



Michel Denis (dir.)

La psychologie cognitive

Éditions de la Maison des sciences de l'homme

Chapitre 4. La perception auditive : notions fondamentales et applications audiologiques

Christian Lorenzi

DOI : 10.4000/books.editionsmsmh.14790

Éditeur : Éditions de la Maison des sciences de l'homme

Lieu d'édition : Éditions de la Maison des sciences de l'homme

Année d'édition : 2012

Date de mise en ligne : 29 novembre 2019

Collection : Cogniprisme

ISBN électronique : 9782735119295



<http://books.openedition.org>

Référence électronique

LORENZI, Christian. *Chapitre 4. La perception auditive : notions fondamentales et applications audiologiques* In : *La psychologie cognitive* [en ligne]. Paris : Éditions de la Maison des sciences de l'homme, 2012 (généré le 16 mars 2020). Disponible sur Internet : <<http://books.openedition.org/editionsmsmh/14790>>. ISBN : 9782735119295. DOI : <https://doi.org/10.4000/books.editionsmsmh.14790>.

Ce document a été généré automatiquement le 16 mars 2020. Il est issu d'une numérisation par reconnaissance optique de caractères.

Chapitre 4. La perception auditive : notions fondamentales et applications audiologiques¹

Christian Lorenzi

- ¹ La psychologie de la perception auditive – et plus spécifiquement la psychoacoustique – naît avec la psychologie expérimentale et cognitive, il y a plus d’un siècle. Son père, Hermann Ludwig von Helmholtz, en décrit les fondements dans sa célèbre *Théorie physiologique de la musique* (1863). Fait singulier, Helmholtz brille autant par ses contributions fondamentales à la compréhension des processus auditifs périphériques et centraux que par son génie technique. Les applications dans le domaine de l’acoustique réalisées par Helmholtz sont nombreuses (résonateurs, sirènes, diapason électrique, etc.) et ouvrent en retour de nouvelles voies de recherche. Depuis sa création il y a maintenant cent cinquante ans, la psychologie de l’audition maintient cette liaison forte entre science et technique, multipliant les applications dans les domaines les plus variés (traitement du signal, reconnaissance de la parole, synthèse musicale, hi-fi, codage et compression de données [MP3], acoustique architecturale, design sonore, réalité virtuelle, etc.) et démontrant sans conteste son utilité pour la société. Ses applications dans le domaine du diagnostic et de la réhabilitation des surdités neurosensorielles sont présentées dans ce chapitre, non seulement en raison de leurs enjeux évidents dans le domaine de la santé publique, mais aussi parce que ce champ d’application – l’audiologie – illustre mieux que tout autre comment des connaissances psychoacoustiques fondamentales peuvent être transférées directement dans les domaines industriel, biomédical et clinique, et être transformées en technologies numériques de pointe allant des simples prothèses amplificatrices à des oreilles bioniques ou implants, véritables interfaces entre cerveau humain et machine. Ce chapitre débute par une brève revue de notions fondamentales en psychoacoustique et présente un modèle de traitement sensoriel du signal sonore en cinq étapes, modèle partagé par la majorité des chercheurs en psychoacoustique à l’heure actuelle. Cette revue se poursuit par un exposé des effets perceptifs associés à toute perte

neurosensorielle et se termine par une description des dispositifs actuels de réhabilitation des pertes auditives.

Notions fondamentales

Les sons

- 2 Les sons de ce monde, qu'il s'agisse de sons de parole, de sons produits par des instruments musicaux ou bien de sons de l'environnement, correspondent à des variations de pression acoustique issues de la mise en vibration d'un objet physique dénommé « source sonore » (le sol, une porte, nos cordes vocales, etc.) par une perturbation mécanique (une « excitation » : choc, frottement, vibration, etc.) et véhiculées dans un milieu donné (le plus généralement, l'air pour les êtres humains). Les sons de ce monde correspondent donc à des phénomènes purement vibratoires, à savoir des ondes qui se propagent depuis la source excitée jusqu'à nos tympans. Ces sons prennent la forme d'événements impliquant ainsi au moins trois éléments, à savoir : 1/une excitation ; 2/une source (caractérisée par ses propriétés physiques de géométrie, de matériau, de type d'interaction avec les autres objets) ; et 3/un médium assurant la transmission de la vibration. Ces trois éléments vont déterminer les caractéristiques spectro-temporelles du signal acoustique (pour une présentation détaillée des notions d'acoustique, cf. Hartmann 2005 ; pour une présentation des enjeux philosophiques liés aux conceptions du son, cf. Casati et Dokic 1994).

Représentations physiques

- 3 Ces variations de pression acoustique – le son – peuvent être représentées de plusieurs manières, mais le choix d'une représentation physique donnée rend certains indices acoustiques plus explicites que d'autres (pour une présentation détaillée des notions de traitement de signal, cf. Hartmann 2005). Ce choix oriente notre compréhension des processus auditifs et conduit à favoriser une théorie dite spectrale, temporelle ou spectro-temporelle de la perception auditive. Le « signal d'onde », qui représente les variations d'amplitude du signal acoustique en fonction du temps, met l'accent sur les informations temporelles ou dynamiques, codées par des mécanismes neuronaux de synchronisation ou verrouillage en phase dans les neurones du nerf auditif, du tronc cérébral et du cortex auditif. Le « spectre d'amplitude » ou « spectre de puissance », qui représente les amplitudes relatives des différentes composantes fréquentielles du son, met l'accent sur les informations fréquentielles, codées spatialement (ou « tonotopiquement ») sur la membrane basilaire (l'épithélium sensoriel) dans la cochlée (l'oreille interne, notre organe sensoriel auditif). Le « spectrogramme », enfin, qui représente les variations d'amplitude de chaque composante fréquentielle du signal en fonction du temps, combine ces deux représentations, mais, comme le spectre de puissance, impose de choisir une certaine fenêtre d'analyse temporelle du signal. Ces différentes représentations ont leurs limites : elles souffrent de l'existence d'un compromis fondamental entre temps et fréquence, dans l'esprit de la célèbre relation d'incertitude d'Heisenberg. Ainsi, une représentation précise des informations temporelles du signal se fera nécessairement au détriment d'une représentation précise

des informations fréquentielles, et *vice versa*. Ceci n'est pas sans effet sur les théories développées pour comprendre les processus auditifs.

Les dimensions auditives

- 4 Ces variations de pression acoustique donnent lieu à des perceptions variées qualifiées d'« auditives » dans notre cerveau. Ces perceptions sont traditionnellement caractérisées par cinq attributs ou dimensions déterminés par la structure spectro-temporelle du signal sonore (pour une présentation détaillée des notions psychoacoustiques, cf. Moore 2003 ou Plack 2005). Les perceptions auditives peuvent ainsi être caractérisées par leur volume sonore ou sonie, attribut permettant de positionner les sons sur une échelle subjective allant du faible au fort. La sonie est principalement déterminée par le niveau de pression acoustique, généralement exprimé en décibels SPL (dB SPL). Toutefois, la sonie dépend également d'autres paramètres acoustiques, comme la structure spectrale, et de la durée du signal. Ces perceptions sont caractérisées par leur hauteur, tonale s'il s'agit d'un son pur (une simple sinusoïde) produit par exemple par un diapason, ou fondamentale s'il s'agit d'un son complexe harmonique comme une voyelle ou une note de violon (pour une revue détaillée de la perception de hauteur, cf. Plack, Oxenham, Fay et Popper 2005). L'attribut de hauteur permet de positionner les sons sur une échelle subjective allant du grave à l'aigu. Cet attribut dépend essentiellement de la structure spectrale (la fréquence, exprimée en hertz [Hz], s'il s'agit d'un son pur, ou le spectre d'amplitude s'il s'agit d'un son complexe) et temporelle (les fluctuations d'amplitude rapides) du signal acoustique. Le niveau sonore et la durée du signal peuvent aussi influencer la perception de hauteur. Ces perceptions sont également caractérisées par leur timbre, cet attribut multidimensionnel de notre perception auditive permettant de discriminer deux sons de hauteur et de sonie identique. Plusieurs dimensions de timbre ont été identifiées, certaines liées à la forme temporelle du signal (le temps d'attaque), et d'autres liées au spectre d'amplitude (la brillance, par exemple). Ces perceptions sont enfin caractérisées par leur durée subjective (elle-même déterminée principalement par la durée physique du signal) et la position subjective de la source sonore dans un espace auditif tridimensionnel, à savoir la position perçue de la source sonore, localisée sur un plan horizontal (azimut) et vertical (élévation). Ces perceptions spatiales sont principalement déterminées par des différences de temps d'arrivée, de phase et d'amplitude du signal acoustique aux deux oreilles, et par les effets de filtrage du signal acoustique produits par l'ensemble constitué par l'oreille externe (le pavillon), la tête, le cou et le tronc (pour une revue détaillée de l'audition spatiale, cf. Blauert 1997).

La perception de scènes auditives

- 5 Sonie, hauteur, timbre, durée, localisation sont des attributs des sources ou « objets » sonores calculés à l'issue ou peut-être parallèlement à une opération qualifiée d'« analyse de scènes auditives » (pour une présentation détaillée, cf. Bregman 1990). En effet, dans la plupart des situations naturelles d'écoute, plusieurs signaux acoustiques constituant une « scène auditive » complexe parviennent simultanément et successivement à nos tympanes. Le système auditif réussit à isoler et à identifier les différentes sources sonores présentes dans cette scène en se fondant sur les caractéristiques acoustiques communes partagées par les signaux issus d'une même

source sonore (harmonicité, synchronie, co-modulations, etc.). L'analyse d'une scène auditive ne consiste pas uniquement à séparer des objets sonores, mais également à regrouper les événements sonores issus de la même source qui se succèdent au sein d'une représentation auditive cohérente généralement dénommée « flux auditif ». Pour regrouper séquentiellement les événements sonores dans un même flux auditif, le système auditif utilise des similitudes ou des différences acoustiques perceptibles (hauteur, timbre, propriétés temporelles). Les processus cérébraux impliqués dans l'analyse en flux auditifs sont encore mal connus, mais ils pourraient impliquer des mécanismes d'adaptation neuronale débutant à un niveau très périphérique du système auditif.

Un modèle fonctionnel des traitements auditifs périphériques

Audibilité

- 6 Nous entendons les sons de ce monde entre environ 20 et 20 000 Hz. Notre sensibilité auditive – telle qu'elle est mesurée par nos seuils auditifs absolus (estimés en décibels, et correspondant à la plus petite amplitude telle que nous détectons tout juste la présence d'un son pur) – est maximale pour les sons de fréquences intermédiaires, comprises entre 1 000 et 5 000 Hz environ. Cette sensibilité optimale aux sons de fréquence intermédiaire n'a pas une origine neurale. Elle résulte d'une amplification sélective du signal acoustique réalisée au moment de sa transmission purement aérienne, puis mécanique, par l'oreille externe (le pavillon et le conduit auditif) et l'oreille moyenne (le tympan et la chaîne ossiculaire : marteau, enclume et étrier) (pour une présentation détaillée des notions anatomiques et physiologiques, cf. Møller 2006 ou Pickles 2008). Le signal est ensuite amplifié par la contraction de cellules ciliées externes (CCE) situées sur la membrane basilaire dans la cochlée (au nombre de 12 000 environ par cochlée). Cette amplification sélective – assimilable à une première étape de filtrage et de rehaussement du signal acoustique – adapte parfaitement notre fenêtre d'audibilité aux caractéristiques spectrales des sons les plus importants pour notre espèce, à savoir la voix, puisque le spectre de la voix humaine « utile » pour la transmission d'informations linguistiques est compris entre environ 50 et 5 000 Hz (l'énergie du signal est essentiellement concentrée en dessous de 500 Hz, puis diminue au-delà de quelques décibels par octave).

Sélectivité fréquentielle

- 7 Les signaux acoustiques amplifiés sélectivement entre 1 000 et 5 000 Hz par l'oreille externe et moyenne sont ensuite décomposés en leurs fréquences constitutives par une batterie de filtres accordés dont les fréquences centrales s'échelonnent entre 20 et 20 000 Hz, et dont les bandes passantes (la largeur des filtres ou la sélectivité) se recouvrent continûment. Ces filtres auditifs sont généralement caractérisés à l'aide de mesures psychoacoustiques de seuils de détection (de sons purs cibles) masqués par des bruits présentant une échancrure centrée sur la fréquence du son cible. Ces « filtres ou canaux auditifs » sont également dénommés « filtres cochléaires » car le filtrage en fréquence réalisé par le système auditif découle des propriétés de l'onde qui se propage depuis la base jusqu'à l'apex sur la membrane basilaire dans la cochlée et des contractions des CCE en réponse au son. Ces filtres auditifs présentent une forme asymétrique (leur pente est plus douce sur le versant basse fréquence que sur le

versant haute fréquence) et leur bande passante (lorsqu'elle est estimée sur une échelle linéaire en hertz) augmente avec la fréquence centrale du filtre. Les filtres auditifs sont donc plus sélectifs (plus étroits) en basse fréquence qu'en haute fréquence, et la capacité à séparer ou résoudre deux composantes fréquentielles au sein d'un son (la capacité à les entendre séparément) est meilleure en basse fréquence. Pour un son complexe harmonique (une voyelle ou un son de violon, par exemple), la limite de résolubilité auditive humaine est estimée entre le 6^e et le 8^e harmonique (au-delà, les harmoniques du son ne peuvent être perçus comme distincts, et sont dits « non résolus » ; ils interagissent au sein du filtre et produisent des battements, c'est-à-dire des fluctuations d'amplitude plus ou moins rapides).

- 8 Cette bande passante correspond approximativement à 10 % de la fréquence du filtre ou un sixième d'octave et peut être estimée avec précision grâce à l'échelle dite des ERB (*Equivalent Rectangular Bandwidth*, ou bande passante rectangulaire équivalente ; cf. Moore 2003) décrite par l'équation suivante :
- 9
$$\text{ERB}(f) = 24,7 (4,37 f / 1\ 000 + 1)$$
- 10 où ERB correspond à la bande passante du filtre (estimée en hertz) et f à sa fréquence centrale (estimée également en hertz ; par exemple, la bande passante du filtre accordé sur une fréquence centrale de 1 000 Hz est de 132,6 Hz). Ainsi, la bande passante des filtres cochléaires s'avère constante lorsqu'elle est estimée sur une échelle logarithmique (c'est-à-dire une échelle de rapport). La sélectivité fréquentielle auditive dépend toutefois du niveau de présentation des sons (leur amplitude) et de leur spectre. En effet, les filtres auditifs sont plus sélectifs à faible niveau de présentation et pour des sons purs, en raison de phénomènes dits de « suppression latérale » (une interaction entre les réponses de la membrane basilaire à deux sons de fréquences et d'amplitudes différentes, aboutissant à rehausser la réponse au son le plus fort) qui apparaissent en réponse à des sons complexes. De fait, les estimations de sélectivité fréquentielle proposées par l'échelle des ERB sont plus adaptées au cas de sons complexes présentés à des niveaux modérés (60-70 dB SPL, le niveau général d'une conversation).
- 11 L'existence de filtres auditifs dévoile donc la capacité du système auditif à réaliser une « transformée de Fourier » – certes limitée – des signaux sonores, c'est-à-dire une capacité à analyser les sons en leurs fréquences constitutives. Elle indique également l'existence d'une forme d'indépendance perceptive le long de l'axe fréquentiel audio. Ainsi, le système auditif présente 32 filtres auditifs non recouvrants entre environ 80 et 8 000 Hz (la plage fréquentielle couverte par le signal de parole), soit 32 canaux de transmission d'informations indépendantes au système auditif central. Cela s'avère amplement suffisant pour assurer une identification précise des signaux de parole (dans le silence tout au moins), puisqu'une intelligibilité parfaite de syllabes, mots ou phrases peut être obtenue lorsque les variations d'amplitude les plus lentes de ces signaux sont transmises dans seulement 4 bandes de fréquence indépendantes.
- 12 La bande passante et la forme asymétrique de ces filtres auditifs contraignent enfin notre capacité à détecter un son en présence d'autres (par exemple, une voix plongée dans un brouhaha, une situation quotidienne d'écoute). Les études psychophysiques de ce phénomène dit de « masquage énergétique » entre sources sonores concurrentes révèlent que notre capacité à détecter un son cible donné en présence d'un son masquant simultané dépend pour une grande partie de la magnitude de la sortie du filtre auditif accordé sur la fréquence du son cible. Cette capacité est d'autant plus

dégradée que le son masquant contribue à cette sortie. Cette contribution est fonction de la séparation fréquentielle entre le son cible et le son masquant (plus ces sons sont proches en fréquence, plus l'effet de masque augmente) et culmine (l'effet de masquage est alors maximal) lorsque le son masque tombe dans la bande passante du filtre accordé sur la fréquence du son cible. Enfin, ces phénomènes de masquage énergétique ne sont pas symétriques le long de l'axe fréquentiel. L'asymétrie des filtres auditifs indiquée plus haut engendre un phénomène de masquage dit « ascendant », correspondant au fait que pour un son de fréquence donnée, les sons de plus basse fréquence présentent un pouvoir de masquage plus important que les sons de plus haute fréquence.

- 13 Comme nous l'avons indiqué plus haut, le filtrage en fréquence réalisé par le système auditif découle des propriétés de l'onde qui se propage depuis la base jusqu'à l'apex sur la membrane basilaire, ainsi que des contractions des CCE en réponse au son. Ces deux phénomènes conduisent à coder géographiquement (c'est-à-dire spatialement) les sons sur la membrane basilaire en fonction de la fréquence. Plus précisément, chaque filtre cochléaire correspond à une portion de longueur constante (0,9 mm) de la membrane basilaire (dont la longueur totale est de 33 mm chez l'homme). Ce principe de codage spatial de l'information fréquentielle dans la cochlée prend le nom de « tonotopie » et indique que le système auditif périphérique calcule de véritables spectres internes ou profils spectraux (généralement dénommés « patterns d'excitation ») en réponse aux sons, correspondant à la distribution d'excitation neurale en fonction de la fréquence centrale des filtres auditifs. Ces patterns d'excitation correspondent à des représentations neurosensorielles périphériques du signal sonore dont le grain est fixé par la résolubilité fréquentielle détaillée plus haut, et dont la forme (pics, creux) véhicule des informations critiques utilisées pour le calcul de la sonie, de la hauteur et du timbre évoqué par le son (les fréquences des formants, dans le cas des voyelles, par exemple).

Compression d'amplitude

- 14 Les variations d'excitation issues de chaque filtre cochléaire sont amplifiées par la contraction des CCE. Or cette amplification n'est pas linéaire, et le gain produit par la contraction des CCE diminue avec l'amplitude du signal acoustique (les sons peu intenses sont donc plus amplifiés que les sons très intenses). Il en résulte une compression des variations d'amplitude du signal, dont les effets s'observent aux niveaux compris entre environ 30 et 80 dB SPL et principalement au-dessus de 1 kHz. Cette compression d'amplitude quasi instantanée réalisée par les mécanismes actifs cochléaires (les CCE) contraint fortement la perception du volume sonore des sons (leur sonie) et explique l'allure des « fonctions de sonie » mesurées par les psychoacousticiens à l'aide de techniques subjectives, à savoir le fait que la sonie évolue comme une fonction puissance d'exposant inférieur à 1 (0,6) du niveau de pression acoustique.

Intégration temporelle

- 15 Les signaux acoustiques filtrés et compressés vont faire l'objet d'une dernière transformation dénommée « transduction ». Les cellules ciliées internes (CCI, au nombre de 3 500 environ par cochlée) situées le long de la membrane basilaire vont

transformer ces signaux – mécaniques à ce stade (vibrations de portions de la membrane basilaire) – en variations de potentiel électrique. Seules les portions positives de ce signal sont transmises par voie synaptique aux fibres du nerf auditif, avec pour résultat une rectification du signal filtré-compressé au niveau de la CCI. Cette étape de transduction, puis de transmission synaptique, aboutit également à une intégration temporelle du signal, assimilable à un filtrage passe-bas de fréquence de coupure estimée à 0,6-1 kHz environ chez la plupart des mammifères (la valeur précise de cette fréquence de coupure n'est toutefois pas connue chez l'homme, mais estimée à 1 kHz). Ce processus déclenche enfin l'émission de trains de potentiels d'action dans les fibres du nerf auditif. Ces fibres sont au nombre de 30 000 environ par nerf, car chaque CCI est connectée à environ 10-20 fibres nerveuses afférentes de seuils distincts (bas, moyen et haut). De fait, le nerf auditif est organisé de manière tonotopique, comme la membrane basilaire, car chaque CCI correspond à un point de la membrane basilaire, et donc à un filtre cochléaire. Les trains de potentiels d'action émis par les fibres du nerf auditif ont la propriété d'être synchronisés de manière très précise (« verrouillés en phase ») sur la période du signal filtré (on parle alors de synchronisation sur la « structure temporelle fine » du signal), autorisant ainsi un deuxième codage – temporel, cette fois – de l'information de fréquence audio. Toutefois, l'intégration temporelle indiquée plus haut limite cette capacité, et le phénomène de synchronisation disparaît au-delà de 5 kHz environ. Au-delà de cette fréquence, la distribution des potentiels d'action est aléatoire et seul le codage spatial de la fréquence est utilisable. Cette observation neurophysiologique est cohérente avec la dégradation brusque des seuils de discrimination fréquentielle mesurés sur des sons purs et la perte de la perception de la similarité des sons lorsque leurs fréquences sont dans une relation d'octave (chroma) au-delà de 5 kHz.

Filtrage de l'enveloppe temporelle

- 16 Selon les objectifs de la modélisation, les signaux transformés à l'issue de ces quatre étapes (amplification, filtrage, rectification et compression) peuvent *in fine* être soumis à une cinquième transformation qui reproduit les effets d'un codage neuronal plus central, aboutissant à l'extraction des fluctuations ou modulations d'amplitude de la sortie de chaque filtre auditif, à savoir l'« enveloppe temporelle » du signal. Cette opération est réalisée de deux manières sur le signal issu de l'étape précédente : soit en appliquant un deuxième filtrage passe-bas de fréquence de coupure égale à $ERB/2$, aboutissant à l'extraction des modulations d'amplitude les plus lentes du signal (inférieures à quelques dizaines ou centaines de hertz) au sein de chaque canal, soit en appliquant une batterie de filtres passe-bandes larges (de bande passante égale à une octave) et accordés sur des fréquences de modulation comprises entre quelques hertz et quelques centaines de hertz. Ceci aboutit à une décomposition de l'enveloppe temporelle du signal, et au calcul d'une forme de « spectre de modulation » du son. Cette dernière étape de traitement est motivée par la découverte du rôle clé (et parfois suffisant) des indices d'enveloppe temporelle dans de nombreux phénomènes auditifs, et tout particulièrement l'identification du signal de parole. Comme indiqué plus haut (*cf.* section « Sélectivité fréquentielle », p. 148), une intelligibilité parfaite de syllabes, mots ou phrases peut être obtenue sur la base de l'enveloppe temporelle de ces signaux extraite au sein de 4 bandes de fréquence indépendantes. Les indices d'enveloppe véhiculent donc l'information phonétique du signal de parole.

- 17 Une application importante de ce principe repose sur le calcul d'un index – l'index de transmission de la parole (*speech transmission index*, ou STI) – qui estime le degré d'atténuation des indices d'enveloppe temporelle d'un signal sonde (ayant certaines caractéristiques du signal de parole) lorsque ce signal sonde est transmis par un canal de communication donné. Cette mesure de la qualité de la transmission de la parole fondée sur l'enveloppe temporelle a connu un grand succès dans le domaine du traitement de la parole et de l'acoustique architecturale.

Effets des lésions cochléaires

- 18 L'étiologie des pertes auditives neurosensorielles (c'est-à-dire d'origine cochléaire) est variée (génétique, virale, pharmacologique, traumatique). La perte auditive est également causée par le vieillissement (presbyacousie). Entre 8 et 15 % de la population des pays européens présentent des signes de perte auditive d'origine neurosensorielle. Cette proportion ne peut que croître en raison du vieillissement de la population de ces pays, et constitue donc un véritable enjeu de santé publique. Les pertes auditives sont caractérisées par des troubles de détection et de discrimination des sons, ainsi que par un inconfort auditif à l'écoute de certains sons. Toutefois, la manifestation majeure des pertes auditives correspond à une dégradation de la capacité à analyser des scènes auditives complexes, comme par exemple suivre une conversation dans un brouhaha. Les études psychoacoustiques réalisées sur des patients malentendants ont permis de dégager quatre grandes conséquences perceptives des lésions cochléaires susceptibles d'expliquer ces troubles d'analyse de scènes. Ces quatre grandes conséquences perceptives de lésions cochléaires sont compatibles avec le modèle en cinq étapes décrit plus haut (pour une présentation détaillée, cf. Moore 1995, 2007).

Trouble de l'audibilité

- 19 Les pertes auditives sont généralement diagnostiquées à partir de l'audiogramme, à savoir la courbe représentant l'évolution des seuils auditifs absolus (normalisés en décibels HL) en fonction de la fréquence audiométrique. Des seuils supérieurs à 20 dB HL signalent la présence d'une perte auditive qui peut être légère (26-40 dB HL), modérée (41-55 dB HL), modérée-sévère (56-70 dB HL), sévère (71-90 dB HL) ou profonde (> 90 dB HL). La configuration de l'audiogramme peut prendre différentes formes : plate, montante (passe-haut), descendante (passe-bas), etc. L'élévation des seuils absolus indique ici une perte d'audibilité (les sons sont perçus comme moins forts) en fonction de la fréquence, aboutissant à une réduction de la bande passante audible (un rétrécissement de la fenêtre d'audibilité). La destruction de CCE (c'est-à-dire de l'amplificateur non linéaire cochléaire) dans le cas de pertes légères à modérées, puis de CCI lorsque la perte est plus sévère, est à l'origine de cette perte d'audibilité et de cette réduction de la bande passante audible. Cette perte engendre des troubles de la perception de la parole dans le silence et dans le bruit (tout particulièrement en limitant la perception des éléments les moins énergétiques du signal de parole, à savoir les consonnes).
- 20 Un vaste programme de recherche a conduit à développer et raffiner un index, dénommé initialement « index d'articulation » (*articulation index*, AI). L'AI constitue une alternative à l'index STI décrit plus haut et permet de prédire l'intelligibilité du signal

de parole à partir de la proportion du signal de parole audible pour un patient présentant une perte auditive donnée (caractérisée par une élévation de ses seuils audiométriques). Cet AI est compris entre 0 et 1, et représente la proportion moyenne du signal de parole audible, par bande de fréquence. Plus l'AI est proche de 1 ou 100 %, meilleure est supposée être l'intelligibilité. Les calculs d'AI prédisent relativement bien les performances d'identification de la parole dans des conditions variées (filtrage, présence de bruit) dans le cas des pertes légères, mais surestiment la performance réelle des patients présentant des pertes modérées-sévères, suggérant que des déficits supraliminaires (au-dessus du seuil absolu) participent au déficit de perception de la parole.

Perte de sélectivité fréquentielle

- 21 Le premier trouble supraliminaire auditif identifié porte sur la dimension fréquentielle et correspond à une perte de la sélectivité fréquentielle auditive. Les mesures psychoacoustiques de masquage permettant de déterminer la forme et la bande passante des filtres cochléaires révèlent que la bande passante de ces filtres est élargie par un facteur 2 dans le cas des pertes modérées et par un facteur 4-5 dans le cas des pertes modérées à sévères. Un élargissement des filtres auditifs d'environ 20 % a été observé dans le cas des pertes légères, mais un grand nombre de données suggèrent que la sélectivité fréquentielle cochléaire n'est pas affectée tant que les seuils absolus sont inférieurs à 30 dB HL. Cet élargissement des filtres auditifs résulte de la destruction partielle ou totale des CCE sur la membrane basilaire. Les conséquences perceptives de cette perte de sélectivité fréquentielle sont multiples. La première correspond à une susceptibilité accrue aux phénomènes de masquage dit énergétique. Ceci est conforme à la plainte récurrente des patients malentendants qui déclarent pouvoir généralement suivre une conversation dans un environnement silencieux, mais ne plus rien comprendre en milieu bruyant. La dégradation de la résolubilité des composantes fréquentielles des sons découlant de la perte de sélectivité fréquentielle aboutit à un véritable lissage ou brouillage des patterns d'excitation évoqués par les sons au niveau périphérique et affecte la perception de la hauteur (tonale comme fondamentale) et du timbre des sons, ainsi que les possibilités d'analyse de scènes auditives complexes où plusieurs signaux sonores sont mélangés. Cette perte de sélectivité aboutit également à une réduction du nombre de canaux fréquentiels indépendants capables de transmettre des informations au système central. Dans le cas d'une perte modérée, le nombre de canaux indépendants entre environ 80 et 8 000 Hz est estimé à 8 (au lieu de 32 pour un sujet normo-entendant). D'une manière générale, ceci explique la moindre robustesse du traitement de signaux complexes tels que la parole lorsque ces signaux sont soumis à différentes formes de distorsions (bruit, filtrage, compression, etc.).

Recrutement de sonie

- 22 Le deuxième trouble supraliminaire auditif identifié porte sur la dimension de l'intensité. Il se manifeste principalement par une intolérance aux sons forts (« Ça me casse les oreilles »). Ce trouble de la perception du volume sonore, dénommé « recrutement de sonie », correspond à un accroissement anormalement rapide de la perception de la sonie en fonction du niveau physique des sons. Le recrutement de

sonie est aujourd'hui compris comme résultant de la destruction des CCE et de la perte de la compression d'amplitude quasi instantanée s'appliquant à la sortie des filtres auditifs. Le recrutement de sonie correspond principalement à un inconfort auditif et n'affecte pas l'intelligibilité des signaux de parole (tout au moins dans le silence).

Troubles du codage temporel

- 23 L'acuité temporelle auditive, c'est-à-dire la limite dans la capacité à détecter des fluctuations d'enveloppe temporelle, est normale chez les personnes présentant une perte auditive neurosensorielle, lorsque l'audibilité des stimuli utilisés pour mesurer cette acuité est bien contrôlée. De manière surprenante, la capacité à détecter des modulations d'amplitude lentes semble même supérieure à celle observée chez les sujets normo-entendants. Ces résultats sont malgré tout – et à première vue – cohérents avec la perte de sélectivité fréquentielle et le recrutement de sonie associés aux pertes cochléaires. En effet, comme il a été indiqué plus haut (cf. section « Représentations physiques », p. 145), la perte de résolution fréquentielle devrait – tout au moins, en théorie – aller de pair avec une amélioration de résolution temporelle. Par ailleurs, la perte de compression cochléaire correspond à une forme d'expansion (ou rehaussement) des fluctuations d'amplitude en sortie des filtres cochléaires (par comparaison avec une cochlée normale) susceptible de conduire à une meilleure sensibilité aux modulations d'amplitude des signaux acoustiques. Toutefois, des données neurophysiologiques et psychoacoustiques récentes ne confortent pas ces deux hypothèses et suggèrent que les facteurs principaux limitant le traitement des indices d'enveloppe sont neuronaux et centraux (c'est-à-dire post-cochléaires). Quelle que soit l'explication de ces observations, un traitement normal (voire meilleur que la normale) des indices d'enveloppe temporelle explique pourquoi l'intelligibilité de la parole dans le silence se maintient chez les personnes malentendantes lorsque les signaux sont amplifiés afin de compenser la perte d'audibilité, et ceci malgré une dégradation substantielle de sélectivité fréquentielle. En effet, comme il a été indiqué plus haut (section « Sélectivité fréquentielle », p. 148), une intelligibilité parfaite de syllabes, mots ou phrases peut être obtenue sur la base de l'enveloppe temporelle de ces signaux extraite au sein de seulement 4 bandes de fréquence indépendantes.
- 24 Si la perception des indices d'enveloppe temporelle semble normale chez les personnes malentendantes, la perception d'informations de structure temporelle fine semble dégradée à la suite de lésions cochléaires. Cela se manifeste essentiellement par des troubles de perception de la hauteur et du timbre des sons, et des troubles de localisation des sons de basse fréquence (localisation fondée sur des différences interaurales de phase des signaux, et donc de structure temporelle fine). Une contribution de ce déficit à la dégradation des capacités d'analyse de scènes complexes est aujourd'hui envisagée, mais reste à confirmer. L'origine de ce déficit est encore méconnue, mais découle vraisemblablement d'altérations des propriétés de synchronisation des décharges des fibres du nerf auditif et/ou d'altérations du recodage central de ces informations de synchronisation neuronale.

Applications audiologiques

- 25 Le modèle de traitement sensoriel en cinq étapes permet de comprendre les conséquences perceptives principales de lésions cochléaires mentionnées ci-dessus, à savoir la perte d'audibilité, le recrutement de sonie, la dégradation de la sélectivité fréquentielle. Il permet également de comprendre certains troubles de perception spatiale et temporelle auxquelles sont associés une susceptibilité accrue aux effets de masquage ainsi que des troubles de perception de la hauteur mélodique et de la compréhension de la parole. Les prothèses auditives numériques et neuroprothèses auditives actuelles implémentent directement ce modèle psychoacoustique afin de compenser (c'est-à-dire, inverser) au mieux les effets des lésions cochléaires. Les mesures psychoacoustiques de seuils absolus, de sélectivité fréquentielle, de fonction de sonie ou encore d'intelligibilité de la parole dans le silence et dans le bruit guident le diagnostic réalisé par le clinicien et le réglage individuel de ces prothèses par l'audioprothésiste, ceci grâce à des batteries de tests cliniques inspirés des mesures psychophysiques originelles.

Prothèses auditives

- 26 Les prothèses auditives visent essentiellement à compenser la perte d'audibilité et le recrutement de sonie en appliquant soit une amplification linéaire, soit une amplification non linéaire (correspondant à un contrôle automatique de gain, AGC) au signal acoustique capté par un microphone et décomposé en plusieurs bandes de fréquence (pour une revue détaillée de la perception sous prothèse auditive, cf. Moore 1995, 2007).
- 27 Le gain apporté par ces prothèses amplificatrices est calculé à partir des mesures de seuils absolus réalisées sur le patient malentendant (par exemple, à partir d'une règle de demi-gain, modifiée afin de limiter les effets de masquage ascendant produits par les sons de basse fréquence). Dans le cas d'une amplification linéaire, les sons les plus intenses issus de chaque bande de fréquence atteignant le seuil d'inconfort sont écrêtés, protégeant ainsi le récepteur auditif d'un éventuel traumatisme acoustique, mais réalisant malheureusement une distorsion du son transmis. Dans le cas d'une amplification non linéaire, le gain appliqué au signal varie en fonction de l'entrée ou de la sortie de la prothèse, et les sons les plus faibles sont plus amplifiés que les sons les plus intenses. Pour chaque bande de fréquence de la prothèse, plusieurs paramètres contrôlent le processus d'amplification non linéaire, à savoir le seuil d'enclenchement de la compression, le taux de compression et les constantes de temps d'attaque (1-100 ms) et de retour (5-10 000 ms) du système AGC. D'après les calculs de l'index STI, les prothèses non linéaires à constantes de temps longues sont *a priori* préférables aux prothèses non linéaires à constantes de temps courtes, car elles minimisent les distorsions des indices d'enveloppe temporelle du signal à l'échelle de temps de la syllabe et maximisent ainsi la réception d'information phonétique. Pourtant, le succès des prothèses non linéaires reste très variable, et ceci quel que soit le choix des constantes de temps du système AGC. En effet, le choix d'une prothèse linéaire ou non linéaire semble dépendre d'un grand nombre de facteurs à la fois sensoriels, cognitifs et écologiques (caractérisant l'environnement sonore quotidien du patient).

- 28 Dans l'ensemble, ces prothèses restaurent une bonne audibilité des sons de l'environnement et des signaux de parole, et une intelligibilité parfaite ou quasi parfaite dans le silence pour bon nombre de pertes auditives modérées ou modérées-sévères. Les prothèses non linéaires rétablissent également un confort auditif au quotidien, indiquant que les systèmes AGC compensent correctement le phénomène de recrutement de sonie. Toutefois, ces prothèses ne restituent pas une perception mélodique et spatiale normale. Elles ne restaurent pas non plus une intelligibilité normale dans le bruit, et ceci plus particulièrement lorsque le bruit de fond présente des fluctuations spectrales et temporelles (comme dans le cas d'un locuteur concurrent ou d'un brouhaha). L'origine de ces limitations est très probablement liée au fait que les prothèses auditives, linéaires ou non, ne restaurent pas une sélectivité fréquentielle normale et une réception normale des informations de structure temporelle fine.
- 29 Les prothèses auditives actuelles tentent de compenser le déficit dans le bruit en réduisant les distorsions des sons cibles induites par les sons masquants par application d'algorithmes de réduction de bruit (par soustraction spectrale, par exemple) ou par atténuation à la source des sons masquants grâce à des microphones directionnels. Toutefois, ces stratégies de réduction de bruit ont un succès limité en condition réelle d'écoute (à savoir, en environnement réverbérant et pour des sons masquants généralement non stationnaires). Des stratégies de rehaussement de la structure spectrale et temporelle des signaux captés visant à compenser la perte de sélectivité fréquentielle et à maximiser la réception des indices d'enveloppe temporelle sont proposées dans les prothèses actuelles, mais à ce jour, leur efficacité n'a pu véritablement être démontrée. On voit les limites de la stratégie d'« inversion » qui vise à compenser les déficits en délivrant un signal dont la structure spectro-temporelle est en quelque sorte modelée en miroir des effets des troubles supraliminaires (afin d'annuler ces effets, une fois les signaux transmis par le système auditif périphérique lésé).

Implants cochléaires et implants du tronc cérébral

- 30 Une oreille bionique ou « implant cochléaire » peut être prescrite dans le cas d'une surdité profonde pour laquelle un appareillage auditif conventionnel n'apporte aucun bénéfice réel (pour une revue détaillée, cf. Zeng, Popper et Fay 2004). Fait remarquable, l'âge d'implantation est aujourd'hui descendu à moins d'un an de vie. L'implant cochléaire est un dispositif numérique stimulant électriquement et directement les fibres nerveuses afférentes du nerf auditif du patient. Le traitement de signal réalisé par le processeur vocal d'un implant cochléaire reproduit parfaitement les différentes étapes de traitement du signal acoustique présentées plus haut. Le signal capté par un microphone est initialement décomposé en plusieurs bandes de fréquences. Il est ensuite compressé en amplitude afin d'adapter l'étendue des variations d'amplitude acoustique (la dynamique acoustique, comprise entre 40 et 80 dB SPL selon la valeur du seuil d'enclenchement et de saturation du dispositif) à la dynamique réduite du nerf auditif en réponse à des stimulations électriques (2-12 dB). L'enveloppe temporelle de ces signaux de bande est ensuite extraite et utilisée pour moduler l'amplitude de trains d'impulsions électriques de cadence fixe, trains d'impulsions délivrés aux fibres du nerf auditif grâce à des électrodes (une vingtaine, environ) implantées chirurgicalement de manière à respecter autant que possible la tonotopie cochléaire.

- 31 Plusieurs stratégies de codage peuvent être implémentées au sein d'un implant cochléaire donné. La stratégie SPEAK (*Spectral Peak*) utilise une cadence de stimulation relativement basse et ne transmet que six à dix maxima spectraux du signal. La stratégie CIS (*Continuous Interleaved Sampling*) utilise une cadence de stimulation relativement rapide sur des canaux de stimulation fixes. La stratégie ACE (*Advanced Combination Encoder*) utilise une cadence de stimulation relativement rapide et transmet beaucoup de maxima spectraux du signal. Si le processeur vocal des implants cochléaires actuels utilise une vingtaine de canaux fréquentiels (et ainsi une vingtaine d'électrodes de stimulation), les données psychoacoustiques révèlent que les patients ne disposent – au mieux – que de 8 canaux de transmission d'information indépendants (comme c'est le cas des personnes atteintes de surdité modérée). Comme nous l'avons indiqué plus haut, ceci est amplement suffisant pour obtenir une intelligibilité parfaite ou quasi parfaite de mots ou de phrases dans le silence, conformément aux observations cliniques et à l'existence d'une corrélation forte entre les mesures d'intelligibilité et de sensibilité auditive aux modulations d'amplitude lentes et rapides (les fluctuations d'enveloppe temporelle) des sons réalisées chez les patients implantés cochléaires. Toutefois, la perception mélodique, la localisation de sources sonores et l'intelligibilité en présence de bruit, de signaux de parole concurrents ou en milieu réverbérant restent extrêmement pauvres, en raison d'une mauvaise résolution et sélectivité fréquentielles et d'un accès limité aux informations de structure temporelle fine du signal (pour chaque canal fréquentiel, celles-ci sont remplacées par un train d'impulsions de cadence fixe). Plusieurs efforts sont menés afin d'augmenter la résolution fréquentielle (le nombre de canaux) et la transmission de la structure temporelle fine des implants, mais les résultats actuels ne sont pas probants.
- 32 Dans le cas où un implant cochléaire ne peut être proposé (car le nerf auditif du patient ne fonctionne pas), une stimulation électrique peut être délivrée sur le premier noyau du tronc cérébral, à savoir le noyau cochléaire, grâce à des électrodes de surface ou des électrodes pénétrantes. La stimulation électrique vise également à respecter autant que possible l'organisation tonotopique du noyau cochléaire. Cet implant, dénommé implant du tronc, permet de restaurer des perceptions auditives plus élémentaires que celles obtenues sous implant cochléaire. L'intelligibilité de la parole dans le silence est généralement pauvre, mais la perception de sonie et de hauteur et l'acuité temporelle des patients appareillés s'avèrent aussi bonnes que celles des patients implantés cochléaires. Les mesures d'intelligibilité dans le silence s'améliorent notablement chez un grand nombre de patients lorsque certaines structures du noyau cochléaire sont intactes, ce qui amène à spéculer sur l'existence d'une voie spécialisée pour le traitement des indices d'enveloppe temporelle (voie du « quoi » auditif).

Prothèses hybrides électro-acoustiques

- 33 Une nouvelle approche de la réhabilitation des personnes malentendantes présentant des pertes neurosensorielles sévères-profondes en haute fréquence vise à restaurer l'identification de la parole dans le bruit en préservant une audition « acoustique » résiduelle dans l'oreille implantée. Ici, une stimulation électrique (E) dans la région haute fréquence (via un implant cochléaire standard) est combinée à une stimulation acoustique amplifiée (A) par prothèse auditive conventionnelle dans la région basse fréquence. Plusieurs études montrent que le fait de préserver après implantation une audition acoustique résiduelle dans l'oreille implantée est un objectif réaliste pour de

nombreux patients présentant une perte auditive sévère en haute fréquence. Les premiers résultats montrent une amélioration importante de l'intelligibilité de la parole dans le bruit chez les patients porteurs de prothèses hybrides A + E. Ces résultats confirment l'idée selon laquelle les avantages de cette nouvelle approche résultent principalement de l'apport d'informations spectrales et temporelles fournies par la stimulation acoustique résiduelle par rapport à une stimulation électrique (plus pauvre sur ces deux dimensions spectrales et temporelles). En effet, la stimulation acoustique en basse fréquence améliore la résolution fréquentielle et peut-être, dans certains cas, la réception d'informations de structure temporelle fine.

Conclusions

- 34 Le modèle de traitement sensoriel auditif décrit par les psychoacousticiens a conduit à plusieurs applications permettant de prédire la compréhension de la parole dans de nombreuses conditions d'écoute (un auditorium, par exemple). Ce modèle permet également de comprendre les conséquences perceptives principales des lésions cochléaires. Le succès des différents dispositifs de réhabilitation implémentés à partir du modèle valide celui-ci en retour et constitue un cas exemplaire de transfert de connaissances fondamentales dans les domaines industriel, biomédical et clinique. Toutefois, l'incapacité de ces dispositifs à restaurer une perception normale des informations mélodiques, de la localisation des sources sonores dans l'espace et des signaux de parole en présence de bruit de fond constitue une importante limite et appelle à un dépassement du modèle. Les traitements spectraux et temporels réalisés par le système auditif périphérique et central sont encore incomplètement compris. Par ailleurs, la contribution des mécanismes d'analyse des scènes auditives et des processus cognitifs (McAdams et Bigand 1994 ; Yost, Fay et Popper 2008) demande à être prise en compte. Ainsi, la contribution de mécanismes ascendants (*bottom-up*) et descendants (*top-down*) permettant de grouper certaines informations et de restaurer d'autres informations manquantes ou masquées, ainsi que la contribution de mécanismes de sélection attentionnelle et celle des traitements en mémoire auditive, semblent essentielles.

BIBLIOGRAPHIE

Références bibliographiques

Blauert, J., 1997. *Spatial Hearing: The Psychophysics of Human Sound Localization*, Cambridge (Mass.), The MIT Press.

Bregman, A. S., 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound*, Cambridge (Mass.), The MIT Press (Bradford Books).

- Casati, R. et Dokic, J., 1994. *La philosophie du son*, Nîmes, J. Chambon.
- Hartmann, W. M., 2005. *Signals, Sound, and Sensation*, 5^e impression corrigée, New York, AIP Press.
- Helmholtz, H. L. von, 1863. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, Braunschweig, Wieweg und Sohn. [Traduction française : *Théorie physiologique de la musique*, Paris, Masson, 1868, réimpression, Sceaux, J. Gabay, 1990.]
- McAdams, S. et Bigand, E., 1994. *Penser les sons*, Paris, PUF (Psychologie et sciences de la pensée).
- Møller, A. R., 2006. *Hearing : Anatomy, Physiology, and Disorders of the Auditory System*, 2^e édition, Amsterdam/Londres/Boston, Academic Press.
- Moore, B. C. J., 1995. *Perceptual Consequences of Cochlear Damage*, Oxford, Oxford University Press (Oxford Psychology Series ; 28).
- Moore, B. C. J., 2003. *An Introduction to the Psychology of Hearing*, 5^e édition, Amsterdam/Londres/Boston, Academic Press.
- Moore, B. C. J., 2007. *Cochlear Hearing Loss : Physiological, Psychological and Technical Issues*, 2^e édition, Hoboken (N.J.), Wiley.
- Pickles, J. O., 2008. *An Introduction to the Physiology of Hearing*, 3^e édition, Bingley, Emerald.
- Plack, C. J., 2005. *The Sense of Hearing*, Mahwah (N.J.), Erlbaum.
- Plack, C. J., Oxenham, A. J., Fay, R. R. et Popper, A. N. (éd.), 2005. *Pitch : Neural Coding and Perception*, New York, Springer (Springer Handbook of Auditory Research ; 24).
- Yost, W. A., Fay, R. R. et Popper, A. N. (éd.), 2008. *Auditory Perception of Sound Sources*, New York, Springer (Springer Handbook of Auditory Research ; 29).
- Zeng, F.-G., Popper, A. N. et Fay, R. R. (éd.), 2004. *Cochlear Implants : Auditory Prostheses and Electric Hearing*, New York, Springer (Springer Handbook of Auditory Research ; 20).

NOTES

1. Nous remercions le CNRS, l'université Paris Descartes, l'École normale supérieure et le Laboratoire de psychologie de la perception (UMR 8158) pour leur soutien. Nous remercions également Daniel Pressnitzer pour ses remarques et suggestions sur une première version de ce chapitre.

AUTEUR

CHRISTIAN LORENZI

Laboratoire de psychologie de la perception (LPP), Université Paris Descartes, Département d'études cognitives de l'École normale supérieure, 29, rue d'Ulm, 75 005 Paris