréquences), mais cela n'est pas systématique et dans certains cas elle est perçue « normalement ». De plus, les réponses sont relativement variables chez un même sujet, ce qui reflète probablement que les sons purs tombant dans les zones mortes évoquent une sensation de hauteur imprécise et peu saillante. Au total, l'effet des zones mortes sur la perception de la hauteur laisse penser que l'information temporelle n'est pas en elle-même suffisante pour coder les sons de basse fréquence, et qu'il faut probablement une concordance correcte entre l'information temporelle disponible dans une fibre et sa « position tonotopique » pour que la hauteur soit perçue normalement (Moore B. C., 2007).

Les lésions cochléaires peuvent entrainer des modifications de la perception de hauteur des sons purs même en l'absence de zones mortes. Une des explications repose sur le fait que le fonctionnement des cellules ciliées internes dans certaines zones de la cochlée peut être altéré sans être totalement aboli. D'une manière générale et conformément à la théorie spatiale, la hauteur perçue d'un son pur tend à se décaler vers les fréquences aigues en cas de surdité sur les basses fréquences et inversement. Néanmoins, ce type de résultat n'est pas constant et chez certains sujets le contraire peut être observé.

5.2 Perception de la hauteur des sons complexes en présence de lésions cochléaires

Les surdités neurosensorielles se caractérisent par une dégradation de la précision du codage de l'information portée par les composantes fréquentielle d'un son complexe, à la fois dans le domaine spectral et dans le domaine temporel. Dans le domaine *temporel*, les lésions cochléaires s'accompagnent généralement d'une diminution de la sensibilité à la structure fine de l'onde sonore. Même si les mécanismes physiopathologiques mis en jeu ne sont pas formellement identifiés, cela peut témoigner notamment d'une altération du verrouillage de phase ou d'une altération dans l'utilisation des informations issues du verrouillage de phase (cf. section 4.4).

Dans le domaine *spectral*, la dégradation du mécanisme actif (secondaire aux lésions des cellules ciliées externes) entraine une diminution de la sélectivité fréquentielle liée à l'élargissement des filtres cochléaires. Les harmoniques de faibles rangs sont normalement très bien résolues par la cochlée d'un auditeur sain, et elles sont essentielles dans la détermination de la hauteur du son complexe (on parle de région dominante, cf. section 2.2.3). En cas de surdité neurosensorielle associée à un élargissement des filtres cochléaires, les harmoniques de faibles rangs seront moins bien (voire pas du tout) résolues. La Figure 13 (Moore B. C., 2007) représente le pattern d'excitation d'un son complexe composé des harmoniques de rangs 4 à 10 d'une fréquence fondamentale de 200 Hz, chez un auditeur sain (trait plein) et en cas de filtres cochléaires trois fois plus larges que la normale (trait pointillé). Les pics successifs correspondant à chaque harmonique dans le pattern d'excitation « normal » disparaissent complètement en cas d'élargissement des filtres.

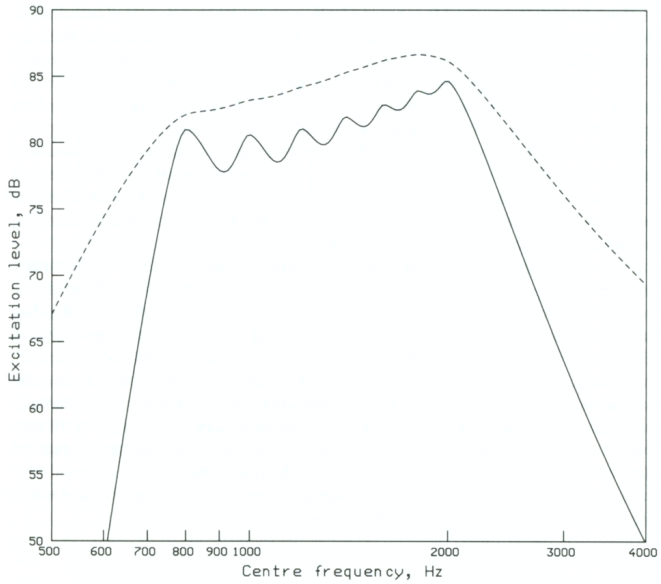


Figure 13. Représentation de patterns d'excitation d'un son complexe composé des harmoniques de rangs 4 à 10 d'une fréquence fondamentale de 200 Hz. La modélisation suppose que les filtres auditifs sont de largeur normale (trait plein), ou trois fois plus larges que la normale (trait pointillé), comme on pourrait le retrouver en présence de lésions cochléaires. Le niveau de chaque composante fréquentielle est de 80 dB SPL. L'individualisation de pics successifs dans le pattern d'excitation correspondant aux différentes harmoniques résolues n'est pas retrouvée dans le cas où les filtres auditifs sont élargis. D'après Moore (2007).

Chez un auditeur sain, un son complexe ne contenant que des harmoniques non résolues est généralement associé à une sensation de hauteur peu précise et peu saillante. On peut donc s'attendre à ce que la diminution de la sélectivité fréquentielle en cas de lésion cochléaire s'accompagne d'une perception plus ambiguë et d'une dégradation de la discrimination de hauteur des sons complexes. De nombreuses études ont évalué les performances de discrimination de hauteur chez des sujets sourds, le plus souvent en mesurant la plus petite différence de f0 détectable entre deux sons complexes successifs (Moore & Peters, 1992; Arehart, 1994; Moore & Moore, 2003). Comme pour beaucoup d'autres aspects de la perception auditive en cas de surdité neurosensorielle, il existe une grande variabilité interindividuelle des résultats, que ce soit dans le niveau de performance globale ou dans l'effet des modifications spectrales des stimuli. Néanmoins, plusieurs résultats importants se dégagent de ces travaux :

(1) Chez certains sujets, la discrimination de hauteur est paradoxalement meilleure pour des sons complexes contenant seulement des harmoniques de rangs élevés (>5) que

pour des sons ne contenant que les harmoniques de faibles rangs. On peut même parfois observer que l'ajout d'harmoniques de faibles rangs à un complexe contenant des harmoniques de rangs élevés *dégrade* les performances. Cela suggère que chez ces sujets sourds, la hauteur est principalement déterminée par les harmoniques de rangs élevés et que l'interaction entre les harmoniques de faibles rangs traitées par des filtres auditifs anormalement larges (qui induit des réponses plus complexes en sortie de filtres) pourrait être à l'origine de la diminution des performances.

(2) Etonnamment, les performances de discrimination de hauteur ne sont que faiblement corrélées à la sélectivité fréquentielle. Les sujets présentant des seuils de discrimination de f0 très élevés n'ont pas nécessairement un élargissement important des filtres auditifs.

Une caractéristique constante a néanmoins été relevée par Oxenham (2008) : tous les sujets dont les filtres auditifs sont anormalement larges présentent une dégradation dans la discrimination de f0 de sons complexes. Il semble donc exister une relation assez forte entre la largeur des filtres auditifs et la perception de la hauteur des sons complexes, même si les études citées ci-dessus mesurant directement le seuil de discrimination de f0 ne l'ont pas clairement mise en évidence. Bernstein & Oxenham (2006) ont eu l'idée de s'intéresser à la dégradation brutale dans les performance de discrimination de f0 typiquement retrouvée chez les auditeurs sains lorsque l'on retire successivement les harmoniques de faibles rangs d'un son complexe. Il est admis que cette dégradation apparait lorsque le son complexe ne contient plus suffisamment d'harmoniques résolues pour permettre une estimation précise de f0. On peut logiquement s'attendre à ce que ce *point de rupture* dans les performances de discrimination soit atteint plus rapidement chez les sujets dont les filtres auditifs sont élargis, car le nombre d'harmoniques correctement résolues est moindre. Les auteurs ont effectivement retrouvé une corrélation significative entre le rang de l'harmonique

correspondant au point de rupture et la largeur des filtres auditifs du sujet (le rang était d'autant plus faible que les filtres auditifs étaient élargis). Ces données indiquent que la diminution de la sélectivité fréquentielle en cas de lésion cochléaire affecte directement la perception de la hauteur des sons complexes.

Au total, les données disponibles indiquent que les sujets atteints de surdité neurosensorielle ont des troubles de la perception de la hauteur des sons complexes se manifestant par une sensation de hauteur moins précise et une dégradation des performances de discrimination de f0. Au moins deux hypothèses peuvent être avancées pour expliquer ces anomalies. D'une part la diminution de la sélectivité fréquentielle limite la capacité du système auditif à identifier les composantes spectrales de faibles rangs, et d'autre part la réduction de la sensibilité à la structure temporelle fine dégrade l'information issue des harmoniques résolues et de certaines harmoniques non résolues.

Ces troubles peuvent potentiellement avoir des conséquences importantes sur la capacité du système auditif à fusionner les composantes fréquentielles d'un son complexe harmonique. Pourtant, à notre connaissance, il n'existe à ce jour aucune donnée expérimentale sur la sensibilité à la fusion harmonique chez les sujets présentant une surdité neurosensorielle ; l'étude présentée dans la section suivante visait principalement à combler cette lacune.

6 CONSEQUENCES DES SURDITES NEUROSENSORIELLES SUR LA SENSIBILITE A L'HARMONICITE

6.1 Introduction

Les surdités de perception –ou neurosensorielles– dominées par la presbyacousie, constituent la cause la plus fréquente de déficience auditive chez l'adulte, bien loin devant les surdités de transmission. Les résultats de l'étude américaine EHLS (*Epidemiology of Hearing Loss Study*) portant sur plus de 3700 personnes âgées de 48 à 92 ans, illustrent les enjeux de santé publique que représente ces pathologies (Cruickshanks, et al., 1998). La prévalence des surdités de perception (définies ici par un seuil audiométrique moyen > 25 dB sur les fréquences 500, 1000, 2000 et 4000 Hz et un Rinne audiométrique < 15 dB) atteignait dans cette cohorte 37.2%, soit dix fois plus que les surdités de transmission (3.7%).

En pratique clinique courante, l'une des plaintes principales des patients présentant une surdité neurosensorielle est la difficulté à comprendre la parole dans un environnement sonore multi-sources. Malgré les progrès importants des appareillages durant ces dernières années, les patients porteurs de prothèses auditives (aide auditive conventionnelle ou implant cochléaire) rapportent également la plupart du temps une gêne dans ce type de situation (réunion, repas au restaurant, etc.). Un auditeur normal, par contre, possède une remarquable capacité à séparer, sans effort, les productions de sources sonores actives simultanément. La réussite de cette « analyse des scènes auditives » (ASA) tient en partie au fait que les éléments successifs produits par une

source donnée possèdent en général des traits physiques qui permettent de distinguer cette source des autres. Il est maintenant bien établi que, chez les auditeurs normo-entendants, l'ASA obéit à deux principes fondamentaux. Le premier veut que des composantes spectrales distinctes mais apparaissant au même instant proviennent d'une même source plutôt que de sources distinctes. Le second veut que des composantes spectrales ayant des relations harmoniques entre elles proviennent également d'une même source plutôt que de sources distinctes. Il est probable que la sensibilité à l'harmonicité constitue, dans la vie quotidienne d'un auditeur normal, un outil important pour l'ASA (cf. Section 2.2.2). Mais paradoxalement, alors que l'on sait que les surdités neurosensorielles rendent l'ASA déficitaire, et alors que l'on sait aussi qu'elles dégradent la capacité à différencier des fréquences voisines, aucune publication à ce jour n'a été consacrée aux conséquences des lésions cochléaires sur la sensibilité à l'harmonicité. Les données disponibles concernant d'une part les mécanismes de la fusion harmonique et d'autre part les conséquences perceptives des surdités neurosensorielles ne permettent pas de prévoir les résultats de l'étude. S'il parait probable qu'un déficit massif dans la discrimination fréquentielle doit entièrement abolir la sensibilité à l'harmonicité, l'effet d'un déficit relativement léger est par contre difficile à prédire actuellement. Cette prédiction dépend en effet de la conception que l'on se fait du mécanisme par lequel un rapport de fréquences entre deux sons purs est auditivement reconnu comme harmonique ou inharmonique, c'est à dire du type de modèle envisagé (cf. Section 2.2.4).

Dans le cas d'un modèle de reconnaissance de pattern, la détection de l'inharmonicité chez les sujets normo-entendants est vraisemblablement limitée seulement par la précision des gabarits, et non par la précision de l'encodage périphérique de la fréquence. En effet, les seuils de détection de l'inharmonicité sont nettement plus grands (i.e., moins bons) que les seuils différentiels de fréquence (Moore, Peters, & Glasberg, 1985). En cas de surdité neurosensorielle, la précision de l'encodage

de la fréquence se dégrade, mais en théorie cette précision peut néanmoins rester meilleure que la précision des gabarits. Si tel est le cas, une surdité neurosensorielle n'aurait pas d'effet délétère sur la détection d'inharmonicité tant que la précision de l'encodage fréquentiel n'est pas très altérée.

Dans le cas des modèles où l'information temporelle est prédominante (comme les modèles de type autocorrélation), le système auditif central détecte que deux fréquences sont en relation harmonique grâce à la similarité entre les informations temporelles véhiculées en réponse à ces deux fréquences. Si tel est le cas, une dégradation dans l'encodage temporel de la fréquence des sons purs devrait avoir pour conséquence une dégradation de la sensibilité à l'harmonicité. Si donc, comme on peut le penser aujourd'hui, les surdités neurosensorielles dégradent l'encodage temporel de la fréquence des sons, elles devraient se traduire par un affaiblissement de la détection d'inharmonicité.

Nous avons donc conçu une étude permettant de comparer la sensibilité à l'harmonicité entre un groupe de sujets normo-entendants et un groupe de sujets atteints de surdités neurosensorielles légères à moyennes. La méthodologie s'inspire de travaux antérieurs réalisés par notre équipe au sujet de la sensibilité à l'harmonicité chez les sujets normo-entendants (Demany & Semal, 1988; Demany & Semal, 1992). Ces travaux ont montré que le seuil différentiel d'un rapport de fréquences entre sons purs simultanés est plus petit (donc meilleur) lorsque le rapport étalon est égal à exactement une octave (2/1) que lorsque le rapport étalon est légèrement plus grand ou légèrement plus petit qu'une octave. Cet effet reflète de façon objective une « reconnaissance » par le système auditif de l'harmonicité du rapport d'octave. Nous cherchons à déterminer dans quelle mesure l'effet en question reste présent chez les sujets atteints de surdité neurosensorielle. Comme cela a été souligné plus haut dans cette thèse, des précautions doivent être prises pour que les sons purs composant les stimuli n'interagissent pas au

niveau de la cochlée. En effet, si des interactions existent, l'octave peut être discriminée de rapports fréquentiels voisins sur la base d'indices de « battement » produits au sein d'un seul filtre cochléaire (cf. Section 3.2). Afin d'éliminer les interactions cochléaires, les sons purs sont présentés dans un bruit connu sous le nom de *Threshold-Equalizing Noise* (*T.E. Noise*) dont le spectre est uniformément masquant pour des sujets normo-entendants (Moore, Huss, Vickers, Glasberg, & Alcántara, 2000).

Le nombre de sujets sourds que nous avons pu tester à ce jour et dont les données sont exploitables reste limité, les inclusions et le recueil des données étant toujours en cours. Nous présentons donc ici les résultats préliminaires, dont certains aspects seront complétés dans la version définitive de l'étude soumise à publication.

6.2 Matériel et méthodes

6.2.1 Participants et critères d'inclusion

Dix-sept sujets répartis en deux groupes (un groupe de sujets normo-entendants et un groupe de sujets atteints de surdité neurosensorielle) ont participé à l'étude à ce jour. Les volontaires âgés de plus de 60 ans sont exclus. Tous les sujets (à l'exception de l'auteur de cette thèse) sont rémunérés pour leur participation. Le premier groupe se compose de neuf sujets normo-entendants (perte tonale moyenne < 20 dB entre 125 et 8000 Hz), dont l'auteur de cette thèse et huit autres sujets recrutés parmi les étudiants et le personnel du Centre Hospitalier Universitaire de Bordeaux.

Le second groupe se compose de huit patients suivis dans le service d'oto-rhino-laryngologie et présentant (sur l'oreille choisie) une surdité neurosensorielle légère à moyenne selon la classification audiométrique des déficiences auditives (cf. section 4.1). Les critères audiométriques d'inclusion (précisés sur la Figure 14) ne retiennent que des pertes audiométriques relativement uniformes de 30 à 50 dB entre 250 et 3000 Hz. Tous les stimuli étant présentés unilatéralement sans masquage controlatéral, les seuils audiométriques de l'oreille controlatérale doivent être supérieurs (*i.e.* moins bons) à ceux de l'oreille testée. Les patients dont le bilan diagnostique a révélé des éléments en faveur d'une atteinte rétrocochléaire sont exclus (notamment une abolition des réflexes stapédiens, un allongement des latences I-III et I-V sur les potentiels évoqués auditifs ou une lésion du nerf cochléo-vestibulaire sur l'imagerie par résonnance magnétique). Malheureusement, les résultats de trois des huit sujets sourds testés se sont révélés inexploitables pour des raisons exposées dans la section 6.2.3. Les données expérimentales du groupe de patients sourds présentées dans les sections suivantes portent donc sur cinq sujets, dont les seuils audiométriques sont indiqués dans la Figure 14.

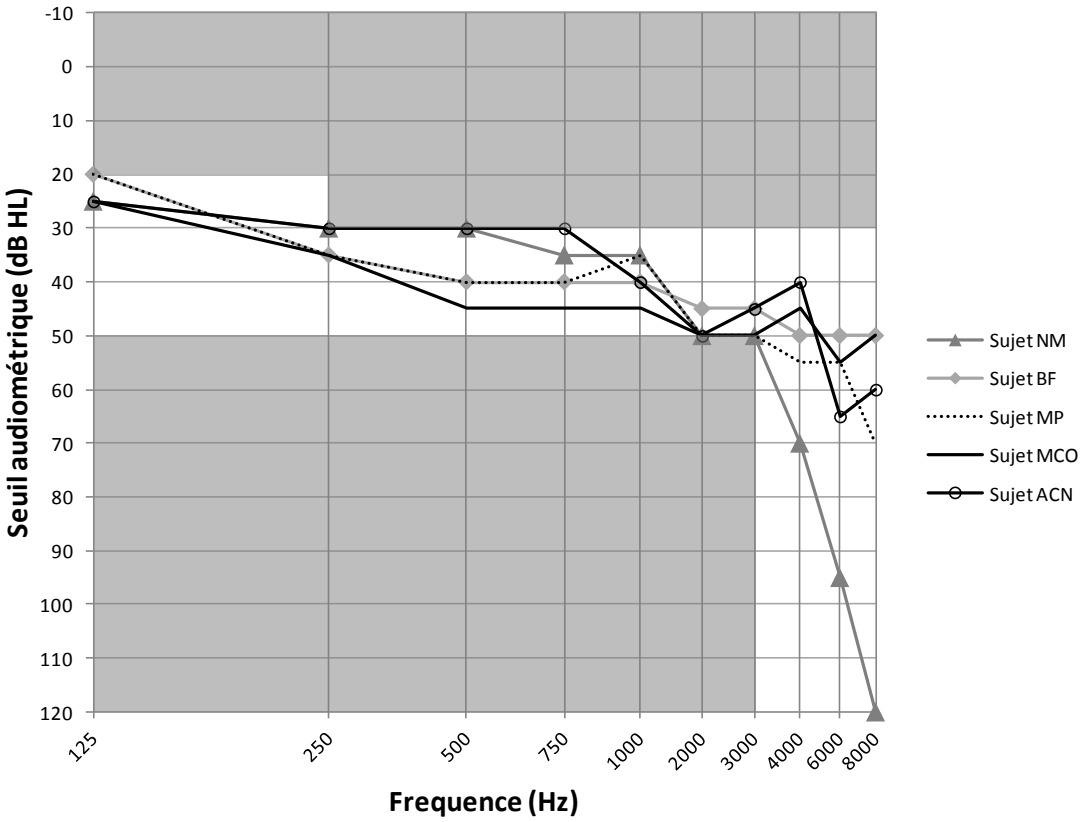


Figure 14. Audiométrie tonale des cinq sujets atteints de surdité dont les données sont exploitables (respectivement nommés NM, BF, MP, MCO et ACN). Les courbes indiquent les seuils absolus en conduction aérienne mesurés sur la meilleure oreille pour 10 fréquences comprises entre 125 et 8000 Hz. Les critères audiométriques d'inclusion correspondent à la zone non grisée sur l'audiogramme, soit une perte de 20 à 50 dB pour les fréquences inférieures à 250 Hz, une perte de 30 à 50 dB entre 250 et 3000 Hz, et une perte ≥ 30 dB pour les fréquences supérieures à 3000 Hz.

L'âge moyen des sujets normo-entendants est de 37 ans (avec des extrêmes allant de 22 à 60 ans), celui des sujets sourds de 53 ans (avec des extrêmes de 45 à 60 ans). Cette différence est statistiquement significative, $t(12) = 2.7$, $P = 0.022$.

6.2.2 Stimuli, tâche et conditions

Chaque stimulus, d'une durée de 200 ms, est composé de trois sons purs de même amplitude présentés unilatéralement et simultanément, dont les fréquences respectives sont égales à F, (2^*oct*).F et [(2^*oct*)^2].F. En d'autres termes, les trois sons purs sont séparés par un intervalle harmonique d'exactement une octave quand le paramètre *oct* est égal à 1, et le degré d'inharmonicité est modifié en faisant varier la

valeur de *oct* autour 1. A chaque essai expérimental, on présente au sujet quatre stimuli, répartis temporellement en deux paires (cf. Figure 15). Chaque paire de stimuli est présentée dans une séquence de bruit continu de 1200 ms (*T.E. Noise* dont le spectre s'étend de 100 à 4000 Hz), débutant 200 ms avant le premier stimulus et se terminant 200 ms après le deuxième. Un silence de 300 ms sépare les deux séquences de bruit. Le paramètre *oct* prend une même valeur dite *étalon* pour trois des stimuli, et une valeur différente dite *cible* pour le second stimulus de l'une des deux paires (choisie au hasard).

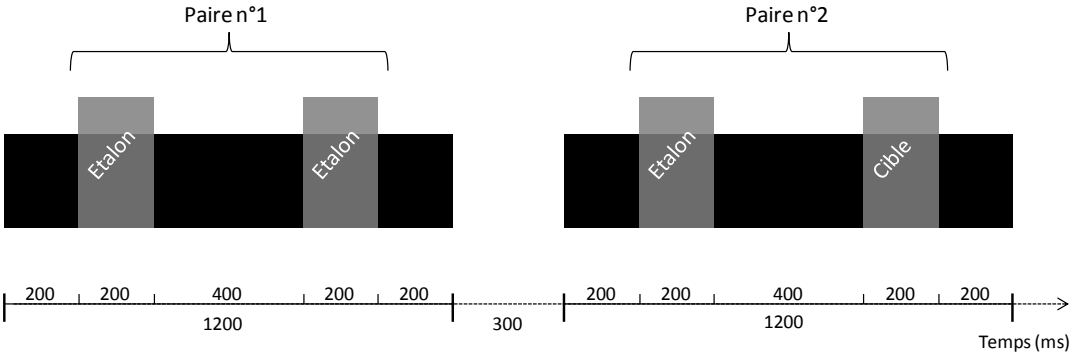
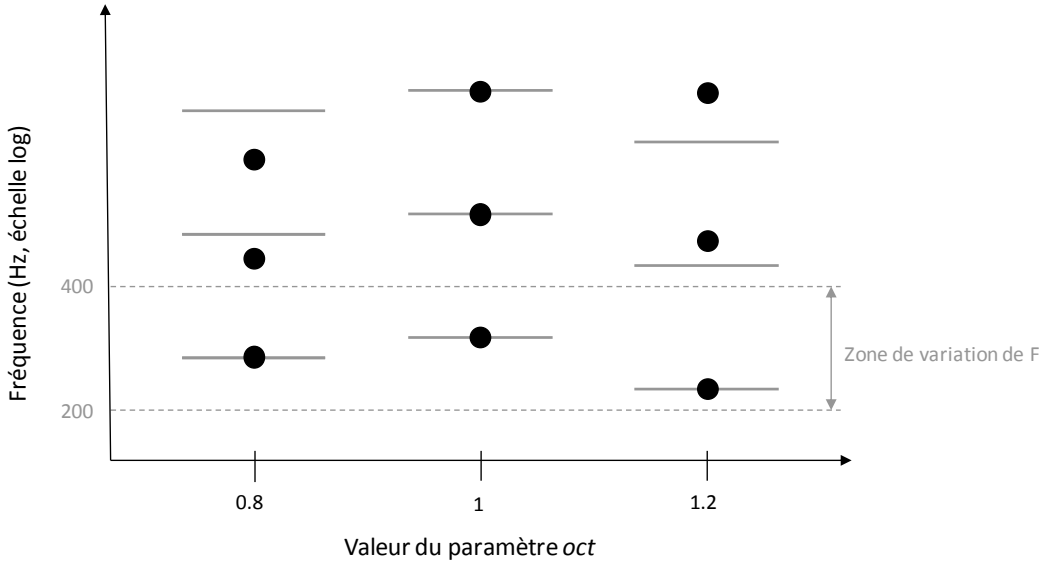


Figure 15. Représentation schématique d'un essai expérimental. Les séquences de bruit sont symbolisées par les rectangles noirs, les stimuli par les rectangles gris. A chaque essai, le sujet doit indiquer dans quelle paire se trouve le stimulus cible, qui est toujours le second stimulus de l'une des deux paires, choisie au hasard (dans cette exemple, la paire n°2). Les durées (en ms) des différents éléments sont indiquées sur l'axe situé en bas de la figure.

La tâche du sujet est d'identifier la paire contenant le stimulus « intrus » (cible), en cliquant sur le bouton correspondant sur un écran d'ordinateur, qui affiche en retour un *feedback* visuel. Dans chaque bloc d'essais, la valeur étalon de *oct* est fixe alors que la cible est variée, avec une procédure « adaptative » automatisée (Kaernbach, 1991; García-Pérez, 2011), de façon à estimer quelle valeur doit prendre la cible pour que la probabilité d'une réponse juste soit égale à 0.78. La différence entre cette valeur et la valeur étalon est retenue comme seuil différentiel pour l'étalon en question.

Chaque sujet est testé dans quatre conditions expérimentales. Dans trois de ces conditions, la valeur de F varie au hasard d'un stimulus à l'autre, entre 200 et 400 Hz. A cause du caractère aléatoire de cette variation, les sujets ne peuvent pas identifier le

stimulus cible autrement que sur la base du paramètre *oct*. En d'autres termes, les sujets sont forcés de juger des rapports de fréquence, plutôt que seulement des fréquences. Dans la quatrième condition, par contre, F prend la même valeur au sein de chaque paire de stimuli (tout en variant cependant aléatoirement d'une paire à l'autre). Il est alors possible d'identifier la cible en détectant seulement une différence entre les deux stimuli de la paire dont elle fait partie. Cette quatrième condition nous permet d'évaluer, chez chaque sujet, la précision de l'encodage de la fréquence. La valeur étalon de *oct* est fixée à 1 dans la quatrième condition, notée condition *1**. Dans les trois autres conditions, les valeurs étalons de *oct* sont respectivement égales à 0.8, 1, et 1.2 (cf. Figure 16). Ces conditions sont respectivement notées conditions *0.8*, *1* et *1.2*. Comparer les seuils différentiels obtenus dans ces trois conditions permet d'évaluer la sensibilité du sujet à l'harmonicité du rapport d'octave.



09

10

11

Figure 16. Représentation schématique du spectre de trois stimuli étalons possibles correspondant respectivement aux conditions *0.8*, *1* et *1.2*. Les trois sons purs composants chaque stimulus sont symbolisés par des points noirs. Les échelons gris matérialisent des intervalles d'exactement une octave. La zone de variation de la composante de plus basse fréquence (F) est comprise entre les deux lignes pointillées.

6.2.3 Choix du niveau d'intensité

Le choix du niveau d'intensité et du rapport signal/bruit est un point crucial du protocole. En effet, il faut s'assurer que chaque composante des stimuli soit présentée à un niveau suffisant pour être correctement perçue, tout en limitant au maximum la possibilité d'interaction périphérique entre composantes, qui augmente avec le rapport signal/bruit. Compte tenu des seuils audiométriques des sujets sourds, le bruit est présenté chez tous les sujets à un niveau global de 62 dB SPL (à l'exception du sujet sourd NM, cf. ci-dessous). Le seuil de détection, dans ce bruit, de sons purs variés en fréquences de 200 à 3000 Hz est déterminé au début de l'expérimentation par une procédure adaptative.

L'utilisation du *T.E. Noise* a permis d'obtenir comme prévu, chez les sujets normo-entendants, un seuil de détection ne variant quasiment pas en fonction de la fréquence. Dans la suite de l'expérience, l'amplitude des sons purs utilisés est fixée pour ces sujets à un niveau de sensation nominal de 15 dB SL (dans une première série de mesures), puis de 5 dB SL (dans une seconde série).

Chez les sujets sourds, le caractère uniformément masquant du *T.E. Noise* s'est révélé moins systématique, ce qui peut refléter notamment la présence de zones mortes cochléaires. Si la variation du seuil de détection en fonction de la fréquence est <5 dB, l'amplitude des sons purs est fixée à 15 dB au dessus du seuil de détection moyen, comme chez les sujets normo-entendants. Si la variation du seuil de détection est comprise entre 5 et 10 dB, l'amplitude des sons purs est fixée à 5 dB au dessus du plus mauvais seuil (ce qui garantit que tous les sons purs sont perçus à un niveau nominal de sensation compris entre 5 et 15 dB SL). Si la variation du seuil est >10 dB, il devient difficile de déterminer un niveau d'intensité garantissant à la fois une détectabilité correcte des composantes et l'absence d'interaction pour l'ensemble des fréquences concernées. Dans ce cas, le seuil de détection est mesuré à nouveau après une

augmentation du niveau du bruit. Pour le sujet NM, un bruit à 70 dB SPL a permis de ramener la variation du seuil de détection à une valeur comprise entre 5 et 10 dB. Pour trois sujets sur les huit inclus initialement, la variation du seuil est restée > 10 dB après plusieurs augmentations successives du niveau du bruit, ce qui rendait impossible la poursuite des mesures.

6.2.4 Procédure

L'expérience se déroule dans une cabine insonore, en quatre à sept séances d'environ une heure et demie chacune. La première séance est consacrée à la réalisation d'une audiométrie tonale liminaire par voie aérienne (selon la méthode des niveaux ascendants), et à la détermination du niveau d'intensité auquel sont présentés les stimuli selon la méthode détaillée ci-dessus. Le coté retenu pour la suite de l'expérimentation est laissé au libre choix pour les auditeurs sains, et dépend des critères audiométriques d'inclusion pour les sujets sourds (cf. section 6.2.1). Pendant les trois séances suivantes, les sujets réalisent entre 3 et 12 blocs d'essais consécutifs dans chacune des quatre conditions (*0.8*, *1*, *1.2* et *1**). L'ordre des conditions peut être *1-0.8-1.2-1** ou *1-1.2-0.8-1**. De cette façon, on limite au maximum pour la condition *1* l'avantage d'un éventuel effet d'apprentissage, et d'autre part on s'assure que les sujets étaient pleinement entrainés à la tâche lors de l'évaluation de la précision de leur encodage fréquentiel (condition *1**).

Dans les conditions *1* et *1**, la valeur cible initiale de *oct* est alternativement supérieure ou inférieure à la valeur étalon (en d'autres termes, on mesure alternativement le seuil de détection d'une inharmonicité résultant d'une augmentation ou d'une diminution de la taille de l'intervalle par rapport à une octave). En revanche, dans la condition *0.8*, la valeur cible initiale de *oct* est toujours inférieure à la valeur étalon ; dans la condition *1.2*, elle est toujours supérieure à la valeur étalon. Cela permet

d'éviter la situation difficilement interprétable d'un stimulus étalon inharmonique à discriminer d'un stimulus cible parfois harmonique, parfois inharmonique (si la valeur cible de *oct* oscille autour de 1).

A l'issue de ces quatre séances, l'expérience est terminée pour les sujets sourds. Les sujets normo-entendants effectuent trois séances supplémentaires, en tout point comparables aux trois séances précédentes, à l'exception du niveau d'intensité des sons purs fixé à 5 dB SL et non plus 15 dB SL.

6.3 Résultats

6.3.1 Sensibilité à l'harmonicité

Les résultats obtenus dans les conditions *0.8*, *1* et *1.2* sont représentés dans la Figure 17. Les symboles gris indiquent les moyennes des seuils différentiels obtenus chez les sujets normo-entendants pour des stimuli respectivement présentés à 15 dB SL et 5 dB SL par composante. Les symboles noirs indiquent les résultats individuels des cinq sujets atteints de surdité neurosensorielle ayant passé l'intégralité de l'expérience. Le niveau du hasard est représenté par les pointillés noirs.

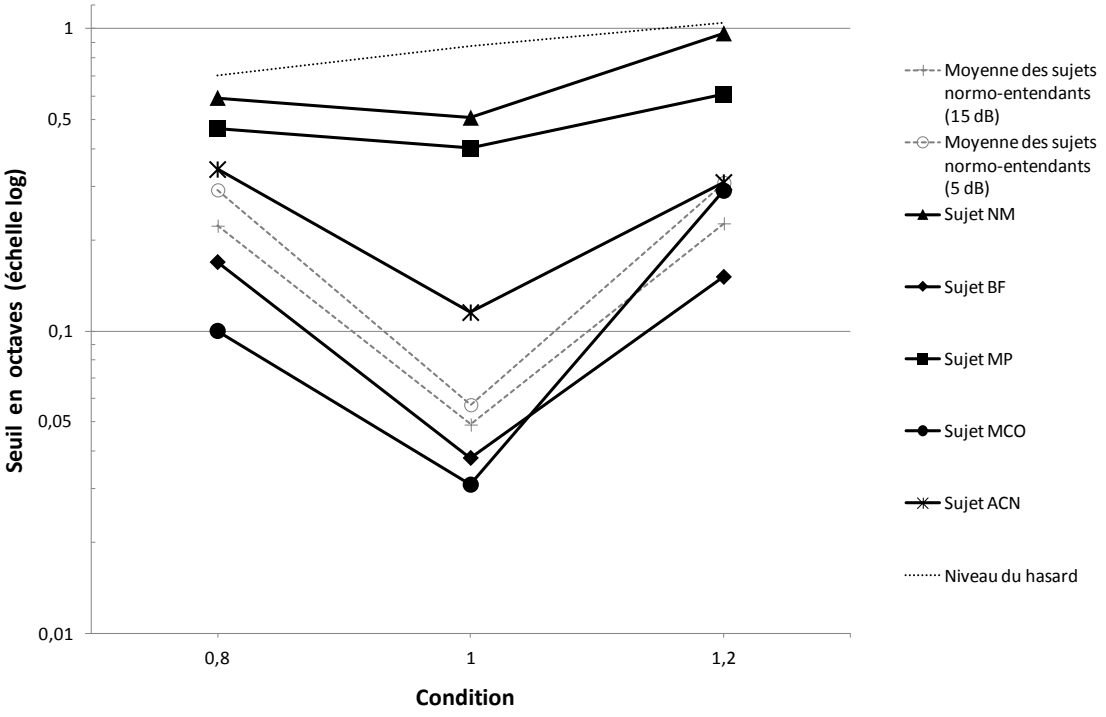


Figure 17. Seuils différentiels (en octaves) obtenus avec les trois valeurs étalons de *oct* (0.8, 1 et 1.2). La moyenne des résultats du groupe des sujets normo-entendants est indiquée pour les deux niveaux d'intensité des stimuli, 5 et 15 dB SL (pointillés gris). Les résultats individuels des cinq sujets atteints de surdité (NM, BF, MP, MCO et ACN) sont notés en noir gras. Le niveau du hasard (obtenu par une simulation d'un million de blocs d'essais où un sujet virtuel répondrait systématiquement au hasard) est indiqué par les pointillés noirs.

Le seuil différentiel entre la valeur étalon et la valeur cible de *oct* s'avère systématiquement meilleur (*i.e.* plus petit) dans le cas où l'étalon est harmonique (condition *1*) que dans le cas où il est inharmonique (conditions *0.8* et *1.2*). Cela se

traduit graphiquement par des courbes en forme de « V », dont le minimum correspond à la condition *1* (étalon harmonique). Ce profil en « V » indique que chez tous les sujets testés, aussi bien les sujets normo-entendants que ceux atteints de surdité neurosensorielle, le système auditif présente une sensibilité particulière à l'intervalle de fréquence harmonique –en l'occurrence, égal à une octave– par rapport à des intervalles relativement proches mais inharmoniques (0.8 et 1.2 octave). On peut qualifier cette sensibilité à l'harmonicité de *relative*, puisqu'elle témoigne du seuil obtenu avec un étalon harmonique *par rapport* à ceux obtenus avec des étalons inharmoniques. Cette sensibilité relative à l'harmonicité correspond graphiquement à la profondeur du « V » observé sur la Figure 17. Elle doit être différenciée de la sensibilité *absolue* à l'harmonicité, qui correspond quant à elle à la valeur du seuil obtenu dans la condition *1* considérée isolément (et donc graphiquement à la position de la « pointe du V » sur l'axe des ordonnées).

Chez les sujets normo-entendants, lorsque les stimuli sont présentés à un niveau de 15 dB SL par composante, les seuils différentiels moyens dans les conditions *0.8*, *1* et *1.2* sont respectivement égaux à 0.23, 0.049 et 0.23 octaves (erreurs standards 0.013, 0.004, 0.017). Lorsque les stimuli sont présentés à 5 dB SL, les résultats semblent uniformément dégradés, avec des seuils différentiels moyens respectivement égaux à 0.29, 0.057 et 0.31 octaves (erreurs standards 0.019, 0.005, 0.024). Les données, exprimées sur une échelle logarithmique, ont été soumises à une analyse de variance à mesures répétées dont le facteur aléatoire était le facteur *Sujet* (9 modalités) et les facteurs fixes étaient le facteur *Condition* (3 modalités ; 0.8, 1 et 1.2) et le facteur *Niveau d'intensité* (2 modalités ; 5 et 15 dB SL). Cette analyse fait apparaître l'existence d'un effet systématique très fort du facteur *Condition* ($F(2,16) = 95.9$, $P < 0.0001$), un effet significatif du facteur *Niveau d'intensité* ($F(1,16) = 8.9$, $P = 0.017$), et l'absence d'interaction significative entre ces deux facteurs ($F(2,16) = 0.9$, $P > 0.4$).

Chez les sujets atteints de surdité, il existe une variabilité interindividuelle marquée dans les résultats. Les deux sujets BF et MCO ont obtenu des seuils différentiels comparables à ceux de certains sujets normo-entendants, et des résultats globalement meilleurs que la moyenne des auditeurs sains. L'analyse statistique de leurs données individuelles (exprimées sur une échelle logarithmique) révèle comme chez les sujets normo-entendants un très fort effet du facteur *Condition*, principalement lié à une différence très significative entre les seuils obtenus dans la condition *1* et ceux obtenus dans les conditions *0.8* et *1.2* ($t(9) = 11.3$, $P < 0.0001$ pour le sujet BF ; $t(11) = 9.8$, $P < 0.0001$ pour le sujet MCO). Le sujet ACN présente des seuils dégradés par rapport au niveau moyen des sujets normo-entendants, mais là encore les seuils obtenus dans la condition 1 sont significativement meilleurs que dans les conditions 0.8 et 1.2 ($t(15) = 6.3$, $P < 0.0001$).

Les deux autres sujets sourds (NM et MP) présentent des seuils différentiels très dégradés dans toutes les conditions, mais les courbes semblent néanmoins conserver un profil en « V » qui témoignerait de la persistance d'une de la sensibilité relative à l'harmonicité du rapport d'octave. Nous avons cherché à vérifier si cette observation graphique se confirme par l'analyse statistique. On retrouve effectivement chez ces deux sujets une différence significative entre les seuils obtenus dans la condition *1* et ceux obtenus dans les conditions *0.8* et *1.2* considérées conjointement ($t(17) = 4.0$, $P = 0.001$ pour le sujet NM ; $t(39) = 2.2$, $P = 0.03$ pour le sujet MP). Néanmoins, en considérant les conditions deux à deux, aucune différence significative n'apparaît entre les conditions *0.8* et *1* ($t(17) = 1.2$, $P > 0.2$ pour le sujet NM ; $t(39) = 1.3$, $P > 0.2$ pour le sujet MP), ce qui remet en cause la vraisemblance du profil en « V » constaté graphiquement. Il convient cependant de prendre en compte le fait que le niveau du hasard est plus bas dans la condition *0.8* que dans la condition *1*. Le seuil des sujets dans la condition *0.8* a pu être ainsi « artificiellement » limité par cet effet plafond, et d'une certaine manière,

sous-estimé. Pour limiter cet écueil, nous avons considéré non pas le niveau absolu du seuil, mais l'écart entre le seuil obtenu et le niveau du hasard dans la condition correspondante. Dans ce cas, l'analyse statistique retrouve effectivement que cet écart est significativement plus grand dans la condition *1* que dans la condition *0.8* ($t(17) = 3.8$, $P = 0.001$ pour le sujet NM ; $t(39) = 3.3$, $P = 0.002$ pour le sujet MP). Cela indique qu'il persiste chez ces deux sujets une sensibilité relative objectivable à l'harmonicité du rapport d'octave.

6.3.2 Effet de la variation aléatoire de F

Comme indiqué dans la section 6.2.2, la condition *1** dans laquelle la valeur de F est la même au sein de chaque paire permet d'évaluer la capacité des sujets à discriminer des fréquences. On peut donc considérer que les seuils différentiels obtenus dans cette

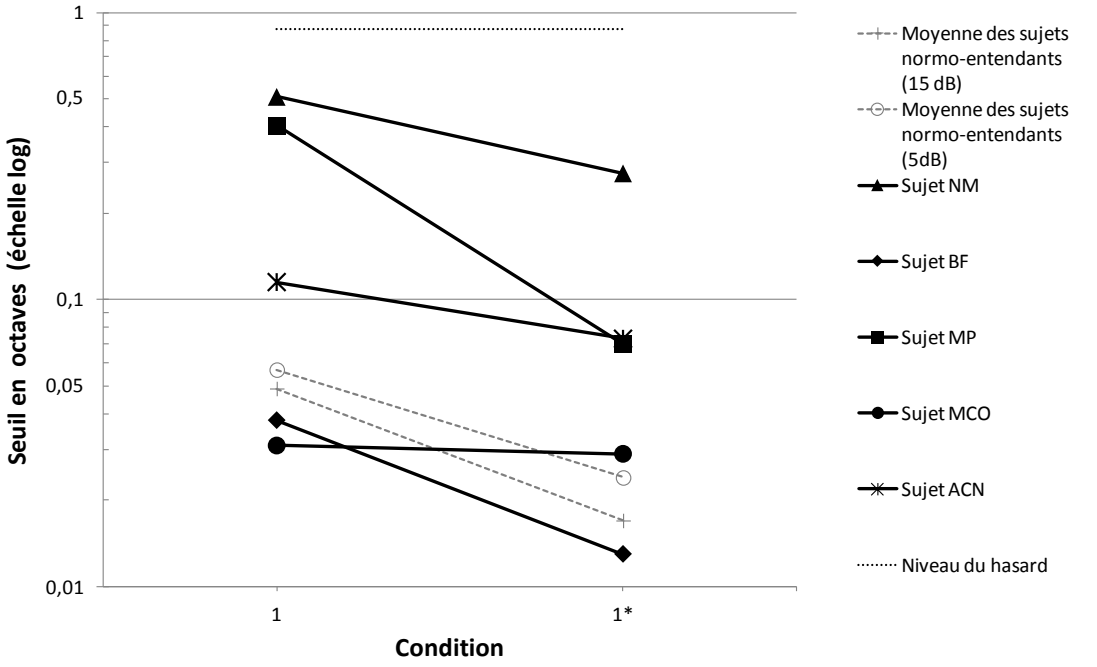


Figure 18. Seuils différentiels, en octaves, obtenus dans la condition *1* (étalon harmonique et variation aléatoire de F) et dans la condition *1** (étalon harmonique sans variation aléatoire de F au sein de chaque paire). La moyenne des résultats du groupe des sujets normo-entendants est indiquée pour les deux niveaux d'intensité des stimuli, 5 et 15 dB SL (pointillés gris). Les résultats individuels des cinq sujets atteints de surdité (NM, BF, MP, MCO et ACN) sont notés en noir gras. Le niveau du hasard (obtenu par une simulation d'un million de blocs d'essais où un sujet virtuel répondrait systématiquement au hasard) est indiqué par les pointillés noirs.

condition sont le reflet de la précision de l'encodage fréquentiel. Dans la condition *1* en revanche, la variation aléatoire de F au sein de chaque paire oblige le sujet à juger des *rapports* de fréquences, et plus précisément dans ce cas à détecter une inharmonicité par rapport à un étalon harmonique. Les seuils différentiels obtenus dans la condition *1* reflétent donc la sensibilité absolue à l'harmonicité.

La Figure 18 présente les seuils différentiels obtenus dans les conditions *1* et *1**. La moyenne des résultats des auditeurs sains est indiquée par les symboles gris, pour chaque niveau d'intensité (5 dB SL et 15 dB SL). Les résultats individuels des sujets atteints de surdité sont représentés par les symboles noirs. Le niveau du hasard est matérialisé par une ligne pointillée noire.

Les données des sujets sains ont été soumises à une analyse de variance utilisant les mêmes facteurs fixes et aléatoire que précédemment, le facteur *Condition* ayant cette fois deux modalités (1 ou 1*). Cette analyse retrouve un fort effet significatif du facteur *Condition* ($F(1,8) = 45.6$, $P = 0.0001$) ainsi que du facteur *Niveau d'intensité* ($F(1,8) = 19.7$, $P < 0.002$). L'interaction entre ces deux facteurs est à la limite de la significativité ($F(1,8) = 4.4$, $P = 0.069$), ce qui indique que lorsque le niveau d'intensité passe de 15 à 5 dB SL, la dégradation des seuils tend à être plus importante dans la condition *1** que dans la condition *1*. Par ailleurs, aussi bien dans la condition *1* que dans la condition *1**, l'abaissement du niveau d'intensité de 15 à 5 dB SL se traduit par une élévation significative des seuils différentiels ($t(53) = 2.3$, $P = 0.024$ dans la condition *1* ; $t(53) = 7.6$, $P < 0.0001$ dans la condition *1**).

Concernant les résultats individuels des sujets sourds, on retrouve comme chez les auditeurs sains, des seuils significativement meilleurs dans la condition *1** que dans la condition *1* chez quatre des cinq sujets testés ($t(18) = 4.8$, $P < 0.0001$ pour le sujet NM ; $t(9) = 7.7$, $P < 0.0001$ pour le sujet BF ; $t(26) = 9.4$, $P < 0.0001$ pour le sujet MP ; $t(12) = 2.2$, $P = 0.05$ pour le sujet ACN). Chez le sujet MCO en revanche, aucune différence

significative n'est retrouvée entre les seuils obtenus dans les deux conditions (t(10) = 0.7, P = 0.51). Ceci indique que, comparativement aux autres, le sujet MCO présente une sensibilité absolue à l'harmonicité plus élevée (et donc un seuil plus bas dans la condition *1*) que ne le « laisse penser » la précision de son encodage fréquentiel.

Plus que la valeur absolue des seuils dans les différentes conditions, il apparaît donc que le rapport entre les seuils obtenus dans la condition *1* et ceux obtenus dans la condition *1** (noté rapport *1/1**) apporte une information pertinente . Ce rapport permet en effet de déterminer si la dégradation prédomine sur la discrimination fréquentielle ou sur la sensibilité absolue à l'harmonicité elle-même. Sur la Figure 18, la valeur de ce rapport peut être indirectement estimée par la pente de la droite reliant les

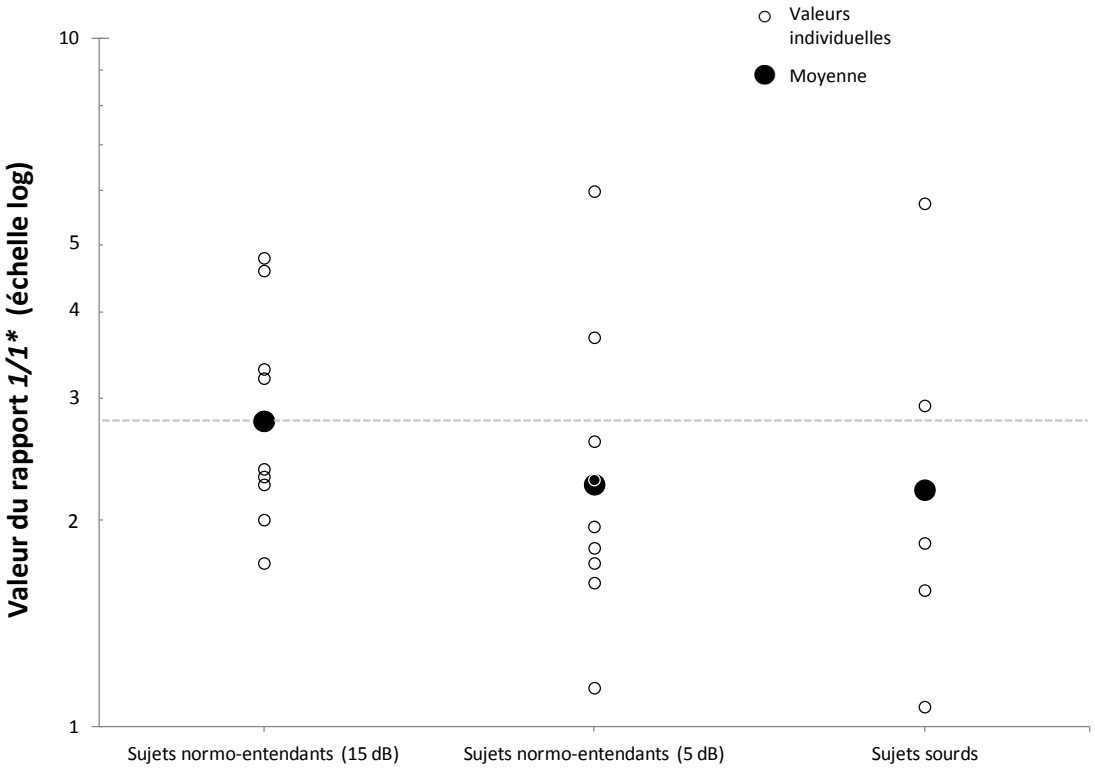


Figure 19. Valeur du rapport entre les seuils mesurés dans la condition *1* et les seuils mesurés dans la condition *1**. Sont représentées les données obtenues pour le groupe de sujets normo-entendants avec des stimuli respectivement présentés à 15 dB SL (à gauche) et à 5dB SL (au centre), et pour le groupe de sujets atteints de surdité neurosensorielle (à droite). Les résultats individuels sont indiqués par des disques blancs, les valeurs moyennes par des disques noirs. La ligne en pointillés matérialise le niveau de référence correspondant au seuil moyen des sujets normo-entendants à 15 dB SL, considéré comme la valeur « normale » du rapport *1/1**.

seuils dans les deux conditions. Néanmoins, il semble judicieux de représenter directement la valeur du rapport *1/1**. La Figure 19 donne les valeurs individuelles (disques blancs) et moyennes (disques noirs) de ce rapport chez les sujets normo-entendants à 15 dB SL et à 5 dB SL, ainsi que chez les sujets sourds. On peut grossièrement considérer que la valeur moyenne retrouvée pour des stimuli présentés au niveau d'écoute confortable de 15 dB SL chez des sujets normo-entendants constitue une sorte de valeur « normale » du rapport 1/1*. Ce niveau de référence, matérialisé par une ligne pointillée, permet d'analyser les données des sujets normo-entendants à 5 dB SL ainsi que celles des sujets sourds :

- Si le rapport est inférieur à la « normale » (au-dessous de la ligne de référence) cela indique que la précision de l'encodage fréquentiel (liée au dénominateur) est proportionnellement plus dégradée que la sensibilité absolue à l'harmonicité (liée au numérateur).
- Si le rapport est supérieur à la « normale » (au-dessus de la ligne de référence) cela indique à l'inverse que la sensibilité absolue à l'harmonicité est proportionnellement plus dégradée que la précision de l'encodage fréquentiel.

Chez les sujets normo-entendants, lorsque les stimuli sont présentés à 5 dB SL, le rapport tend à être inférieur à la valeur de référence, même si, comme cela a été indiqué plus haut, la différence est à la limite de la significativité ($P = 0.069$). Ce résultat peut s'expliquer par le fait qu'une diminution du rapport signal/bruit entraine une dégradation la précision de l'encodage fréquentiel. Il indique néanmoins que cette élévation des seuils de discrimination de fréquences tend à être plus importante que la dégradation de la sensibilité absolue à l'harmonicité.

Chez les sujets atteints de surdité, les résultats présentent une grande variabilité interindividuelle qui limite la formulation de conclusions définitives. Néanmoins, un seul des cinq sujets (MP) présente un rapport 1/1* nettement au-dessus de la ligne de

référence, alors que la valeur moyenne du rapport est légèrement inférieure au niveau de référence, comme chez les sujets normo-entendants à 5 dB SL. Ces résultats vont contre l'idée que les surdités neurosensorielles dégradent spécifiquement la sensibilité absolue à l'harmonicité.

6.4 Discussion

Ce travail constitue à notre connaissance la première étude portant sur la sensibilité à l'harmonicité chez les patients atteints de surdité neurosensorielle. Même si des contraintes expérimentales propres à ces pathologies ont pour l'instant limité le nombre d'inclusions et les données exploitables, les résultats préliminaires apportent d'ores et déjà des éléments de réponse à certaines questions essentielles relatives à l'impact de ce type de surdité sur la sensibilité à l'harmonicité.

Les surdités neurosensorielles dégradent-elles la sensibilité à l'harmonicité ?

Les résultats obtenus dans les conditions expérimentales *0.8*, *1* et *1.2* indiquent que la sensibilité relative à l'harmonicité reste statistiquement significative chez tous les sujets sourds (si l'on tient compte du niveau du hasard dans les différentes conditions), même chez les sujets dont les seuils de discrimination fréquentielle sont très dégradés. Elle apparaît donc comme un mécanisme « robuste » qui, de manière inattendue, peut être parfaitement préservé chez des sujets présentant pourtant une perte audiométrique moyenne supérieure à 45 dB sur les fréquences 500-4000 Hz. Même si elle est parfois partiellement altérée chez certains sujets, la sensibilité relative à l'harmonicité subsiste néanmoins à un niveau suffisant pour être objectivable expérimentalement (à l'image des sujets NM et MP).

Chez ces derniers sujets, les niveaux de sensibilité absolue à l'harmonicité sont par contre très altérés, jusqu'à 8 à 10 fois supérieurs à la moyenne des auditeurs sains (0.51 octave chez le sujet NM contre environ 0.05 octave chez les sujets normo-entendants). On peut s'interroger sur le bénéfice potentiel que le système auditif pourrait tirer d'un mécanisme de détection d'inharmonicité dégradé à ce point. Dans une situation habituelle de communication avec plusieurs locuteurs simultanés, la

sensibilité à l'harmonicité est un élément important de l'analyse de la scène auditive chez un auditeur sain (cf. sections 2.2.2 et 6.1). Un seuil de détection d'inharmonicité voisin d'une demi-octave est vraisemblablement très rarement suffisant pour déterminer si une composante fréquentielle appartient ou non à une série harmonique donnée dans notre environnement sonore habituel. Il faut cependant souligner que cette valeur du seuil représente un niveau moyen de sensibilité absolue à l'harmonicité sur un large registre de fréquences, et il n'est donc pas exclu que le seuil soit sensiblement meilleur pour certaines fréquences que pour d'autres. Ainsi, il est possible que même à un niveau aussi dégradé, le « reliquat » de sensibilité absolue à l'harmonicité observé par exemple chez les sujets NM et MP reste utilisable par le système auditif dans l'analyse de certaines scènes auditives.

Il est par ailleurs nécessaire de reconsidérer l'interprétation des résultats en prenant en compte le problème d'éventuels battements de consonance imparfaite liés aux interactions cochléaires entre les composantes fréquentielles des sons complexes inharmoniques. Pour les sujets sourds dont les seuils différentiels sont comparables à ceux des sujets normo-entendants (BF, MCO et dans une moindre mesure ACN), il est légitime de supposer que le mécanisme utilisé pour détecter une inharmonicité est similaire à celui des auditeurs sains. L'Expérience 2 présentée dans l'article « *Auditory discrimination of frequency ratios: The octave singularity* » (cf. section 3.2) démontre que les sujets normo-entendants n'utilisaient pas d'indice de battement pour différencier un rapport de fréquence harmonique d'un rapport inharmonique ; de plus, les conditions expérimentales sont ici encore moins favorables à la détection de battements dans la mesure où les stimuli sont présentés dans du *T.E. Noise* et à un faible niveau d'intensité (entre 5 et 15 dB SL). En revanche, certains sujets sourds que nous avons testés (principalement NM et MP) présentent des seuils différentiels globalement très élevés dans toutes les conditions, ce qui témoigne d'une altération sévère du fonctionnement

cochléaire pouvant comprendre notamment un élargissement des filtres auditifs. Ce dernier peut favoriser les interactions au niveau cochléaire, et rendre la détection de battements de consonance imparfaite plus facile. Si tel était le cas, l'amélioration significative du seuil différentiel dans la condition *1*, interprétée jusqu'alors comme le témoin d'une sensibilité relative à l'harmonicité, pourrait simplement refléter la détection de battements dans les stimuli cibles, alors qu'ils sont par définition absents dans les stimuli étalons parfaitement harmoniques. Néanmoins, il existe un argument fort allant contre l'idée que ces sujets puissent utiliser un indice de battements dans la condition *1*. En effet, la capacité du système auditif à détecter des battements dépend directement de la fréquence des ces battements, qui doit rester dans des valeurs relativement faibles. A titre de comparaison, la détection de modulation d'amplitude de sons purs (qui peut être grossièrement assimilée à la perception de battements) est optimale pour une fréquence de modulation de 4 Hz (Zwicker E. , 1952). Le sujet dont les seuils sont les plus dégradés (sujet NM) a un seuil différentiel égal à 0.51 octave dans la condition *1* ; dans le cas le plus favorable à la détection de battements, c'est à dire dans le cas où la fréquence des battements est la plus faible possible avec nos stimuli, cette dernière reste au minimum voisine de 115 Hz (en considérant que la valeur cible de *oct* soit égale à 0.5l octave, et que le sujet détecte une interaction entre les deux premières composantes du stimulus cible de fréquences 200 et 285 Hz). Ainsi, dans l'hypothèse où le masquage par du bruit et la présentation des stimuli à un niveau faible n'aient pas totalement éliminé les interactions périphériques chez les sujets dont les seuils sont très dégradés, les battements sont quoi qu'il en soit bien trop rapides pour être audibles.

La variabilité interindividuelle des résultats

La grande variabilité interindividuelle que nous observons chez les sujets sourds est un élément largement rapporté dans les travaux étudiant les diverses conséquences

des atteintes auditives neurosensorielles, notamment dans le domaine de la perception de la hauteur ou de la sensibilité à la structure temporelle fine (Lacher-Fougère & Demany, 2005; Moore B. C., 2007). Elle témoigne en partie de la diversité des lésions du système auditif périphérique englobée par la notion de surdité neurosensorielle. Nos critères audiométriques d'inclusion sont stricts, et les sujets sourds que nous avons testés peuvent être considérés comme relativement comparables en termes d'audiométrie ; la perte moyenne en audiométrie tonale s'échelonne entre 40 et 47 dB selon les critères du BIAP (cf. section 4.1), et les résultats de l'audiométrie vocale dans le silence (effectuée au cours du suivi clinique en dehors de notre protocole expérimental) sont également compris dans une gamme étroite avec un seuil d'intelligibilité variant entre 40 et 45 dB. Nos résultats indiquent donc que les niveaux de sensibilité absolue à l'harmonicité peuvent être très variables dans un groupe de patients qui serait probablement jugé particulièrement homogène par un clinicien. Cette idée selon laquelle les seuils audiométriques (qui restent pourtant au cœur du bilan des surdités en pratique clinique courante) ne reflètent que très imparfaitement le degré de l'atteinte du système auditif est depuis quelques années soutenue par de nombreux auteurs. Une dégénérescence des fibres du nerf auditif (Kujawa & Liberman, 2009), ou un défaut de synchronisation neuronale (Henry & Heinz, 2012) peuvent entrainer des troubles auditifs « supraliminaires » qui ne sont pas décelés par un examen audiométrique habituel. Nos données indiquent que la sensibilité absolue à l'harmonicité pourrait faire partie de ces mécanismes supraliminaires qui ne sont pas ou peu corrélés au seuil auditif absolu.

Cette variabilité interindividuelle ne doit pas être considérée comme un obstacle à la compréhension des mécanismes responsables de la sensibilité à l'harmonicité, mais bien comme un résultat à part entière. On peut aisément concevoir qu'il existe un lien direct entre la grande variabilité des conséquences cliniques de surdités audiométriquement équivalentes, et la présence chez certains sujets d'atteintes auditives

supraliminaires. Humes (2013), qui s'est intéressé à l'étude des différences interindividuelles dans les résultats de l'appareillage auditif conventionnel –par opposition au modèle traditionnel fondé sur la comparaison de groupes– soutient l'idée que cette approche est particulièrement pertinente pour les cliniciens qui « *prennent en charge des patients particuliers et non le patient "moyen" ou "typique"* ».

L'effet de l'âge

L'âge moyen des sujets du groupe normo-entendant est malheureusement significativement plus faible que celui des sujets sourds, dans la limite des inclusions faites à ce jour. Cette différence est notamment liée au fait que le profil de surdité que nous ciblions reste très rare chez les patients jeunes. Cela introduit incontestablement un biais, dans la mesure où le déficit de sensibilité absolue à l'harmonicité mis en évidence expérimentalement pourrait être lié en partie à l'âge des sujets et non à leur atteinte auditive. Si l'influence du vieillissement est restée longtemps peu explorée en psychoacoustique, des données relativement récentes indiquent que l'âge en tant que tel peut être significativement lié à une dégradation de certaines capacités du système auditif chez des sujets pourtant audiométriquement normaux. Alain et al. (2001) et Grube, Cramon & Rübsamen (2003) ont montré que le seuil de détection d'une composante fréquentiellement déplacée au sein d'une série harmonique augmente significativement avec l'âge, même en l'absence de surdité. Moore et al. (2012) et Hopkins & Moore (2011) ont par ailleurs mis en évidence que la sensibilité à la structure temporelle fine diminue également avec l'âge, même en l'absence d'élévation des seuils audiométriques absolus. Comme ces dernières données pouvaient le suggérer, on retrouve dans le groupe de sujets sains que nous avons testés une corrélation significative entre l'âge des sujets et les seuils obtenus dans la condition *1* (qui reflètent la sensibilité absolue à l'harmonicité) lorsque les stimuli sont présentés à 15 dB SL ($r = 0.78$, $P = 0.014$). Cette

corrélation est à la limite de la significativité lorsque les stimuli sont présentés à 5 dB SL ($r = 0.66$, $P = 0.052$). Il est donc crucial pour la poursuite de l'étude de veiller à apparier au mieux les sujets des deux groupes du point de vue de l'âge, notamment en augmentant le nombre de sujets âgés de plus de 45 à 50 ans dans le groupe des auditeurs sains.

Implications pour les grands modèles de la perception de la hauteur

Nous n'avons pas la prétention de répondre par nos données à la question du mécanisme sous-tendant la perception de la hauteur et la fusion harmonique, qui est débattue depuis plusieurs décennies. Néanmoins, il est possible d'analyser avec lequel de ces grands modèles nos résultats s'accordent le mieux. Comme cela a été précisé, une théorie fondée sur la reconnaissance de pattern suppose qu'il est possible que des sujets conservent une sensibilité absolue à l'harmonicité normale, malgré une dégradation de la précision dans l'encodage des fréquences. Aucun des résultats que nous avons obtenus ne va dans ce sens. Chez les auditeurs sains tout d'abord, la manipulation du niveau d'intensité relative du bruit par rapport aux stimuli permet dans une certaine mesure de simuler une dégradation de l'encodage périphérique de la fréquence, sans pour autant modifier la précision des éventuels gabarits internes dont dispose le système auditif. Les résultats montrent effectivement que la diminution du rapport signal/bruit augmente significativement les seuils différentiels de fréquence (Figure 18, condition *1**), mais cela s'accompagne également d'une altération significative de la sensibilité absolue à l'harmonicité (Figure 18, condition *1*). Par ailleurs, chez les sujets atteints de surdité que nous avons testés, aucun ne présente à la fois une dégradation notable dans la discrimination de fréquence et une sensibilité absolue à l'harmonicité normale. Au total, le type de résultat qui aurait constitué un argument fort en faveur d'un modèle fondé sur la reconnaissance de pattern n'a pas pu être mis en évidence pour l'instant chez les sujets

testés. Au contraire, l'ensemble de nos données est en faveur d'une dégradation concomitante de la discrimination fréquentielle et de la sensibilité à l'harmonicité, ce qui semble plus facilement explicable par un modèle reposant sur un mécanisme de type autocorrélation.

7 CONCLUSION GENERALE

Les travaux présentés dans cette thèse apportent des données nouvelles sur la sensibilité à l'harmonicité chez les sujets normo-entendants et chez les sujets atteints de surdité neurosensorielle. Chez les auditeurs sains, les résultats indiquent que l'octave harmonique et l'octave mélodique sont analysées par des mécanismes aux propriétés différentes. L'octave harmonique a un statut perceptif particulier, et elle est reconnue par un mécanisme capable de discriminer un très faible écart à l'octave, sans pour autant préciser le sens de cet écart ; ce n'est pas le cas de l'octave mélodique. Cette constatation va contre l'idée que la fusion harmonique repose sur la reconnaissance de patterns spécifiques par des gabarits internes, puisque de tels gabarits seraient théoriquement utilisables de manière équivalente dans les deux conditions.

Il était raisonnable de penser que les sujets atteints de surdité neurosensorielle puissent présenter une dégradation importante de la sensibilité à l'harmonicité, qui aurait pu rendre compte d'une partie de la gène auditive rapportée par ces patients dans les environnements sonores multi-sources. En fait, nos données préliminaires portant sur cinq sujets sourds montrent qu'il est nécessaire de dissocier la sensibilité absolue et la sensibilité relative à l'harmonicité. Le niveau de sensibilité absolue, défini comme le seuil de détection d'une inharmonicité, présente une très grande variabilité interindividuelle qui s'étend de seuils comparables à ceux des sujets normo-entendants jusqu'à des seuils dix fois plus élevés. Néanmoins, nos résultats indiquent que les surdités neurosensorielles n'entrainent pas d'atteinte spécifique de la sensibilité absolue à l'harmonicité, sa dégradation étant avant tout la conséquence de l'altération de la discrimination fréquentielle.

La sensibilité relative à l'harmonicité correspond au fait que les sujets détectent plus facilement une différence de rapports de fréquences entre un stimulus harmonique et un stimulus inharmonique, qu'entre deux stimuli inharmoniques. De manière surprenante, elle est conservée chez tous les sujets sourds que nous avons testés, même en présence d'une dégradation importante de la discrimination fréquentielle, et dans des conditions expérimentales qui rendent la détection de battements quasi impossible. Il est vraisemblable que la cette sensibilité relative à l'harmonicité, particulièrement robuste, corresponde au mécanisme spécifique de reconnaissance de l'octave harmonique que nous avons mis en évidence dans la première série d'expériences portant sur des auditeurs sains.

Certains aspects de nos résultats restent néanmoins à éclaircir, et soulèvent de nouvelles questions. Premièrement, le protocole expérimental choisi ici explorait la sensibilité à l'harmonicité de manière globale sur une large gamme de fréquences. Un niveau de sensibilité absolue à l'harmonicité très dégradé peut donc aussi bien correspondre à une atteinte homogène sur toutes les fréquences qu'à une atteinte plus hétérogène, préservant des « îlots » fréquentiels pour lesquels la sensibilité absolue reste correcte. Ce type de dépendance fréquentielle a déjà été mis en évidence dans des tâches auditives différentes (Gabriel, Koehnke, & Colburn, 1992). Deuxièmement, la corrélation entre la dégradation de la sensibilité absolue à l'harmonicité et l'âge des sujets normo-entendants amène à s'interroger sur la difficulté à isoler les conséquences directes des lésions cochléaires des diverses atteintes plus centrales liées au vieillissement. Un meilleur appariement en âge entre sujets audiométriquement normaux et sujets sourds est une solution qui a elle-même des limites, dans la mesure où certains sujets âgés dont les seuils absolus sont normaux présentent nénamoins une atteinte auditive supraliminaire.

Bibliographie

Alain, C., McDonald, K., Ostroff, J., & Schneider, B. (2001). Age-related changes in detecting a mistuned harmonic. *J. Acoust. Soc. Am., 109*, pp. 2211-2216.

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

ANSI. (1994). *American National Standard Acoustical Terminology, ANSI S1.1-1994*. New York: American National Standards Institute.

Arehart, K. H. (1994). Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners. *J. Acoust. Soc. Am., 95*, pp. 3574-3585.

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *Am. J. Psychol., 84*, pp. 147-166.

Baken, R., & Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice* (2nd ed.). London: Singular Publishing Group Inc.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature, 436*, pp. 1161-1165.

Bernstein, J. G., & Oxenham, A. J. (2005). An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination. *J. Acoust. Soc. Am., 117*, pp. 3816-3831.

Bernstein, J. G., & Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *J. Acoust. Soc. Am., 120*, pp. 3929-3945.

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*, pp. 788-801.

Bregman, A. S., & Doehring, P. (1984). Fusion of simultaneous tonal glides: The role of parallelness and simple frequency relations. *Percept. Psychophys., 36*, pp. 251-256.

Brunstrom, J. M., & Roberts, B. (2000). Separate mechanisms govern the selection of spectral components for perceptual fusion and for the computation of global pitch. *J. Acoust. Soc. Am., 107*, pp. 1566-1577.

BIAP, Bureau International d'Audiophonologie. (1997). Page Internet *Recommandation biap 02/1 bis*. Consultée le 12 Juillet 2013 sur biap.org: http://www.biap.org/fr/recommandations/65-ct-2-classification-des-surdites

Burns, E. M., & Turner, C. (1986). Pure-tone pitch anomalies. II. Pitch-intensity effects and diplacusis in impaired ears. *J. Acoust. Soc. Am., 79*, pp. 1530-1540.

Burns, E. M., & Viemeister, N. F. (1976). Nonspectral pitch. *J. Acoust. Soc. Am., 60*, pp. 863-869.

Carlyon, R. P. (1998). Comments on "A unitary model of pitch perception" [J. Acoust. Soc. Am. 102, 1811-1820 (1997)]. *J. Acoust. Soc. Am., 104*, pp. 1118-1121.

Carlyon, R. P., & Gockel, H. E. (2008). Effects of harmonicity and regularity on the perception of sound sources. In *Auditory perception of sound sources.* (W. A. Yost, A. N. Popper, & R. R. Fay Eds.) New York: Springer, pp. 191-213.

Carlyon, R. P., & Shackleton, T. M. (1994). Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? *J. Acoust. Soc. Am., 95*, pp. 3541-3554.

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

Cramer, E. M., & Huggins, W. H. (1958). Creation of pitch through binaural interaction. *J. Acoust. Soc. Am., 30*, pp. 413-417.

Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E., Klein, R., Mares-Perlman, J. A., et al. (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. The Epidemiology of Hearing Loss Study. *Am. J. Epidemiol., 148*, pp. 879-886.

Dai, H. (2000). On the relative influence of individual harmonics on pitch judgment. *J. Acoust. Soc. Am., 107*, pp. 953-959.

de Cheveigné, A. (1993). Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing. *J. Acoust. Soc. Am., 93*, pp. 3271-3290.

de Cheveigné, A., & Pressnitzer, D. (2006). The case of the missing delay lines: synthetic delays obtained by cross-chanel phase interaction. *J. Acoust. Soc. Am., 119*, pp. 3908-3918.

Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *J. Acoust. Soc. Am., 76*, pp. 57-66.

Demany, L., & Semal, C. (1992). Detection of inharmonicity in dichotic pure-tone dyads. *Hear. Res., 61*, pp. 161-166.

Demany, L., & Semal, C. (1990). Harmonic and melodic octave templates. *J. Acoust. Soc. Am., 88*, pp. 2126-2135.

Demany, L., & Semal, C. (1988). Dichotic fusion of two tones one octave apart: evidence for internal octave templates. *J. Acoust. Soc. Am., 83*, pp. 687-695.

Dobbins, P. A., & Cuddy, L. L. (1982). Octave discrimination: An experimental confirmation of the "streched" subjective octave. *J. Acoust. Soc. Am., 72*, pp. 411-415.

Evans, E. F., & Harrison, R. V. (1976). Correlation between outer hair cell damage and deterioration of cochlear nerve tuning properties in the guinea pig. *J. Physiol., 256*, pp. 43-44.

Fishman, Y. I., & Steinschneider, M. (2010). Neural correlates of auditory scene analysis based on inharmonicity in monkey primary auditory cortex. *J. Neurosci., 30*, pp. 12480-12494.

Gabriel, K., Koehnke, J., & Colburn, H. (1992). Frequency dependence of binaural performance in listeners with impaired binaural hearing. *J. Acoust. Soc. Am., 91*, pp. 336-347.

García-Pérez, M. A. (2011). A cautionary note on the use of the adaptive up-down method. *J. Acoust. Soc. Am., 130*, pp. 2098-2107.

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am., 54*, pp. 1496-1516.

Grube, M., von Cramon, D., & Rübsamen, R. (2003). Inharmonicity detection. Effects of age and contralateral distractor sounds. *Exp. Brain. Res., 153*, pp. 637-642.

Hall, D. A., & Plack, C. (2009). Pitch processing sites in the human auditory brain. *Cereb. Cortex, 19*, pp. 576-585.

Halsey, K., Fegelman, K., Raphael, Y., Grosh, K., & Dolan, D. F. (2005). Long-term effects of acoustic trauma on electrically evoked otoacoustic emission. *J. Assoc. Res. Otolaryngol., 6*, pp. 324-340.

Harrison, R. V., & Evans, E. F. (1979). Some aspects of temporal coding by single cochlear fibres from regions of cochlear hair cell degeneration in the guinea pig. *Arch. Otorhinolaryngol., 224*, pp. 71-78.

Heinz, M. G., Colburn, H. S., & Carney, L. H. (2001). Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. *Neur. Comput., 13*, pp. 2273-2316.

Henry, K. S., & Heinz, M. G. (2012). Diminished temporal coding with sensorineural hearing loss emerges in background noise. *Nat. Neurosci., 15*, pp. 1362-1364.

Hopkins, K., & Moore, B. C. (2007). Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *J. Acoust. Soc. Am., 122*, pp. 1055-1068.

Hopkins, K., & Moore, B. C. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *J. Acoust. Soc. Am., 130*, pp. 334-349.

Houtsma, A. J., & Goldstein, J. L. (1972). The central origin of the pitch of pure tones : Evidence from musical interval recognition. *J. Acoust. Soc. Am., 51*, pp. 520-529.

Houtsma, A. J., & Smurzynski, J. (1990). Pitch identification ans discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am., 87*, pp. 304-310.

Humes, L. (2013). Individual differences research and hearing aid outcomes. *Semin. Hear., 34*, pp. 67-73.

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *Neuroimage, 50*, pp. 1202-1211.

Huss, M., & Moore, B. C. (2005). Dead regions and pitch perception. *J. Acoust. Soc. Am., 117*, pp. 3841-3852.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Percept. Psychophys., 49*, pp. 227-229.

Kale, S., Micheyl, C., & Heinz, M. G. (2013). Effects of sensorineural hearing loss on temporal coding of harmonic and inharmonic tone complexes in the auditory nerve. *Adv. Exp. Med. Biol.*, pp. 109-118.

Kalluri, S., Depireux, D. A., & Shamma, S. A. (2008). Perception and cortical neural coding of harmonic fusion in ferrets. *J. Acoust. Soc. Am., 123*, pp. 2701-2716.

Keilson, S. E., Richards, V. M., Wyman, B. T., & Young, E. D. (1997). The representation of concurrent vowels in the cat anesthetized ventral cochlear nucleus: evidence for a periodicity-tagged spectral representation. *J. Acoust. Soc. Am., 102*, pp. 1056-1071.

Kemp, D. T. (2002). Otoacoustic emissions, their origin in cochlear function, and use. *Br. Med. Bull., 63*, pp. 223-241.

Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after "temporary" noise-induced hearing loss. *J. Neurosci., 29*, pp. 14077-14085.

Lacher-Fougère, S., & Demany, L. (2005). Consequence of cochlear damage for the detection of interaural phase differences. *J. Acoust. Soc. Am., 118*, pp. 2519-2526.

Lacher-Fougère, S., & Demany, L. (1998). Modulation detection by normal and hearing-impaired listeners. *Audiology, 37*, pp. 109-121.

Larsen, E., Cedolin, L., & Delgutte, B. (2008). Pitch representations in the auditory nerve: Two concurrent complex tones. *J. Neurophysiol., 100*, pp. 1301-1319.

Li, X., Nie, K., Imennov, N. S., Won, J. H., Drennan, W. R., Rubinstein, J. T., et al. (2012). Improved perception of speech in noise and Mandarin tones with acoustic simulations of harmonic coding for cochlear implants. *J. Acoust. Soc. Am., 132*, pp. 3387-3398.

Liberman, M. C., Dodds, L. W., & Learson, D. A. (1986). *Basic and applied aspects of noise-induced hearing loss.* (R. J. Salvi, D. Henderson, R. P. Hamernik, & V. Colletti, Eds.) New York: Plenum.

Liberman, M. C., Dodds, L. W., & Learson, D. A. (1986). Structure-function correlation in noise-damaged ears: A light and electron-microscopic study. In *Basic and Applied Aspects of Noise-Induced Hearing Loss.* (R. J. Salvi, D. Henderson, R. P. Hamernik, & V. Colletti Eds.) New York: Plenum, pp. 163-176.

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

Licklider, J. C. (1951). A duplex theory of pitch perception. *Experientia, 7*, pp. 128-133.

Lorenzi, C., Debruille, L., Garnier, S., Fleuriot, P., & Moore, B. C. (2009). Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. *J. Acoust. Soc. Am., 125*, pp. 27-30.

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

McAlpine, D., Jiang, D., & Palmer, A. R. (2001). A neural code for low-frequency sound localization in mammals. *Nat. Neurosci., 4*, pp. 396-401.

Meddis, R., & Hewitt, M. J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *J. Acoust. Soc. Am., 89*, pp. 2866-2882.

Meddis, R., & O'Mard, L. (1997). A unitary model of pitch perception. *J. Acoust. Soc. Am., 102*, pp. 1811-1820.

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

Micheyl, C., Xiao, L., & Oxenham, A. J. (2012). Characterizing the dependence of pure-tone frequency difference limens on frequency, duration, and level. *Hear. Res., 292*, pp. 1-13.

Miller, R. L., Schilling, J. R., Franck, K. R., & Young, E. D. (1997). Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers. *J. Acoust. Soc. Am., 101*, pp. 3602-3616.

Moore, B. C. (1982). *An Introduction to the Psychology of Hearing* (2nd ed.). London: Academic Press.

Moore, B. C. (2007). *Cochlear Hearing Loss : Physiological, Psychological and Technical Issues* (2nd ed.). Chichester: John Wiley & Sons Ltd.

Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception for normal-hearing and hearing-impaired people. *J. Assoc. Res. Otolaryngol., 9*, pp. 399-406.

Moore, B. C., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *J. Acoust. Soc. Am., 91*, pp. 2881-2893.

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

Moore, B. C., Glasberg, B. R., & Shailer, M. J. (1984). Frequency and intensity difference limens for harmonics within complex tones. *J. Acoust. Soc. Am., 75*, pp. 550-561.

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

Moore, B. C., Huss, M., Vickers, B. R., Glasberg, B. R., & Alcántara, J. I. (2000). A test for the diagnosis of dead regions in the cochlea. *Br. J. Audiol., 34*, pp. 205-224.

Moore, B. C., Peters, R. W., & Glasberg, B. R. (1985). Thresholds for the detection of inharmonicity in complex tones. *J. Acoust. Soc. Am., 77*, pp. 1861-1867.

Moore, B., Glasberg, B., Stoev, M., Füllgrabe, C., & Hopkins, K. (2012). The influence of age and high-frequency hearing-loss on sensitivity to temporal fine structure at low frequencies. *J. Acoust. Soc. Am., 131*, pp. 1003-1006.

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

Nelson, E. G., & Hinojosa, R. (2003). Presbyacusis: A human temporal bone study of individuals with flat audiometric patterns of hearing loss using a new method to quantify stria vascularis volume. *Laryngoscope, 113*, pp. 1672-1686.

Ohm, G. S. (1843). Über die Definition des Tones, nebst daran geknüfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen (On the definition of a tone and related theory of a siren and similar tone-producing devices). *Annalen der Physik und Chemie, 59*, pp. 513-565.

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends Amplif., 12*, pp. 316-331.

Oxenham, A. J., Bernstein, J. G., & Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proc. Natl. Acad. Sci. U.S.A., 101*, pp. 1421-1425.

Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A., & Santurette, S. (2011). Pitch perception beyond the traditional existence region of pitch. *Proc. Natl. Acad. Sci. U.S.A., 108*, pp. 7629-7634.

Palmer, A. R. (1988). The representation of concurrent vowels in the temporal discharge patterns of auditory nerve fibers. In *Basic Issue in Hearing.* (H. Duifhuis, J. W. Jorst, & H. Wit Eds.) London, pp. 244-251.

Palmer, A. R., & Russell, I. J. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hear. Res., 24*, pp. 1-15.

Pantev, C., Elbert, T., Ross, B., Eulitz, C., & Terhardt, E. (1996). Binaural fusion and the representation of virtual pitch in the human auditory cortex. *Hear. Res., 100*, pp. 164-170.

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

Patterson, R. D., Allerhand, M., & Giguère, C. (1995). Time-domain modelling of peripheral auditory processing: A modular architecture and a software plateform. *J. Acoust. Soc. Am., 98*, pp. 1890-1894.

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

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J. Neurosci., 24*, pp. 6810-6815.

Plack, C. J. (2005). *The Sense of Hearing.* New York: Psychology Press Taylor & Francis Group.

Plomp, R. (1967). Beats of mistuned consonances. *J. Acoust. Soc. Am., 42*, pp. 462-474.

Plomp, R. (1967). Pitch of complex tones. *J. Acoust. Soc. Am., 41*, pp. 1526-1533.

Ritsma, R. J. (1962). Existence region of the tonal residue. *J. Acoust. Soc. Am., 34*, pp. 1224-1229.

Roberts, B., & Brunstrom, J. M. (2001). Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral strech. *J. Acoust. Soc. Am., 110*, pp. 2479-2490.

Roberts, B., & Brunstrom, J. M. (2003). Spectral pattern, harmonic relations, and the perceptual grouping of low-numbered components. *J. Acoust. Soc. Am., 114*, pp. 2118-2134.

Rose, J. E., Brugge, J. F., Anderson, D. J., & Hind, J. E. (1968). *Hearing Mecanisms in Vertebrates.* (A. V. S. de Reuck, & J. Knight Eds.) London.

Schouten, J. F. (1940). The residue and the mechanism of hearing. *Proceedings of the Koninklijke Nederlands Akademie van Wetenschap, 43*, pp. 991-999.

Schouten, J. F., Ritsma, R. J., & Cardozo, B. L. (1962). Pitch of the residue. *J. Acoust. Soc. Am., 34*, pp. 1418-1424.

Schuknecht, H. F. (1993). *Pathology of the ear.* Philadelphia: Lea and Febiger.

Sek, A., & Moore, B. C. (1995). Frequency discrimination as a function of frequency, measured in several ways. *J. Acoust. Soc. Am., 97*, pp. 2479-2486.

Sellick, P. M., Patuzzi, R., & Johnstone, B. M. (1982). Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique. *J. Acoust. Soc. Am., 72*, pp. 131-141.

Semal, C., & Demany, L. (1990). The upper limit of "musical" pitch. *Music Perception, 8*, pp. 165-176.

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*, pp. 2631-2644.

Sinex, D. G. (2008). Responses of cochlear nucleus neurons to harmonic and mistuned complex tones. *Hear. Res., 238*, pp. 39-48.

Sinex, D. G., & Li, H. (2007). Responses of inferior colliculus neurons to double harmonic tones. *J. Neurophysiol., 98*, pp. 3171-3184.

Sinex, D. G., Guzik, H., Li, H., & Henderson Sabes, J. (2003). Responses of auditory nerve fibers to harmonic and mistuned complex tones. *Hear. Res., 182*, pp. 130-139.

Stevens, S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *J. Acoust. Soc. Am.*, pp. 185-190.

Terhardt, E. (1974). Pitch, consonance, and harmony. *J. Acoust. Soc. Am., 55*, pp. 1061-1069.

Tramo, M. J., Cariani, P. A., Delgutte, B., & Braida, L. D. (2001). Neurobiological foundations for the theory of harmony in western tonal music. *Ann. NY Acad. Sci., 930*, pp. 92-116.

Verschuure, J., & van Meeteren, A. A. (1975). The effect of intensity on pitch. *Acustica, 32*, pp. 33-44.

Viemeister, N., Rickert, M., & Stellmack, M. (2001). Beats of mistuned consonances: Implications for auditory coding. In *Physiological and Psychophysical Bases of Auditory Function.* (D. Breebart, A. Houtsma, A. Kohlrausch, V. Prijs, & R. Schoonhoven Eds.) Maastrict: Shaker Publishing, pp. 113-120.

Vinay, & Moore, B. C. (2007). Prevalence of dead regions in subjects with sensorineural hearing loss. *Ear Hear., 28*, pp. 231-241.

Woolf, N. K., Ryan, A. F., & Bone, R. C. (1981). Neural phase-locking properties en the absence of outer hair cells. *Hear. Res., 4*, pp. 335-346.

Yost, W. A. (2009). Pitch perception. *Atten. Percept. Psychophys., 71*, pp. 1701-1715.

Zwicker, E. (1952). Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones. *Acustica*, pp. 125-133.

Zwicker, E. (1970). Masking and Psychological Excitation as Consequences of the Ear's Frequency Analysis. In *Frequency Analysis and Periodicity Detection in Hearing.* (R. Plomp, & G. F. Smoorenburg Eds.) Leiden, pp. 376-394.

Zwicker, U. (1984). Auditory recognition of diotic and dichotic vowel pairs. *Speech Commun., 3*, pp. 265-277.

SERMENT D'HIPPOCRATE

Au moment d'être admis(e) à exercer la médecine, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité.

Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences.

Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences.

Je donnerai mes soins à l'indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

Admis(e) dans l'intimité des personnes, je tairai les secrets qui me seront confiés. Reçu(e) à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les moeurs.

Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité.

Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré(e) et méprisé(e) si j'y manque.