



La Perception Auditive

(cours P.A.M.U.)

Alain Goyé

Janvier 2002

TABLE

PHYSIOLOGIE ET PERCEPTION DE L'INTENSITÉ SONORE	3
1. Acoustique et perception auditive	3
1.1. Rôles de la perception auditive dans la vie courante	3
1.2. Place de la perception auditive dans les domaines scientifique et industriel	4
1.3. Grandeurs perceptives	5
2. Physiologie de l'audition	6
2.1. La tête	6
2.2. L'oreille externe	6
2.3. L'oreille moyenne	7
2.4. L'oreille interne	9
2.5. Le nerf cochléaire (ou nerf auditif)	11
2.6. Les voies auditives centrales	11
3. Etude de la sensation d'intensité sonore	12
3.1. Le champ audible	12
3.2. Echelle de bruit	14
3.3. Les seuils auditifs	14
3.4. Les courbes d'isotonie	15
3.5. L'audiogramme	16
3.6. La Sonie	17
3.7. Le Seuil Différentiel d'Intensité	19
3.8. Facteurs de variation de la sonie	21
4. Masquage et Bandes Critiques	26
4.1. Les Bandes Critiques	26
4.2. Le Masquage	28
4.3. Les filtres auditifs	30
5. Perception d'intensité et perte auditive	33
5.1. Le recrutement ou rattrapage de la sonie	34
5.2. L'adaptation pathologique	35
5.3. L'altération des filtres auditifs	35
PERCEPTION DE LA HAUTEUR DES SONS	38
1. Définitions	38
2. Hauteur tonale des sons purs	39
2.1. L'échelle de tonie : le Mel	39
2.2. Facteurs de variation de la hauteur tonale	41
2.3. Seuils différentiels de hauteur	42
3. Hauteur des sons complexes	44
3.1. Hauteurs virtuelles de sons harmoniques	44
3.2. Sons non harmoniques	45
3.3. Ecoute analytique et écoute synthétique	46
4. Modèles pour la perception de la hauteur	47
4.1. La tonotopie et les modèles de reconnaissance de formes	47
4.2. Le codage temporel	48
4.3. Nécessité d'un modèle mixte	48

5. Intervalles et gammes	49
5.1. Construction des gammes	49
5.2. L'oreille absolue	51
5.3. Préférences	51
PERCEPTION DU TIMBRE DES SONS	52
1. Consonance et dissonance	52
1.1. Consonance des sons purs	52
1.2. Consonance des sons complexes	53
2. La notion de timbre	55
2.1. Timbre et composition spectrale	55
2.2. Timbre et enveloppe temporelle	55
3. Vers un espace des timbres	56
ECOUTE BINAURALE ET PERCEPTION DE L'ESPACE SONORE	57
1. La localisation des sons	57
2. Autres (in-)capacités binaurales	65
3. L'étendue subjective des sources	68
ANNEXE : MÉTHODES POUR L'ÉTUDE DES SENSATIONS AUDITIVES	71

Physiologie et perception de l'intensité sonore

1. Acoustique et perception auditive

L'acoustique est une science passionnante par son interaction avec une multiplicité de domaines très variés. Dans certains de ces domaines les problèmes posés sont si complexes que l'acoustique ne se développe véritablement avec profit que depuis quelques dizaines d'années. Cette complexité concerne aussi bien les phénomènes physiques, que les mécanismes du vivant, en particulier ceux de la perception auditive, qui ont évidemment une grande importance dans la vie de tous les jours. La branche de l'acoustique qui traite de la perception des sons s'appelle la psychoacoustique. Cette science étudie le fonctionnement de l'appareil auditif considéré comme une « boîte noire », en recherchant les relations entre les propriétés physiques des sons reçus et les sensations évoquées. Evidemment, elle est éclairée et complétée par l'étude des mécanismes mis en jeu, dans la physiologie et psychologie. Dans tous ces domaines il nous reste encore beaucoup à apprendre.

1.1. Rôles de la perception auditive dans la vie courante

La perception auditive remplit un grand nombre de fonctions diverses :

- information sur l'environnement, les objets qui nous entourent : le bruit d'un choc sur un objet nous donne des indications sur son matériau, sa structure interne, son élasticité... Par exemple (tiré d'un article de Stephen McAdams) : une pile de vaisselle glisse d'une table, les assiettes s'entrechoquent dans l'air avant de s'écraser sur une surface dure où toutes se cassent, sauf une. A partir des seuls indices sonores, tout auditeur baignant dans une culture utilisant ces objets dans le même environnement, saura reconnaître sans difficulté la nature de l'événement, les objets intervenant dans celui-ci et la transformation qu'ils y subissent.

- alerte, détection, informations sur la proximité et la direction des sources sonores. La manière dont nous percevons ces sources sont perçues différemment suivant le contexte la situation de l'auditeur, comme le montrent les deux exemples suivants :

1. un coup de klaxon automobile sera généralement ignoré ou perçu comme gênant si l'on est chez soi... à moins que sa voiture soit mal garée devant la maison. Le même coup de klaxon prendra une signification différente si l'on traverse la rue, ou si l'on est au volant, attendant distraitement au feu rouge.

2. dans une conversation que nous n'écoutions pas jusqu'alors, la simple prononciation de notre nom peut focaliser soudain notre attention.

- reconnaissance de sources : la caractéristique sonore d'une source nous permet de la reconnaître parmi de multiples semblables. On parle de signature acoustique de la source. Nous reconnaissons

ainsi les voix, les pas de personnes connues, les cris d'un bébé, la sonnerie personnalisée de notre téléphone mobile, le bruit du moteur du camion-poubelles... Des spécialistes, les « oreilles d'or », sont entraînés à reconnaître la signature acoustique des bateaux et sous-marins.

- appréhension d'espaces clos : les sons perçus, combinés avec la connaissance a priori que nous avons de leur source, nous renseignent sur l'espace environnant : son volume, les revêtements muraux, sa forme, etc.

- la notion de confort acoustique est essentielle dans les lieux voués à l'audition (amphis, auditoriums) mais prend également de plus en plus d'importance dans tous les lieux de vie (habitations, habitacles de véhicules, espaces de travail, lieux de réunion...)

- la communication enfin est un des rôles essentiels de l'audition car l'homme est un animal social.

La communication sonore passe par:

1. l'intelligibilité proprement dite de la parole, mais aussi:
2. le timbre, qui donne des indications sur le locuteur (âge, sexe, état de fatigue, identité).
3. l'intonation, qui contribue au sens (interrogation, exclamation... sans parler des langues tonales) mais aussi qui exprime l'humeur ou les sentiments.

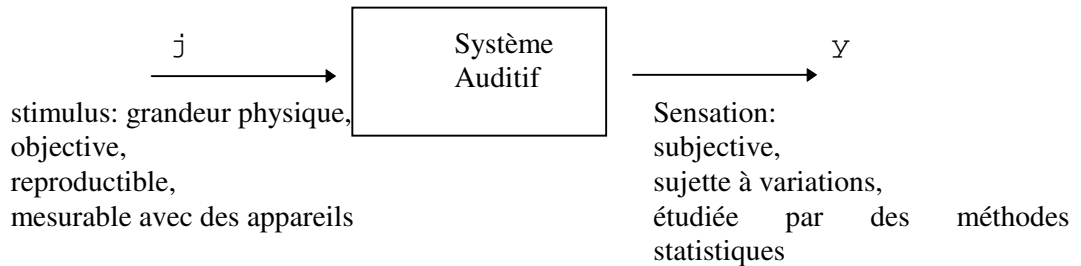
1.2. Place de la perception auditive dans les domaines scientifique et industriel

La psychoacoustique trouve des applications dans tous les domaines où intervient le son audible ; on peut citer, sans être exhaustif :

- Audiovisuel : enregistrement, diffusion, reproduction sonores...
- Architecture et environnement : réduction du bruit, étude et contrôle des « ambiances » sonores...
- Télécommunications : codage et transmission de messages sonores, qui sont essentiellement de deux types : vocal et musical.
- Informatique: toutes les applications multimédia, du jeu à la visioconférence.
- Médecine : évaluation et réhabilitation des déficiences de la fonction auditive (aides auditives, implants cochléaires qui permettent par stimulation électrique du nerf auditif d'évoquer des sensations sonores chez des sourds profonds).
- Acoustique sous-marine : reconnaissance de signatures acoustiques (« oreilles d'or »)
- Automobile : champ acoustique de l'habitacle, réduction des bruits dans certaines zones et certaines gammes de fréquences, esthétique des bruits (une grande attention est portée par exemple à la qualité sonore du claquement de portière qui peut avoir un effet plus ou moins sécurisant).
- Industrie : casques et bouchons anti-bruit pour le confort et la sécurité des personnes.
- Aéronautique : spatialisation des sons au casque pour la différenciation et l'intelligibilité des informations sonores.
- Communication des plongeurs sous-marins : ceux-ci peuvent se parler et s'entendre grâce à un vibreur placé dans l'embout respiratoire ; la transmission se fait également par ondes acoustiques.
- Musique: synthèse sonore (l'enjeu est de fabriquer de nouveaux instruments pour de nouveaux sons et une meilleure manipulation des sons correspondant à nos paramètres perceptifs), restauration d'enregistrements, facture instrumentale...

1.3. Grandeurs perceptives

A chaque grandeur physique du son on peut essayer d'associer une grandeur subjective, une grandeur de la perception:



Les principales grandeurs perceptives et leurs correspondants physiques (lorsqu'ils existent) sont regroupées dans le tableau ci-dessous :

Grandeur	Unité	Perception de	Vocabulaire	Grandeur physique principale associée	Unité physique
isotonie	phones	égale intensité	Fort / faible	-	dB
sonie	sones	intensité / loudness		niveau SPL	dB
tonie	tones / mels	hauteur / pitch	Grave / aigu	fréquence	Hz
<i>multiplés</i>	- - -	timbre	« mat, chaud, cuivré... »	- - -	- - -
chronie	-	durée / duration	Court / long	temps	s

Ces associations ne sont pas rigoureuses : on constate expérimentalement que la sonie (intensité subjective) dépend du niveau SPL mais aussi de la (ou les) fréquence(s), du timbre et de la durée du son. Il en va de même pour les autres grandeurs perceptives.

Notons encore que malgré l'existence d'un vocabulaire spécialisé certains abus de langage courants peuvent favoriser les confusions ; ainsi on parle souvent d'intensité perçue plutôt que de sonie. Cette « intensité » là n'est PAS l'intensité acoustique NI le niveau d'intensité acoustique !

2. Physiologie de l'audition

2.1. La tête

Le son doit d'abord contourner la tête avant d'arriver à l'oreille. Il subit à cette occasion:

- une **atténuation** (sélective en fonction des fréquences),
- un **retard** qui est différent sur les deux oreilles, et qui se traduit pour les sons périodiques par des différences de phase :

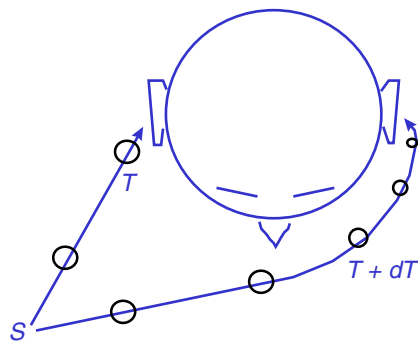


Figure : retard et atténuation du son par la tête du sujet

Cette atténuation et ce retard jouent un rôle essentiel dans la localisation des sources (tant dans l'évaluation de la direction que de la distance).

Ils constituent un filtrage auquel s'ajoute celui de l'oreille externe.

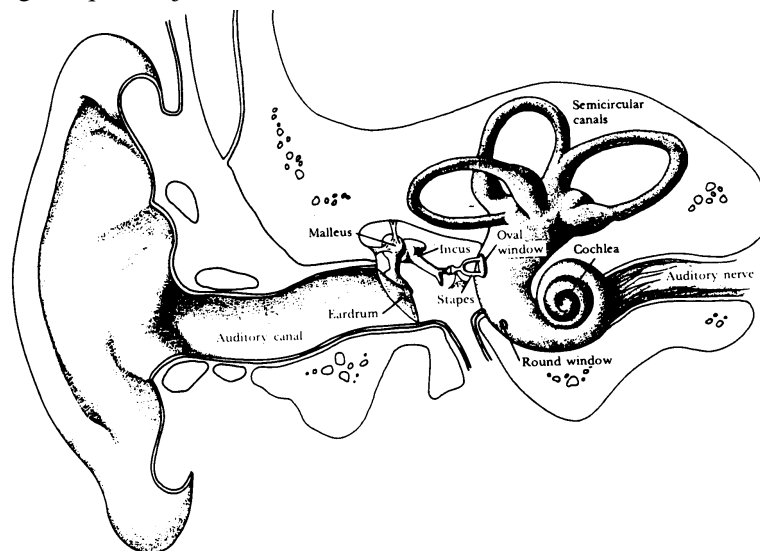


Figure : vue générale de l'oreille

2.2. L'oreille externe

Celle-ci se compose du pavillon et du conduit auditif externe.

- le pavillon (orientable chez certains animaux, mais pas chez l'homme) réalise un filtrage sélectif suivant la direction d'incidence du son et sa fréquence.

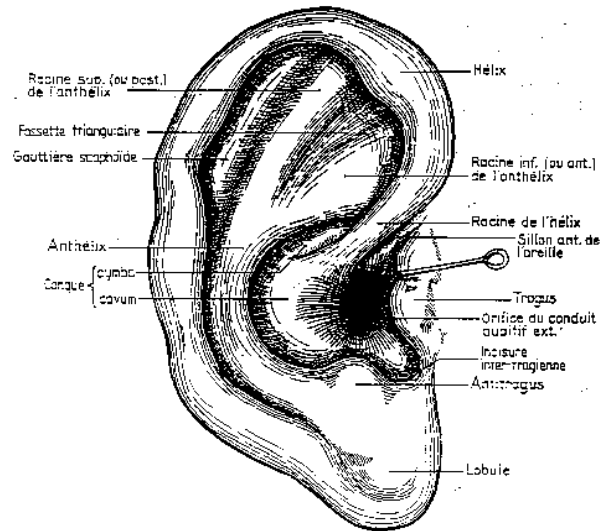
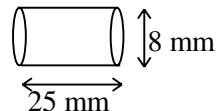


Figure : le pavillon de l'oreille humaine

- le conduit auditif externe, peut être modélisé comme un cylindre fermé à une extrémité par le tympan :



Le C.A.E. joue un rôle de guide d'onde, jusqu'au tympan.

L'ensemble de l'oreille externe a pour effet une augmentation de l'intensité sonore au niveau du tympan, de quelques dB entre 1,5 et 7 kHz. (avec un pic vers 6 kHz dû au pavillon, et un autre vers 2,5 kHz dû au C.A.E.).

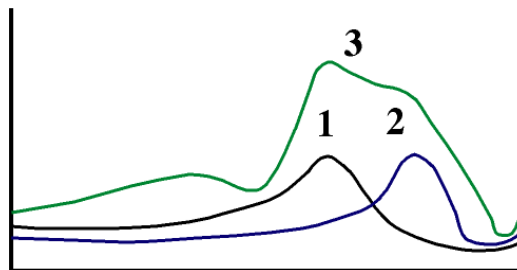


Figure : les résonances du pavillon (2 : $\sim +10$ dB vers 6000 Hz), du C.A.E. (1 : $\sim +10$ dB vers 2500 Hz) et totale de l'oreille externe (3) pour un son incident dans l'axe du pavillon.

L'ensemble (tête et oreille externe) joue un rôle dans la perception de l'espace. C'est pourquoi on utilise dans de nombreux cas une tête artificielle (à défaut d'une vraie tête) pour faire des test ou des enregistrements.

- L'analyse et la reproduction des fonctions de filtrage remplies par ces deux éléments permettent la spatialisation artificielle du son, selon une méthode étudiée plus loin dans ce cours.

2.3. L'oreille moyenne

L'oreille moyenne est une cavité d'air dans un os, le **rocher**, qui renferme un système articulé de trois **osselets** (les plus petits os du corps humain):

- le **marteau** (20 g.) : au contact du tympan. Tenu par des ligaments
- l'**enclume** (25 g.) : Tenu par des ligaments
- l'**étrier** (5 g.) : attaché au bout de l'étrier, et solidaire de la fenêtre ovale

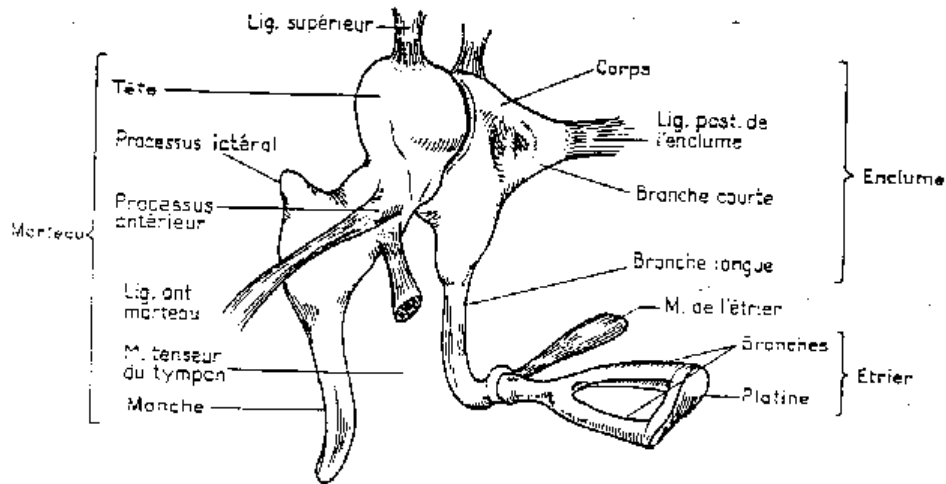


Figure : les osselets de l'oreille moyenne

L'oreille moyenne joue un rôle **d'amplification et d'adaptation d'impédance** par :

- un effet de levier, qui est en fait assez faible chez l'homme.
- et surtout, le rapport des surfaces du tympan (65 mm^2) et de la fenêtre ovale (3 mm^2) environ de 20.

Au total, la pression au niveau de la fenêtre ovale est ~ 24 fois plus grande qu'au niveau du tympan, ce qui représente un gain de 27,5 dB. Ce gain est essentiel pour l'adaptation d'impédance entre les milieux aérien (oreille moyenne) et liquide (oreille interne), sans laquelle 99% de l'énergie serait réfléchi au niveau de l'interface.

On estime que c'est sur la plage des fréquences de la parole (250-6000 Hz) que l'adaptation d'impédance est la meilleure et que sur les fréquences 1000-2000 Hz, 46 % de l'énergie est transmis.

L'oreille moyenne joue également un rôle essentiel de **protection** de deux manières :

- la limitation mécanique naturelle des mouvements des osselets, attachés les uns aux autres et tenus par des ligaments.
- une limitation par contraction d'un muscle lié à l'étrier.

C'est le **réflexe stapédien**, qui a pour effet de protéger l'oreille interne contre les bruits de trop forte intensité. Il se déclenche, aux fréquences usuelles, autour de 80 - 90 dB, et produit une atténuation jusqu'à 40 dB. Il diminue ainsi la fatigue auditive et améliore le pouvoir de discrimination de l'oreille aux fortes intensités.

Note 1.: un autre muscle, lié au marteau, participe à ce rôle de protection en se contractant.

Note 2.: le son peut également être transmis directement par vibration du rocher et des os de l'ensemble de la boîte crânienne ; c'est la conduction osseuse, qui explique en particulier que notre propre voix nous paraît différente « en direct » et enregistrée.

Cette conduction osseuse est utilisée pour la transmission du son dans certaines prothèses auditives comme la *B.A.H.A.* (*Bone-Anchored Hearing Aid*).

Note 3.: la trompe d'Eustache, qui s'ouvre à la déglutition, permet ainsi l'équilibrage des pressions de part et d'autre du tympan, ce qui est nécessaire pour optimiser les vibrations de celui-ci. La trompe d'Eustache est aussi un lieu où sont sécrétés des agents anti-infectieux.

2.4. L'oreille interne

C'est dans l'oreille interne que l'énergie mécanique est transformée en énergie bioélectrique, c'est-à-dire en potentiels d'action nerveux.

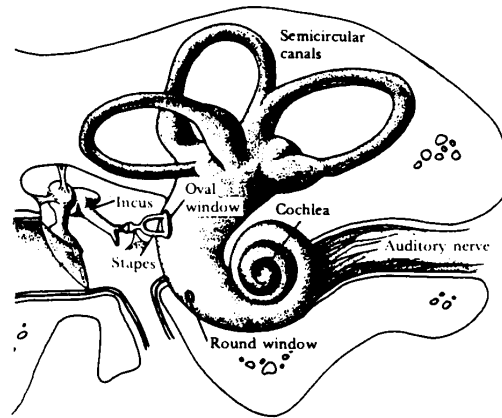


Figure : vue extérieure de l'oreille interne

L'oreille interne se compose :

- de l'appareil vestibulaire, comprenant les trois canaux semi-circulaires visibles sur la figure, qui joue un rôle important pour l'équilibre mais n'intervient pas dans l'audition.
- de la cochlée, qui a globalement la forme d'un canal en colimaçon, d'une longueur déroulée de 2,5 à 3 centimètres, divisé en deux dans sa longueur par une lame osseuse à laquelle s'attachent deux membranes : la **membrane basilaire** et la **membrane tectorielle**.

La membrane basilaire sépare le canal de la cochlée en deux rampes remplies de liquide. Ces rampes communiquent au sommet du limaçon, l'**apex**, par un orifice : l'**hélicotréma**, qui assure l'équilibre des pressions.

A l'extrémité du limaçon, la base, ces rampes sont fermées et séparées de l'oreille moyenne chacune par une membrane, respectivement : la **fenêtre ovale**, sur laquelle s'appuie l'étrier, et la **fenêtre ronde**, qui est libre.

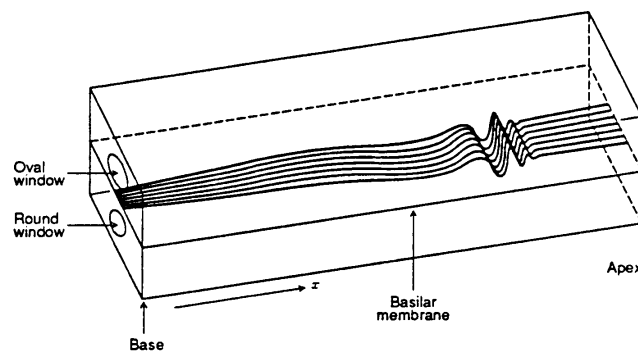


Figure : schéma de la cochlée déroulée, divisée par la membrane basilaire

Le liquide remplissant la cochlée est incompressible. Lorsque l'étrier transmet à la fenêtre ovale les vibrations sonores, les variations de pression sont appliquées simultanément sur toute la longueur de la membrane basilaire. Cependant celle-ci a des propriétés mécaniques variables : plus mince et raide vers la base, elle devient progressivement plus large et souple vers l'apex. En conséquence, les variations de pression provoquent l'apparition d'ondes progressives de déformation de la membrane, de la base vers l'apex. L'enveloppe de ces ondes varie avec la composition fréquentielle du son. Pour un son pur, cette enveloppe présente un maximum en un

point d'autant plus proche de la base (resp. de l'apex), que le son est aigu (resp. grave). C'est ce qu'on appelle la **tonotopie** cochléaire.

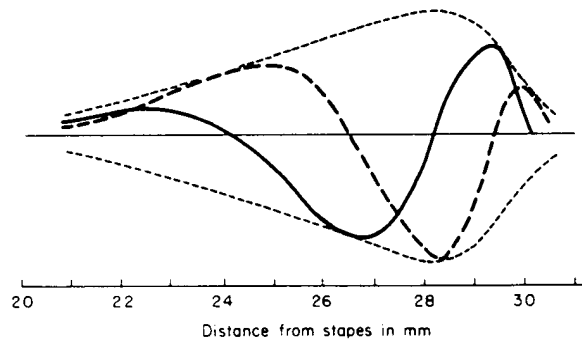


Figure : onde progressive le long de la membrane basilaire

Entre la membrane basilaire et la membrane tectorielle se trouve l'organe de Corti, schématisé ci-dessous. Le principal élément en est les **cellules ciliées**, (ou cellules de Corti), cellules implantées dans la membrane basilaire qui se prolongent par des cils.

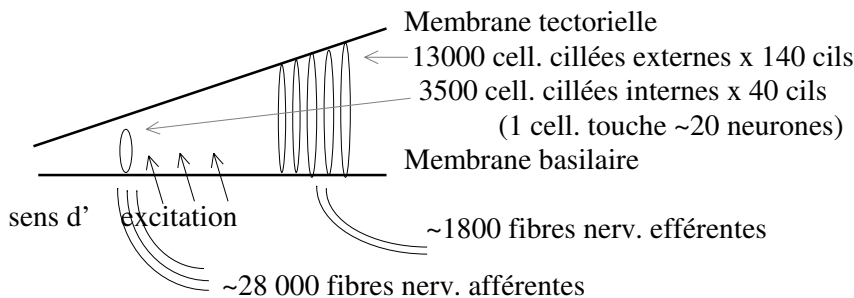


Figure : coupe schématique de l'organe de Corti.

Les cellules ciliées sont de deux types:

- les **cellules ciliées internes (CCI)**, dont les cils flottent librement dans le liquide ; elles sont alignées sur une seule rangée. Quand les vibrations de la membrane basilaire les agitent, elles libèrent des ions qui vont dépolariiser les terminaisons du nerf auditif (vers le centre du limaçon).

Ces cellules sont peu nombreuses (environ 3500), fragiles (les sifflements d'oreille traduisent souvent qu'elles ont souffert d'une exposition au bruit) et ne se régènèrent pas une fois détruites. Il faut donc les ménager : leur disparition équivaut à une surdité totale.

Les CCI ont des propriétés essentielles pour la perception, en particulier:

- la **phase réfractaire** (durée minimale de récupération entre deux décharges d'ions)
- le seuil d'excitation, qui présente une résonance pour une fréquence donnée: la « **fréquence caractéristique** » de la cellule.

Cependant les CCI seules ne rendent pas compte de la très bonne sélectivité fréquentielle de notre oreille.

- les **cellules ciliées externes (CCE)**, dont l'extrémité des cils est solidaire de la membrane tectorielle ; elles sont réparties sur 3 rangées.

Elles ont des propriétés remarquables aux conséquences fondamentales, à travers deux mécanismes actifs :

1. les **contractions rapides** qui se produisent de manière réflexe, en phase avec les vibrations de la membrane basilaire, pour amplifier celles-ci très sélectivement et augmenter ainsi:

- la **sensibilité**, avec un gain de 40 à 50 dB, surtout aux faibles niveaux ;
- la **sélectivité** : grâce à elles, un petit nombre de CCI est excité dans une zone étroite et libère un maximum d'énergie.

2. les **contractions lentes**, qui passent par le système efférent médian et agissent sur les deux oreilles. Les contractions lentes ont un effet d'atténuation des contractions rapides. Elles permettent d'expliquer certains phénomènes (adaptation de la sonie, protection contre les stimuli trop intenses, otoémissions = émission de sons par l'oreille).

Il faut donc retenir que ce sont les CCE qui rendent l'audition **forte et claire**.

2.5. Le nerf cochléaire (ou nerf auditif)

Le nerf cochléaire forme, avec la juxtaposition du nerf vestibulaire, la 8^{ie} paire crânienne.

Ses fibres transmettent les potentiels d'action au système nerveux central.

La réponse d'une fibre en fonction de la fréquence constitue une « **courbe d'accord** » : ces courbes présentent des résonances très fortes, du fait de la tonotopie de la membrane basilaire, rehaussée par l'action des CCE.

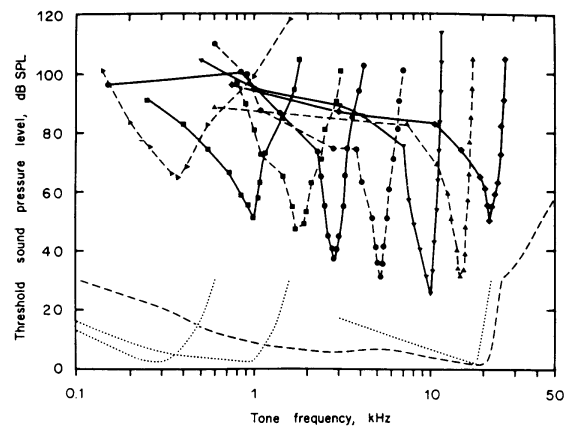


Figure : courbes d'accord des fibres du nerf auditif mesurées des chats anesthésiés par Evans (1975) : seuil d'excitation (dB SPL) en fonction de la fréquence (kHz, échelle logarithmique).

2.6. Les voies auditives centrales

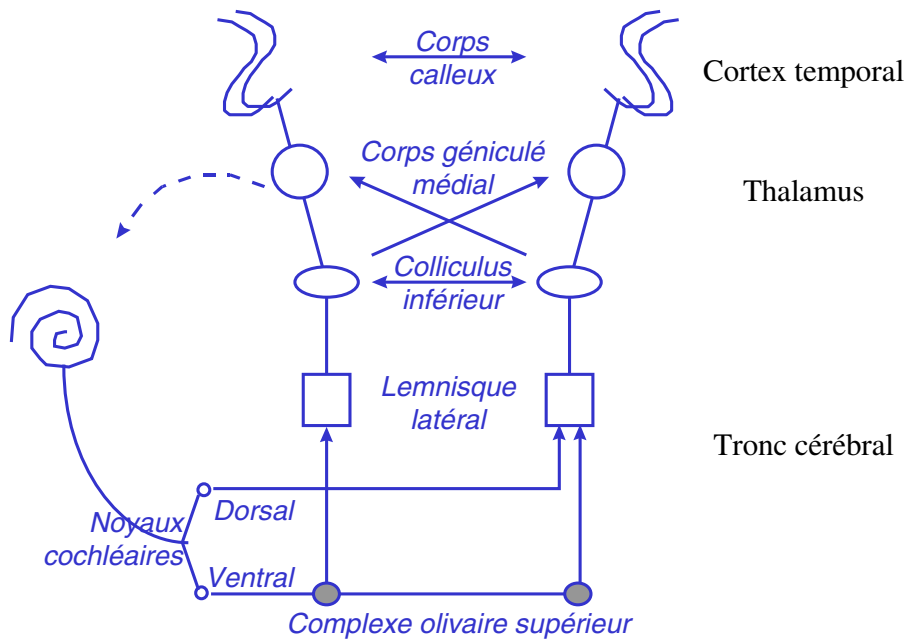


Figure : représentation schématique des voies auditives centrales

De la périphérie vers le centre, l'influx nerveux émis dans la cochlée traverse les étages suivants :

- les noyaux cochléaires qui réalisent une analyse et un codage des propriétés physiques du son. On trouve là de nombreuses connexions de type convergente-divergente (un grand nombre d'afférences et un grand nombre d'efférences) Il y a là des neurones spécialisés, sensibles soit à une certaine fréquence, soit à une certaine intensité ; d'autres sont sensibles à des variations de fréquence ou d'intensité, voire à certains sons complexes. On y retrouve une tonotopie (regroupement géographique des neurones suivant les caractères du son auxquels ils sont sensibles, par exemple la fréquence).
- les noyaux olivaires, qui jouent un rôle dans la localisation. C'est là qu'ont lieu les premières afférences binaurales (homolatérales activatrices, controlatérales inhibitrices) ; certains neurones sont sensibles à des différences interaurales de temps (latences, phases) ou d'intensité. De ces centres partent des connexions vers d'autres noyaux nerveux comme le nerf VII qui transmet le réflexe stapédien.

Les centres immédiatement supérieurs sont moins bien connus :

- le lemnisque latéral, dans lequel on retrouve toujours une certaine tonotopie,
- le colliculus inférieur, où l'on observe des cartes en azimuth (direction d'incidence du son), qui joue donc un rôle important dans la localisation.
- dans le corps géniculé médial (immédiatement sous-cortical), on trouve ensuite des étages répondant exclusivement à différents sons complexes, tels que des voyelles ou des consonnes pour la compréhension de la parole.
- enfin, les aires auditives corticales réalisent un traitement complexe dans lequel on retrouve de nombreuses connexions de type convergente-divergente. Le traitement y est largement parallèle, et réparti sur les aires avec une large interconnexion des aires y compris à travers les hémisphères ; il n'y a pas de spécificité exclusive des aires, et même leurs rôle et fonctionnement peuvent largement évoluer par plasticité.

Ces éléments sont résumés dans le tableau suivant :

noyaux cochléaires bulbaires dorsal et ventral	décodage de l'intensité, la durée, la fréquence
complexe olivaire supérieur et colliculus inférieur	localisation du son dans l'espace (fondée sur les connexions bilatérales)
thalamus et cortex cérébral	décodage plus complexe, interprétation; intervention de la mémoire acquise... détermination de la réponse comportementale

Figure : rôle essentiel des voies auditives centrales

3. Etude de la sensation d'intensité sonore

3.1. Le champ audible

Les champs de l'audition, de la musique et de la parole sont représentés sur la figure ci-dessous, dans le plan harmonique.

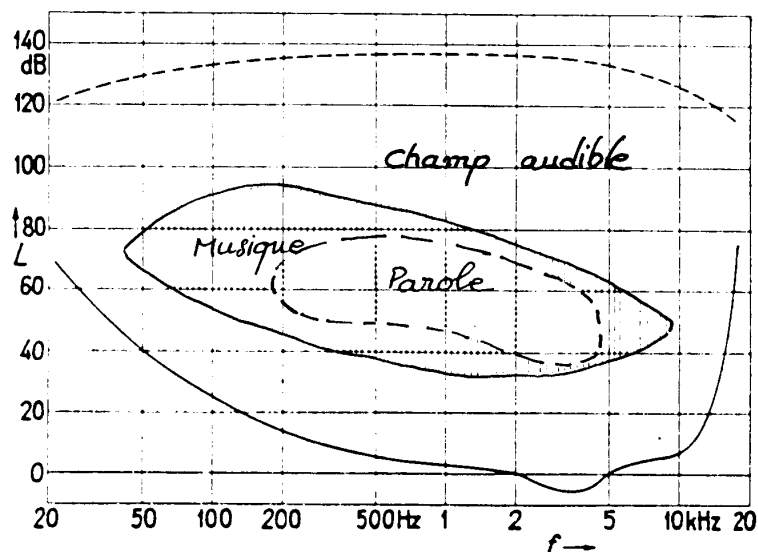


Figure : champ audible, champs de la musique et de la parole (abscisses : fréquences en Hz, ordonnées : niveaux sonores en dB).

Le champ audible est délimité inférieurement par les seuils auditifs, supérieurement par les seuils maximums de confort ou les seuils de douleur. Ces derniers sont rarement mesurés car la mesure implique un risque pour l'oreille ; aussi la limite (en pointillés) est-elle définie de manière moins précise.

Entre ces deux limites, on voit qu'aux fréquences moyennes la dynamique de l'oreille est de près de 120 dB ; cela signifie que nous sommes capables d'entendre des sons dans un rapport de puissances de 1 à 10^{12} (soit 1000 milliards) !

En fréquences, le champ audible s'étend environ de 20 à 20 000 Hz. En pratique ces limites, surtout vers les hautes fréquences, sont valables que pour des sujets jeunes et en bonne santé. Avec l'âge, notre sensibilité auditive décroît, particulièrement dans les aigus. La perte auditive courante chez les sujets âgés s'appelle la presbycusie.

Il faut noter également que les courbes de seuils auditifs sont des moyennes statistiques sur un grand nombre de sujets, dont on a tiré des normes. Il est courant pour des jeunes d'entendre dès un niveau inférieur de 10 dB à la norme, et on ne commence à parler de perte auditive que pour des seuils supérieurs d'au moins 20 dB à la norme.

Du reste notre capacité à entendre les sons aigus se traduit surtout de manière qualitative dans la vie courante : la présence des fréquences supérieures à 8-10 kHz dans la musique joue plus sur la « dureté » de l'enveloppe temporelle des sons impulsifs que sur le timbre des instruments (voir le cours sur le timbre).

Lorsque l'on descend vers les graves (au dessous de 16-20 Hz), la sensation sonore est progressivement remplacée par une perception tactile des vibrations, dont nous pouvons quasiment suivre la forme d'onde. La période devient trop grande pour que le son soit intégré comme tel par l'oreille, dont la « constante de temps » est de l'ordre de 50 ms.

Dans certains cas il est possible que ces sons graves soient également détectés par les produits de distorsion (harmoniques) fabriqués par l'oreille moyenne, dont la réponse n'est pas parfaitement linéaire.

Le champ de la musique s'étend de 50 à 10 000 Hz environ, celui de la parole est plus restreint : l'essentiel de l'énergie est entre 200 à 5000 Hz, et la restriction à la bande [300, 3400] Hz qui est celle du téléphone altère peu l'intelligibilité pour un sujet qui entend normalement.

3.2. Echelle de bruit

Comme nous venons de le voir, la dynamique du champ audible d'environ 100 à 120 dB, ce qui signifie que le rapport des sons extrêmes que l'oreille peut percevoir sans douleur est de l'ordre de 10^5 à 10^6 .

L'échelle de bruit suivante indique le niveau sonore moyen de quelques situations courantes:

	dB (A)	
	140	Avion au décollage
	130	
	120	<i>Seuil de douleur</i>
Concert - discothèque	110	
	100	
Restaurant scolaire	90	<i>Seuil de danger</i>
	80	Ronflement / Automobile
Salle de classe	70	
	60	Fenêtre sur rue
	50	
	40	Salle de séjour
Chambre à coucher	30	
	20	Vent léger
	10	
	0	<i>Seuil d'audibilité</i>

3.3. Les seuils auditifs

Deux méthodes peuvent être utilisées pour mesurer le niveau minimum audible des sons purs, en fonction de leur fréquence (et généralement sur une seule oreille) :

- au casque, et on mesure alors la "*Minimum Audible Pressure*" ou *MAP*, la pression acoustique reçue étant contrôlée après étalonnage à l'aide d'un micro - sonde placé au niveau du tympan du sujet.
- en chambre anéchoïque (i.e. une salle aux parois totalement absorbantes) et en champ libre : on mesure alors le "*Minimum Audible Field*" ou *MAF*, et la pression est contrôlée ensuite en plaçant à l'emplacement de la tête du sujet, des microphones portés par une tête artificielle.

On observe que les seuils sont meilleurs (plus bas) dans ce second cas, car au casque le bruit de la circulation sanguine du sujet est amplifié et masque le son qu'on cherche à détecter.

Les variations du seuil d'audition avec la fréquence constituent donc la courbe des seuils auditifs, qui se confond avec ce que nous appelons plus loin l'isotonique 0 dB.

La forme générale de cette courbe s'explique par les caractéristiques des différentes parties de l'oreille :

1. l'ensemble pavillon et conduit auditif, en amont du tympan, est un canal de transmission acoustique qui possède comme on l'a vu une fréquence de résonance vers 3 kHz, et accroît globalement la sensibilité de 1 à 5 kHz,
2. les caractéristiques de transmission de l'oreille moyenne, et
3. la répartition des neurones dans la cochlée : un plus grand nombre de neurones sont associés aux fréquences moyennes (50 à 5000 Hz), et la sensibilité globale à ces fréquences, liée au nombre moyen de décharges nerveuses, en est améliorée.

3.4. Les courbes d'isophonie

Après les niveaux minimums audibles, on s'intéresse aux variations de la sensation d'intensité sonore, en recherchant d'abord l'ensemble des niveaux (physiques) qui donnent une même sensation d'intensité.

Pour cela on établit une unité : le **phone**, défini par référence au niveau physique d'un son de 1 kHz :

Un son de N phones est un son donnant une sensation d'égale intensité avec un son pur de fréquence 1 kHz et de niveau N dB SPL.

Ainsi, par définition, à 1000 Hz, l'échelle des phones se confond avec celle des dB.

L'isophonie est comme la sonie légèrement différente selon la méthode de mesure (casque ou champ libre).

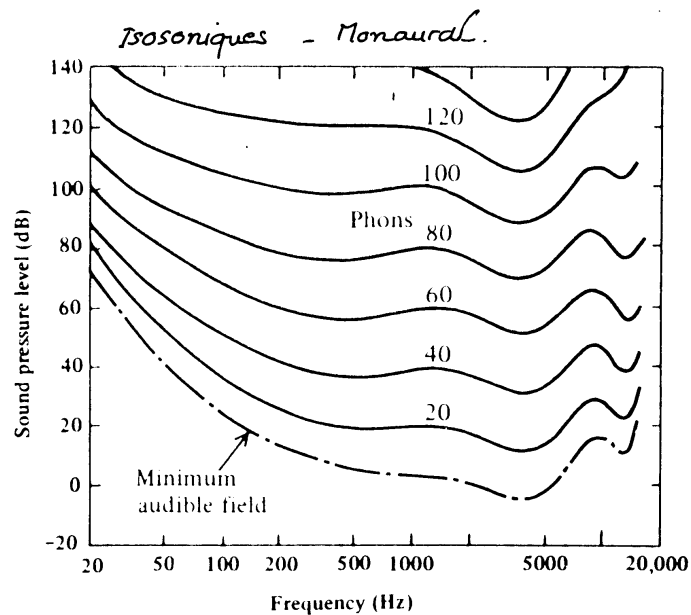


Figure : courbes d'isophonie monaurales (1 seule oreille), mesurées en champ libre.

On remarque que ces courbes sont plus concaves à niveau faible, plus plates à fort niveau : autrement dit, la relation entre le niveau en dB et la sensation provoquée dépend moins de la fréquence pour les niveaux élevés.

Une conséquence de ce fait est la réduction de la dynamique de l'oreille dans les niveaux graves et très aigus : l'augmentation du niveau sonore y fait passer plus rapidement d'un niveau « faible » à un niveau « fort », puis au seuil de douleur (limite supérieure du champ audible).

Ces isophonies ont diverses applications :

- dans la mesure des niveaux sonores, les filtres de pondération utilisés, vus au chapitre précédent, sont définis par des gains inversement proportionnels (i.e. des courbes renversées, en dB) aux niveaux des courbes isophonies à 40, 70 ou 100 dB.
- dans les amplificateurs HI-FI, on trouve souvent un filtre « contour » (« *loudness* ») qui permet de relever les graves (et un peu les aigus) pour corriger les intensités relatives des sons enregistrés à niveau élevé (concert) lors d'une écoute à bas niveau (salon).

3.5. L'audiogramme

On mesure les seuils d'audition de sons purs, le plus souvent au casque, à l'aide d'un audiomètre. L'audiomètre fait automatiquement le rapport du seuil d'audition mesuré au seuil moyen donné par la norme ; le résultat est exprimé en dB *HL* (*Hearing Level*) ; la courbe obtenue pour un sujet « normal » est donc une droite à 0 dB *HL*.

Les méthodes les plus courantes en audiométrie tonale sont :

- la présentation de sons pulsés, de niveaux fixes pour chacune des fréquences testées (octaves de 125 à 4000 ou 8000 Hz, et parfois 2 ou 3 fréquences intermédiaires). Le sujet indique chaque fois qu'il entend le son, en levant le doigt car la parole perturbe l'attention nécessaire.
- la présentation, aux mêmes fréquences que ci-dessus, de sons pulsés dont le niveau varie continûment, en fonction de la réponse du sujet qui appuie sur un bouton tant qu'il entend les sons. On change la fréquence testée lorsqu'un nombre suffisant d'inversions de sa réponse permet de déterminer par moyennage une valeur de seuil assez précise.
- la présentation de sons pulsés, dont la fréquence balaie lentement le champ audible, et dont le niveau, là encore, varie continûment, vers le haut ou vers le bas selon que le sujet appuie ou non sur un bouton indiquant qu'il entend. C'est l'audiogramme de Békésy. Les seuils sont obtenus comme la courbe des valeurs moyennes des niveaux sonores enregistrés.

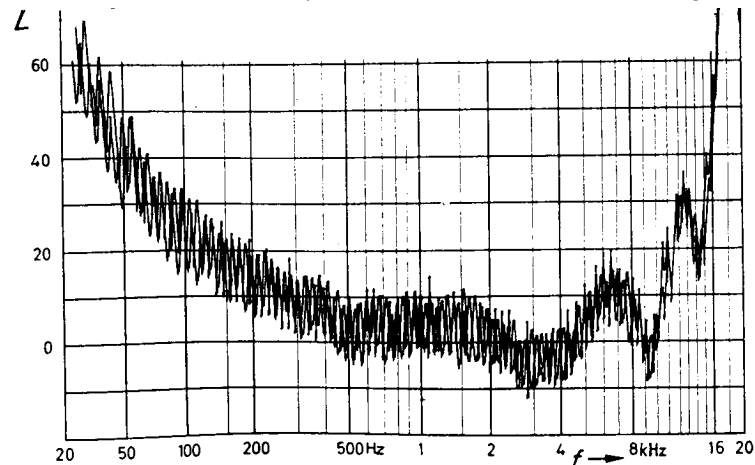


Figure : audiogramme de Békésy.

L'habitude dans le milieu médical veut qu'on exprime la perte auditive en représentant, aux diverses fréquences de mesure, le seuil auditif en dB *HL* mesuré sur une échelle négative. Les audiogrammes auront donc classiquement les formes ci-dessous :

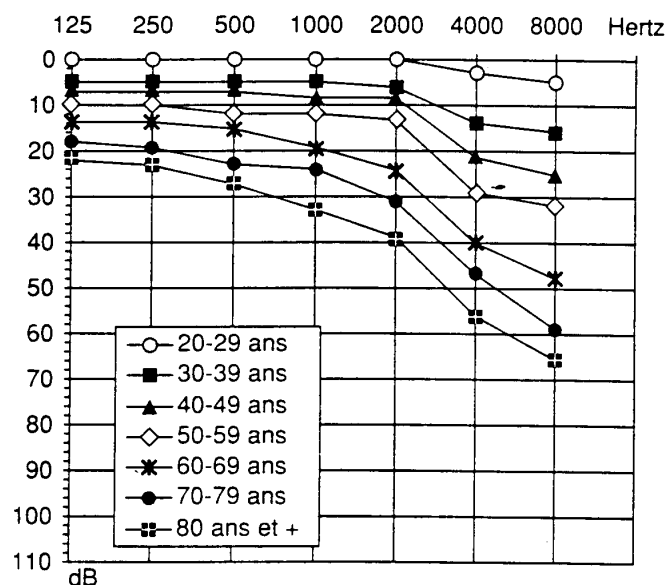


Figure : audiogrammes moyens en fonction de l'âge (presbycusis). La perte auditive (seuil en dB HL) est notée vers le bas ; on peut se représenter l'audiogramme comme l'atténuation d'un filtre que la déficience auditive superposerait aux caractéristiques de l'oreille normale.

Les audioprothésistes parlent parfois à leurs patients de pertes en « pourcentage d'audition ». Il s'agit d'un langage imagé, pour exprimer une mesure qui est en fait une moyenne des pertes en dB sur différentes fréquences, pondérées en fonction de leur importance relative dans la parole.

3.6. La Sonie

La sonie mesure le rapport (subjectif) entre deux sensations d'intensité sonore. Son unité est le **son**. Comme il s'agit d'une échelle relative, là encore sa définition nécessite une référence : par définition,

- à la fréquence de 1000 Hz, et au niveau de 40 dB SPL, on a une sensation unité : 1 sone. (ce qui revient encore à dire qu'on a 1 sone pour 40 phones).
- le niveau sonore pour lequel le son semble « deux fois plus fort » est le niveau de 2 sones, et ainsi de suite.

La mesure de l'échelle de sonie peut faire appel à différentes méthodes. Nous en citons deux, en notant certains de leurs avantages et inconvénients :

- méthode par ajustement : dans celle-ci, on présente au sujet alternativement un son de référence et un second son, dont il doit ajuster le volume jusqu'à ce que l'intensité du second son lui semble être dans un rapport donné (2 fois plus fort, ou 2 fois moins fort...) avec le son de référence. Cette méthode combine les difficultés de toutes les méthodes de jugement relatif, et des méthodes où le sujet ajuste le niveau du stimulus (voir annexe sur les méthodes d'étude des grandeurs psychoacoustiques).
- méthode par estimation d'amplitude : nous détaillons un peu cette méthode dont la démonstration est faite durant le cours, avant d'en exposer les résultats.

Test par estimation d'amplitude

Le test de sonie "par estimation d'amplitude" a été décrit par Stevens en 1971.

Il consiste à présenter en alternance :

- un son de référence, qui est un bruit large bande ;
- des sons de niveaux relatifs variables par rapport à la référence.

La séquence de présentation de chaque couple de sons peut être la suivante :

- son de référence durant 1 s.,
- silence durant 0,25 s.,
- son comparé 1 s.,
- silence 2,25 s. (durant lequel le sujet peut noter son estimation d'intensités relatives).

Au total une vingtaine de paires de sons peuvent être présentées pour construire une échelle de sonie. Ces paires sont présentées dans un ordre aléatoire. On sait que sans cela l'ordre de présentation aurait une influence sur les résultats (voir plus loin les facteurs de variation de la sonie).

Il est également utile de présenter au préalable les niveaux extrêmes (le plus fort et le plus faible) utilisés durant le test ; cela aide le sujet à ajuster son échelle de notations.

Les instructions données sont les suivantes :

si l'amplitude du son de référence se voit arbitrairement attribuer la valeur 100, quelle est l'amplitude relative du deuxième son de chaque paire ? Si le son est jugé 2 fois plus fort, on devra noter 200, s'il est 4 fois plus fort : 400, 2 fois moins fort : 50, etc.

Traitement des résultats

A partir des résultats bruts du test, on peut d'abord construire l'histogramme des moyennes : on calcule les moyennes des réponses données pour chaque paire de sons présentée, et on compare les moyennes pour les paires dans lesquelles les niveaux étaient identiques. On vérifie ainsi la cohérence des réponses.

On peut alors tracer la loi psychophysique : variation de la sonie (i.e. des réponses données) en fonction de l'intensité relative au son de référence, en dB. (voir l'annexe sur les méthodes d'étude des sensations auditives).

Dans le cas où le test a été réalisé avec des sons purs de fréquence fixe, les résultats suggèrent que la loi de variation de la sonie S en fonction de la pression acoustique p , pourrait être assez bien approchée par une loi puissance.

On recherche alors l'exposant a de celle-ci : si $S = k.p^a$, alors $\log S = \log k + a \log p$: on trace donc $\log S$ en fonction du niveau en dB, qui est proportionnel à $\log p$.

On trouve en mesurant la pente de la droite obtenue : $a = 0,6$; c'est la loi de Stevens.

Loi de Stevens

Cette loi expérimentale exprime le résultat du test précédent, réalisé par Stevens avec des sons purs de fréquence 1000 Hz ; elle peut s'énoncer ainsi :

→ **La sensation est comme la puissance 0,6 de l'excitation**

Ici la 'sensation' est la sonie en sones, et 'l'excitation' est la pression acoustique en Pa. (comme l'exposant est inférieur à 1, la courbe correspondante est convexe).

Il existe cependant une relation équivalente à cette loi, plus facile à retenir :

→ **La sonie double tous les 10 dB**

Par exemple : à 1 kHz et à 60 dB, on a une sonie de ... 4 sones. (il faut se souvenir que par définition, à 40 dB on a 1 sone).

Ainsi, sur une échelle de 100 dB, on a un rapport d'intensités subjectives de 2^{10} soit environ 1000 entre les niveaux extrêmes.

Pour démontrer l'équivalence, on montre que $p^{0.6}$ double tous les 10 dB : si les niveaux sonores N_2 et N_1 sont différents de 10 dB, cela s'écrit : $20 \log(p_2 / p_0) = 20 \log(p_1 / p_0) + 10$ donc :
 $\log(p_2 / p_1) = 0,5$ soit : $(p_2 / p_1)^{0,6} = 10^{0,3} = 2$.

Exercice: Détermination du coefficient k de la relation $S = k \cdot p^{0.6}$: à partir du choix arbitraire qui a été fait (on a 1 sone pour 40 phones), on a:
à 1 kHz et à 40 dB : $p = 100, p_0 = 2 \cdot 10^{-3}$ Pa.
Donc 1 (sone) = $k \cdot (2 \cdot 10^{-3})^{0.6}$ d'où : $k = 41,6$.

En réalité, une étude plus fine montre que l'approximation de la loi de Stevens n'est pas bonne pour les niveaux faibles (inférieurs à 30 dB environ), où la sonie réelle est inférieure à celle prédite ; la courbe de la sonie en fonction du niveau en dB s'infléchit alors vers le bas, comme le montre la figure :

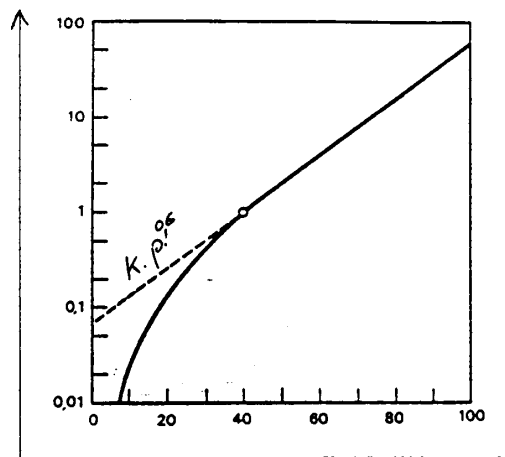


Figure : la loi de Stevens (sonie ou intensité perçue en fonction du niveau sonore en dB) à la fréquence de 1000 Hz

Une difficulté apparaît dans le test de l'échelle des sones comme dans de nombreux tests psychoacoustiques : notre perception est « déformée » par une tendance naturelle et inconsciente essayer de deviner les propriétés de la source, et d'analyser notre perception, plus que de simplement en rendre compte.

Malgré les précautions prises, le résultats de tests psychométriques dépendent toujours de la méthode et des conditions expérimentales. Ainsi, selon une autre méthode conçue pour éliminer le biais, (prise en compte seulement des réponses symétriques, une seule réponse par sujet), Warren a trouvé que la sonie doublait tous les 6 dB seulement.

3.7. Le Seuil Différentiel d'Intensité

Dans la vie courante, pour beaucoup de grandeurs subjectives, les variations de niveau sont tout aussi importantes, et même souvent plus significatives, que le niveau lui-même. On étudie donc avec soin les Différences juste perceptibles (DJP) ou "Just Noticeable Differences" (JND) ou encore Seuils Différentiels (SD) ou "Difference Limens" (DL).

C'est le cas de la sonie, dont les variations nous renseignent entre autres sur les mouvements de la source. On définit donc le Seuil Différentiel d'Intensité (SDI) comme la plus petite différence de niveau audible.

Diverses méthodes peuvent être employées pour l'étudier, par exemple :

- les sons modulés en amplitude (la modulation optimale est environ 4 Hz ; c'est la fréquence de modulation la plus importante dans l'enveloppe de la parole).

- les sons successifs séparés par des silences (sons de 500 ms et silences de 100 à 300 ms en général)
- la détection d'un incrément d'intensité dans un son continu
- la discrimination du son le plus fort dans une séquence fort / faible d'ordre aléatoire (méthode dite « deux alternatives à choix forcé » ou 2AFC, seuil à 75%, voir annexe)

Les résultats diffèrent :

- suivant les méthodes ; par exemple, la méthode de présentation successive de sons séparés par des silences donne de moins bons résultats que celle des sons modulés, car elle mobilise plus la mémoire.
- et suivant les sons utilisés, suivant par exemple qu'il s'agit de sons purs ou de bruits large bande.

Le SDI DI mesuré peut être exprimé en dB, relativement à l'intensité absolue du signal, par :

$$SDI_{dB} = 10 \log_{10} \frac{I + DI}{I}$$

Les résultats principaux sont les suivants :

- pour les bandes de bruit, c'est la **loi de Weber**: « $DI / I = \text{constante}$ » ; autrement dit, donc le SDI en dB est constant, légèrement inférieur à 1 dB.
- pour les sons purs, la loi de Weber ne s'applique plus : la fonction $DI = f(I)$ présente une pente de l'ordre de 0,9. Ainsi la discriminabilité des sons purs augmente avec le niveau. On a trouvé, pour un son de 100Hz:
 - $DI = 1,5$ dB à 20 dB SPL,
 - $DI = 0,7$ dB à 40 dB SPL,
 - $DI = 0,3$ dB à 80 dB SPL

L'ordre de grandeur de ces résultats est à retenir : **le SDI est de l'ordre de 1 dB.**

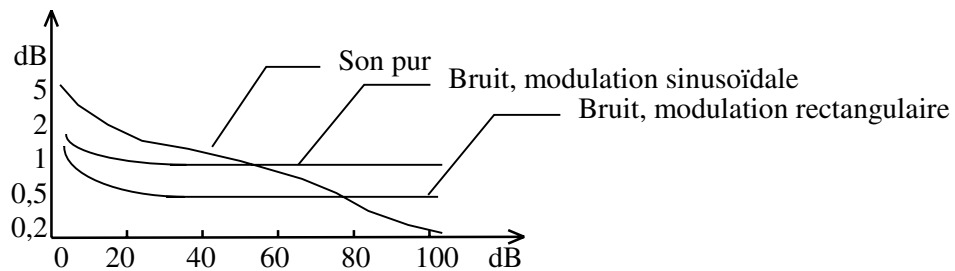


Figure : seuils différentiels d'intensité mesurés pour différentes conditions

Physiologiquement, il semble que l'explication de ces résultats fasse intervenir le changement du pattern d'activité nerveuse en fonction du niveau de stimulation.

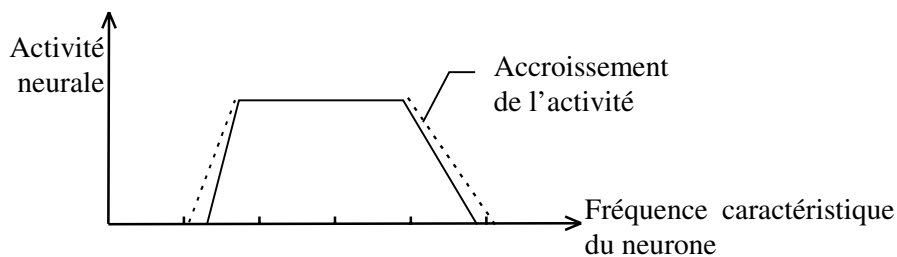


Figure : élargissement du pattern d'activité nerveuse en fonction de l'intensité sonore

Plus précisément, la dynamique du système auditif semble s'expliquer par une combinaison de trois facteurs:

- les variation du taux de décharge des fibres nerveuses situées en bordure du pattern d'excitation. (au centre de celui-ci, toutes les fibres sont excitées et saturées ; comme toute fibre nerveuse elles obéissent à la loi du « tout ou rien »).
- une augmentation du degré de verrouillage de phase¹ des messages nerveux au fur et à mesure qu'un son pur émerge et se différencie d'un bruit.
- l'exploitation, aux niveaux supérieurs du système nerveux, des propriétés statistiques des décharges des neurones.

3.8. Facteurs de variation de la sonie

Niveau du signal

Nous avons étudié plus haut la fonction de sonie et mis en évidence la loi de Stevens. En fait cette loi n'est valable, comme nous l'avons déjà mentionné, que pour des niveaux supérieurs à 30 dB. Pour les niveaux plus faibles, elle peut être corrigée en la loi suivante, proposée par Scharf (1978) :

$$S = k \cdot (p - p_r)^{0,6}$$

L'introduction de p_r dans la formule restitue l'inflexion de la courbe de sonie aux bas niveaux. A 1 kHz, la pression de correction p_r vaut 45 μ Pa (attention : il ne faut pas confondre celle-ci avec la pression acoustique p_o qui sert de référence dans la définition du niveau sonore en dB !) Pour d'autres fréquences, la fonction de sonie suit une loi similaire mais les paramètres k et p_r sont différents.

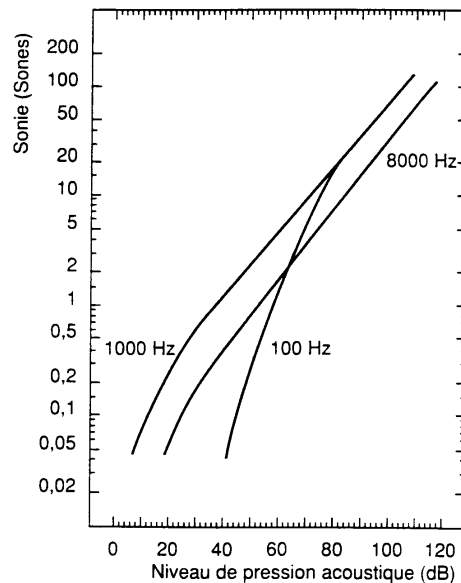


Figure : fonctions de sonie de sons purs à différentes fréquences

Nature du signal

Nous venons de voir les fonctions de sonie de sons purs. Dans le cas d'un *bruit large bande*, les résultats expérimentaux montrent que :

¹ Le verrouillage de phase est la tendance des fibres du nerf auditif, à décharger au rythme de l'excitation si celle-ci est périodique. Il permet ainsi un codage temporel de l'information sur les fréquences, qui est utile à la perception des hauteurs, comme on le voit plus loin.

- la sonie n'obéit plus à une loi de puissance
- la sonie augmente plus rapidement aux faibles niveaux
- la sonie d'un bruit est toujours supérieure à celle d'un son pur, à intensité physique égale.

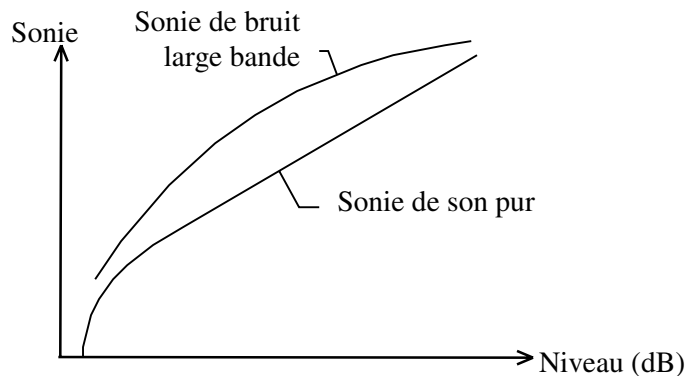


Figure : comparaison des fonctions de sonie de sons purs et de bruits large bande

L'étude de l'influence de la composition spectrale sur la sonie des bruits, abordée plus loin, débouchera sur la notion essentielle de bandes critiques.

Forme et mode de présentation des stimuli

Les résultats des tests de sonie varient également :

- suivant que les stimuli sont présentés en ordre aléatoire, croissant, ou décroissant ; dans ce dernier cas ils dépendront également du niveau initial, par un phénomène similaire à l'"adaptation de la sonie", décrite plus loin.

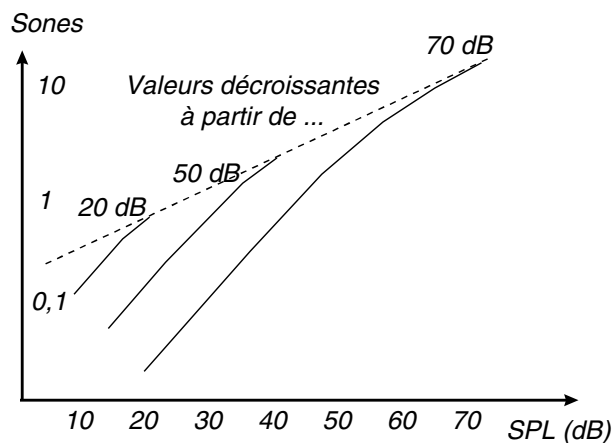


Figure : variations de la fonction de sonie avec l'ordre de présentation des stimuli ; exemple de présentations par valeurs décroissantes à partir de différents points de départ.

- suivant la forme des stimuli, par exemple s'il s'agit de sons pulsés ou de sons continus.

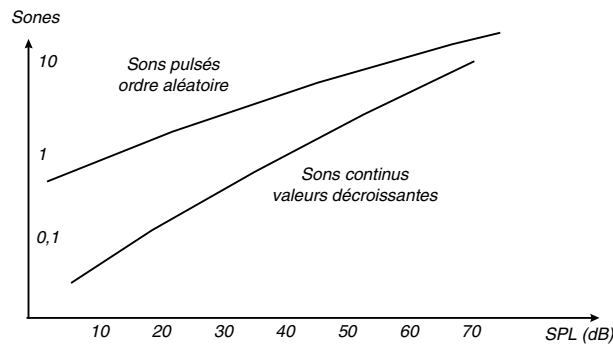


Figure : variations de la fonction de sonie avec la forme et le mode de présentation des stimuli : sons pulsés présentés en ordre aléatoire, ou sons continus présentés par valeurs décroissantes.

- les résultats varient encore avec la gamme des stimulations présentées (son amplitude) , la première stimulation présentée, les instructions données, la symétrie de la gamme des réponses demandées...

Durée du signal

L'expérience montre que la sonie augmente avec la durée du signal présenté, jusqu'à atteindre un plateau pour une durée d'environ 200 ms ; au-delà, on peut observer une légère décroissance.

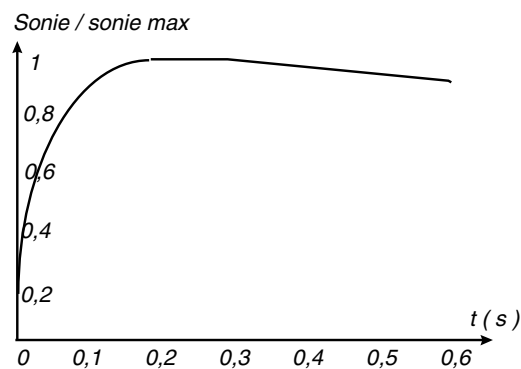


Figure : variation de la sonie avec la durée du stimulus

En fait, la **durée d'intégration**, au bout de laquelle la sonie maximale est atteinte, varie avec la fréquence du stimulus : elle est ainsi d'environ 375 ms à 250 Hz, mais descend lorsque la fréquence s'élève, jusqu'à valoir 150 ms à 8000 Hz.

Cette durée d'intégration peut être mise en regard de la variation de la *détectabilité* du stimulus en fonction de sa durée : pour des sons purs pulsés de 1000 Hz, on trouve que la détectabilité du signal est optimale pour des durées variant de 15 à 150 ms, mais inférieure pour des durées plus grandes ou plus petites.

L'étude de la variation de la sonie sur des durées longues (supérieures à 500 ms) met en évidence un autre phénomène, illustré par la petite décroissance au-delà du plateau sur la figure ci-dessus.

Cette étude peut se faire par deux méthodes :

- la méthode *SDLB* ou "*Simultaneous Dichotic Loudness Balance*", qui consiste à faire entendre en continu le son testé dans une oreille et, par intervalles (par exemple 3 mn), à faire entendre dans

l'autre oreille un son dont le sujet ajuste le niveau pour le rendre subjectivement égal à celui du son continu. On enregistre les variations du niveau donné au son de comparaison au fil du temps. Cette méthode, simple dans son principe, présente l'inconvénient de ne pas prendre en compte l'existence d'interactions interaurales.

- une méthode d'estimation directe (monaurale), utilisée par Scharf.

Les résultats sont les suivants :

- la sonie ne varie généralement pas, sauf aux faibles niveaux (inférieurs à 30 dB), où elle décroît légèrement avec le temps ; c'est ce qu'on appelle l'**adaptation simple de la sonie**.

- cette adaptation est maximale pour des sons purs H.F.

- elle reste controversée et inexpliquée ; elle est sujette à de grandes variations interindividuelles.

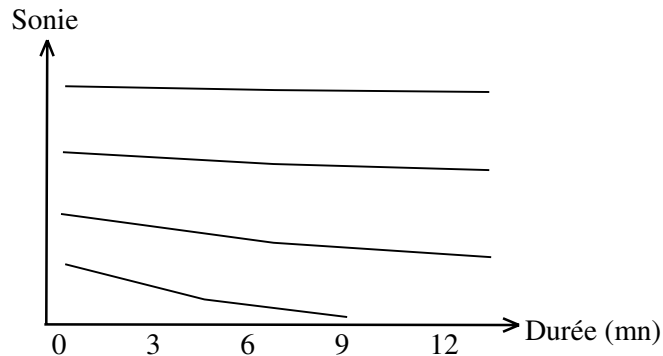


Figure : l'adaptation simple de la sonie

Les mécanismes de cette adaptation sont probablement liés à ceux mis en jeu par un autre phénomène, celui de la **fatigue auditive post-stimulatoire**.

La méthode de mesure de cette fatigue est la suivante :

- on réalise une première mesure du seuil auditif à la fréquence de test.

- on présente la stimulation "fatigante", généralement de fréquence proche de la fréquence de test, de niveau et de durée importants.

- immédiatement après la fin de la stimulation fatigante (ce qui constitue la difficulté pratique de la procédure), on mesure de nouveau le seuil à la fréquence de test. Le déplacement observé du seuil par rapport à la mesure initiale, s'appelle "*Temporary Threshold Shift*" ou *TTS*. Ce déplacement est influencé, en particulier, par :

- l'intensité, la durée, la fréquence du stimulus fatigant
- la fréquence du stimulus-test
- l'intervalle de temps entre le stimulus fatigant et le test

L'évolution du *TTS* en fonction de ce dernier intervalle, fait apparaître des courbes de récupération diphasiques, telles que le montre la figure ci-dessous.

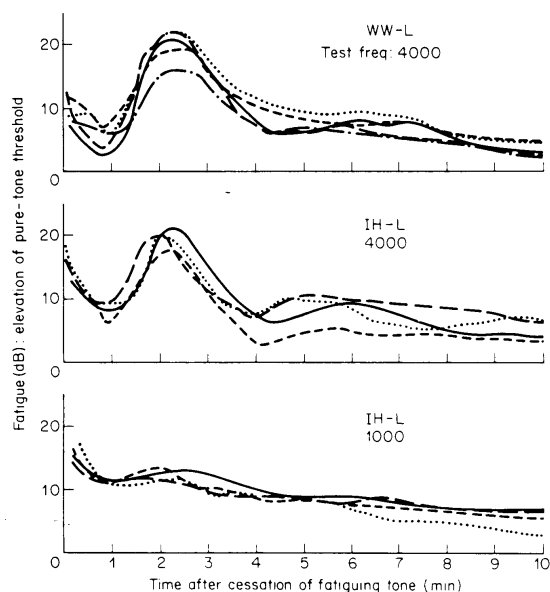


Figure : « temporary threshold shifts » ou élévation des seuils auditifs en fonction de la durée après cessation du son fatigant (en minutes). Son fatigant : 500 Hz, 120 dB, 3 mn; son test : 4 kHz (haut et milieu), 1 kHz (bas).

Ce processus diphasique suggère que deux mécanismes seraient impliqués :
 l'un, rapide, lié essentiellement à l'activité neuronale, et
 l'autre, plus lent, qui fait probablement intervenir des changements métaboliques (récupération des cellules ciliées...).

Effet de l'attention

L'expérience a montré que la focalisation de l'attention sur certains paramètres améliore les performances. Par exemple, la performance de détection d'une sinusoïde dans le bruit est meilleure si la fréquence de la sinusoïde est connue a priori : le sujet focalise alors son écoute sur la fréquence qu'il cherche à entendre.

Enfin l'expérience, la motivation et l'entraînement des sujets ont aussi une influence sur les résultats. Ces facteurs doivent être soigneusement pris en compte pour ne pas biaiser les résultats d'un test.

Effet de la composition spectrale

Zwicker a eu l'idée d'étudier les variations de la sonie avec la largeur de bande du signal. Pour cela, il a constitué des bruits de bande variable en additionnant des sons purs de fréquences proches, autour d'une fréquence de référence, en maintenant le niveau *SPL* constant.

Dans l'expérience telle qu'elle est reproduite durant le cours, on a :
 - un son d'origine à la fréquence 1000 Hz, de largeur de bande de 15% (930 - 1075 Hz),
 - puis une augmentation de la largeur de bande par 7 pas de 15%, l'amplitude décroissant de manière à maintenir l'énergie totale constante.

On observe les résultats suivants :

- au début, un changement dans la qualité du son, mais pas de changement de sonie
- à partir d'une certaine largeur de bande (160 Hz autour de 1 kHz), la sonie commence à augmenter.

Cette augmentation suit les courbes représentées ci-dessous :

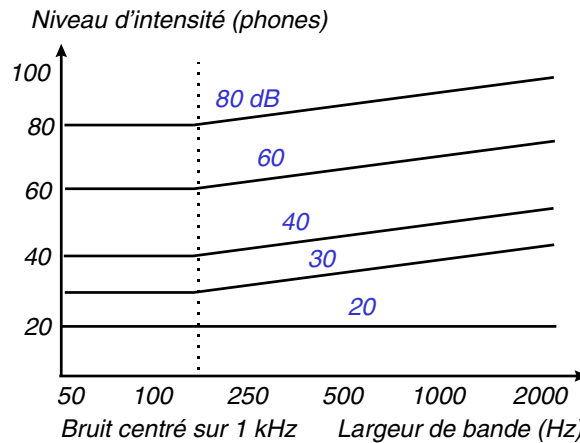


Figure : variations de la sonie avec la largeur de bande du signal, pour différents niveaux d'intensité sonore

4. Masquage et Bandes Critiques

4.1. Les Bandes Critiques

Mise en évidence

Les Bandes Critiques sont mises en évidence par le test précédent, dont les résultats suggèrent que :

- un signal de bande inférieure à 160 Hz excite toujours une zone identique de la cochlée, et donc un même nombre de fibres nerveuses.
- au-delà de cette largeur, d'autres zones de la cochlée, et de nouvelles fibres nerveuses, sont touchées par l'excitation, ce qui explique que la sensation d'intensité croît.

Ainsi il semble que la cochlée réagit comme un filtre, centré sur l'excitation, de largeur 160 Hz pour la fréquence centrale de 1000 Hz qui était testée.

La répétition de l'expérience à différentes fréquences, confirme cette idée que le système auditif se comporte, vis à vis de la sensation d'intensité, comme un jeu de filtres de fréquence centrale variable, de manière à être toujours centrée sur l'excitation.

L'étude des capacités de résolution fréquentielle de l'oreille, liées à la perception des hauteurs qui est étudiée plus loin, corrobore cette interprétation.

Variation de la largeur des BC en fonction de la fréquence centrale

Lorsque l'on fait varier la fréquence centrale du bruit dans le test précédent, il est intéressant d'observer comment varie la fréquence à partir de laquelle la sensation d'intensité commence à croître, c'est à dire la largeur de la bande critique à la fréquence considérée.

On trouve que cette largeur :

- est à peu près constante, et égale à 100 Hz, pour les fréquences inférieures à 500 Hz,
- puis augmente régulièrement en fonction de la fréquence centrale, et devient grossièrement proportionnelle à celle-ci, dans un rapport 0,2 :

$$\begin{aligned} Df &\gg 100 \text{ Hz} && \text{pour } f < 500 \text{ Hz,} \\ Df / f &\gg 0,2 && \text{pour } f > 500 \text{ Hz.} \end{aligned}$$

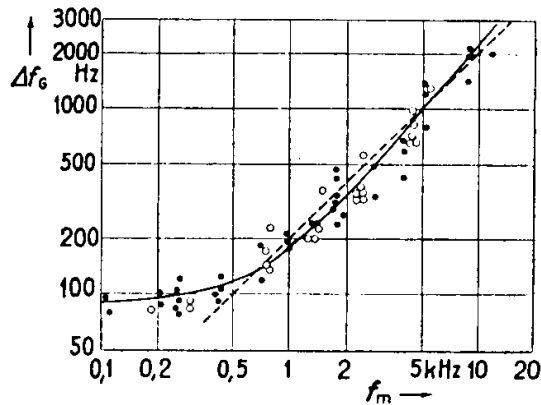


Figure : variation de la largeur des bandes critiques avec la fréquence centrale

On peut bien sûr trouver une fonction analytique qui rende plus finement compte des données expérimentales. Par exemple :

$$dF = 25 + 75(1 + 1,4F^2)^{0,69}$$

Une première échelle des fréquences : les Barks

Sur la base des résultats précédents, Zwicker a déduit une échelle de fréquences proportionnelle à la largeur des bandes critiques, de la manière suivante (E. Zwicker et R. Feldtkeller : Psychoacoustique, Ed. Masson, 1981) :

« L'ouïe peut former une bande critique en n'importe quel point de l'échelle des fréquences. En les rangeant arbitrairement l'une à côté de l'autre, on trouve dans la zone de fréquences de 20 Hz à 16 kHz, 24 bandes critiques ».

L'indice de la bande critique dans laquelle se trouve une fréquence donnée, transformé en une variable continue, devient une nouvelle échelle de fréquences : le **Bark** (du nom de H. Barkhausen, qui s'est intéressé parmi les premiers à la mesure de la sonie).

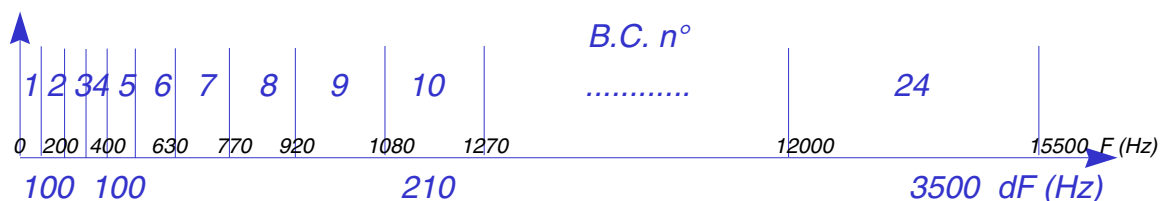


Figure : l'échelle de fréquences des Barks ; de haut en bas, les nombres représentent respectivement : les indices, les fréquences de séparation, et les largeurs des bandes critiques.

A la suite de ces travaux il est courant encore d'entendre dire que notre perception décompose le son selon 24 bandes critiques. Rien n'est plus trompeur. Il faut comprendre, comme le note Zwicker, que ces bandes critiques se « forment » en n'importe quel point, concrètement là où se trouve l'excitation sonore.

Quant au nombre de 24, il est comme le dit aussi Zwicker, tout à fait arbitraire, dans la mesure où il découle de la juxtaposition de bandes critiques dont les positions ne sont choisies qu'en fonction de leur largeur.

Cette juxtaposition suppose en outre l'hypothèse implicite, que les bandes critiques seraient comme autant de filtres passe-bande de formes parfaitement rectangulaires. Cette hypothèse simple ne correspond évidemment pas à la réalité, comme nous allons le voir.

4.2. Le Masquage

La notion de masquage est étroitement liée à la perception de l'intensité autant qu'à celle des hauteurs, qui sera étudiée plus loin ; nous avons vu avec la physiologie de la cochlée, grâce notamment aux cellules ciliées externes, qu'un « canal » auditif (i.e. une fibre du nerf auditif) ne répond qu'à une excitation située dans une zone fréquentielle précise. C'est ce qu'on appelle la sélectivité fréquentielle de l'oreille.

La mesure de cette sélectivité fréquentielle par des tests objectifs sur des animaux, donne les courbes d'accord neurales que nous avons présentées (voir 'physiologie'). Ces courbes d'accord permettent d'expliquer les phénomènes de masquage que nous décrivons ci-dessous.

Principe

Le principe du phénomène de masquage est qu'un son fort en masque (empêche de percevoir) un autre plus faible, surtout s'ils sont dans des fréquences proches.

La courbe délimitant la zone du plan « fréquence / amplitude » dans laquelle un autre son ne sera plus perçu, est appelée courbe de masquage ou « **pattern de masquage** » de ce masque.

Le pattern de masquage d'un son pur ou d'un son à bande étroite présente une pente raide du côté grave, et une pente plus faible du côté aigu. Le masquage est donc plus important dans cette zone, ce qu'on résume en disant « **les graves masquent les aigus** ».

Nous devons noter également que le masquage existe encore, sous certaines conditions, entre deux sons non simultanés mais séparés par un bref intervalle de temps. On parle alors de **masquage séquentiel** par opposition au cas courant du **masquage simultané**. Le masquage séquentiel est dit :

- *proactif*, lorsque le masque précède le son masqué. C'est le cas le plus important. Il met en évidence des mécanismes d'inhibition de l'excitabilité de la cochlée, par une excitation immédiatement antérieure.
- *rétroactif*, lorsque le masque précède le son masqué. Ce masquage qu'on qualifierait « d'anticausal » en traitement du signal, ne peut s'expliquer que par l'interférence des intégrations temporelles des deux signaux concurrents.

Expérience de base et patterns de masquage

L'expérience élémentaire d'étude du masquage consiste à mesurer le pattern de masquage d'un son donné, de la manière suivante :

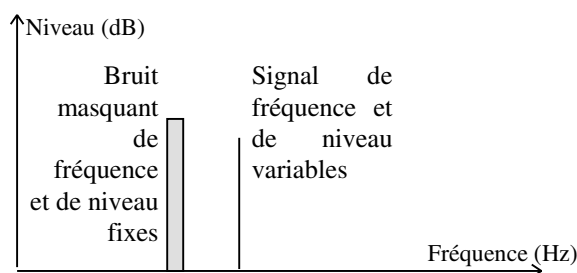


Figure : mesure de courbe (ou pattern) de masquage d'un bruit à bande étroite.

Si le son variable est un son pur, on évitera que le son masque soit également un son pur, car les effets de battements lorsque les fréquences sont proches ou multiples l'une de l'autre, perturbent alors la mesure.

Voici les patterns de masquage de quelques sons simples et complexes :

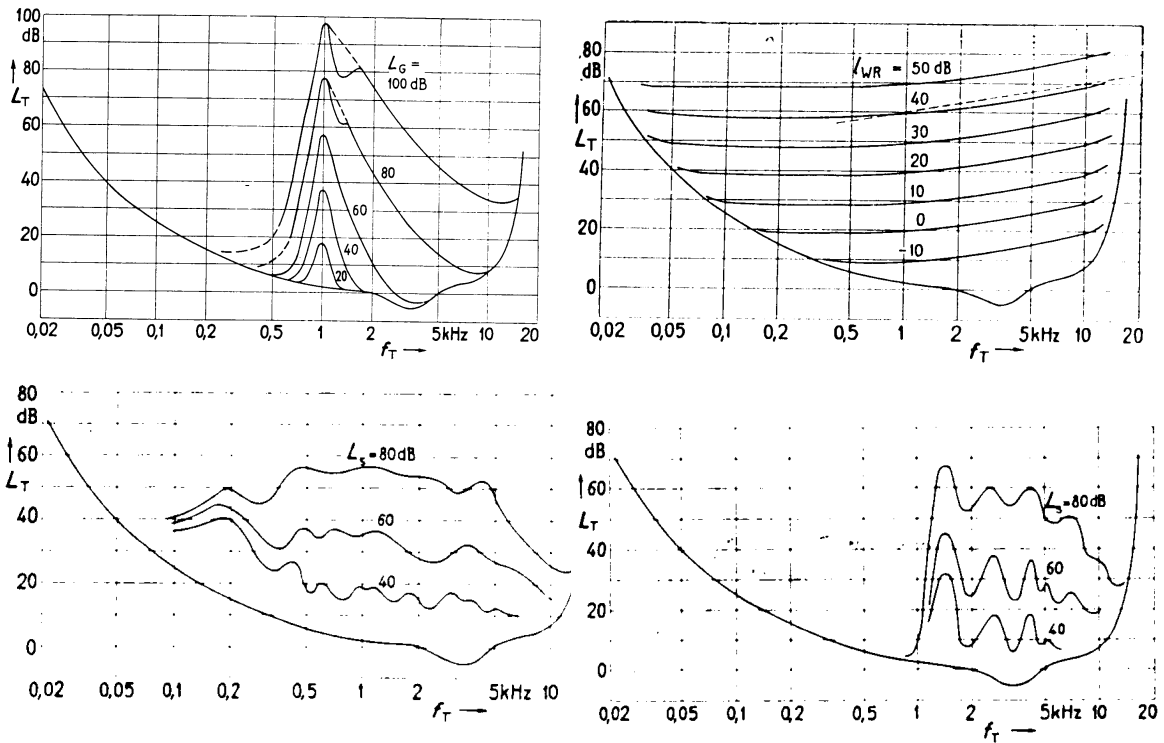


Figure : courbes ou patterns de masquage de différents sons à divers niveaux, de haut en bas et de gauche à droite : bruits à bande étroite, bruits blancs, sons graves de violons, sons aigus de violons.

Interprétation physiologique : le pattern d'excitation

Comme le laissent deviner les courbes d'accord neurales, le son fort ou « masque » produit une réponse plus ou moins forte dans les divers canaux auditifs voisins de sa ou ses fréquences propres ; l'enveloppe de cette réponse constitue son « pattern d'excitation », qu'on peut assimiler à l'enveloppe des vibrations de la membrane basilaire. Sa forme s'explique en effet (en partie) par la sélectivité en fréquence des déformations de la membrane basilaire, renforcée par les mécanismes actifs où interviennent les cellules ciliées externes.

Si elle est assez forte, l'excitation du masque recouvre celle que produirait le son faible, qui se retrouve « masqué ».

Cette interprétation implique que le pattern de masquage coïncide approximativement avec le pattern d'excitation du masque. Cette hypothèse a été corroborée par des tests et des observations physiologiques.

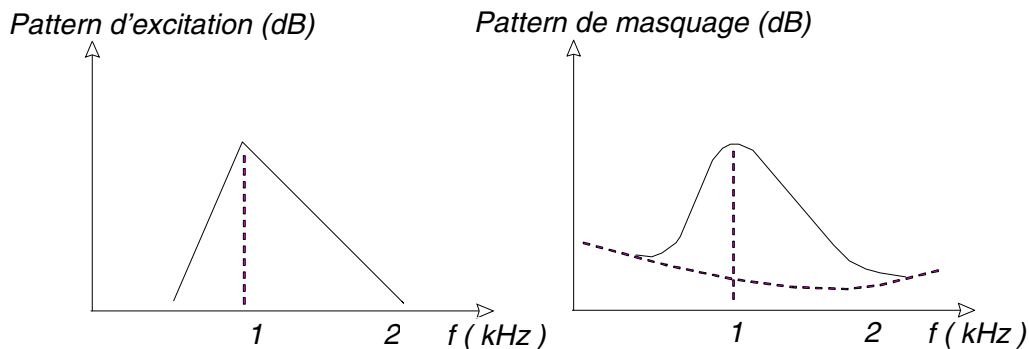


Figure : pattern d'excitation et pattern de masquage

Méthode de calcul de la sonie

Le modèle des bandes critiques de Zwicker et la notion de masquage, permettent de déduire une méthode d'évaluation de la sonie produite par un son donné, de la manière suivante :

1. on estime l'excitation dans chacune des 24 bandes critiques², ce qui peut être fait par la loi de Stevens (pour des sons purs) ou une loi approchante (pour des sons complexes ou des bruits). Notons que la répartition de ces bandes critiques, selon l'échelle des Barks, est grossièrement similaire à l'échelle de tiers d'octave utilisée dans la mesure des niveaux sonores.
2. on corrige cette estimation pour tenir compte de l'effet de masque, qui peut être négligeable, partiel ou total, suivant la sonie relative des bandes critiques adjacentes.
3. on somme les excitations des 24 bandes critiques pour obtenir la sonie totale.

Bien que ne reproduisant pas fidèlement les processus de l'oreille, cette méthode d'estimation de la sonie donne généralement d'assez bons résultats (en accord avec les résultats expérimentaux).

Notons qu'une évaluation du masquage entre bandes adjacentes, selon le même principe, est utilisée dans tous les procédés de codage audio courants (MPEG, ATRAC du minidisc, MP3...) dans le but de compresser l'information sonore en éliminant ce qui est inaudible pour l'oreille humaine. On trouvera un exemple sur Internet à l'adresse : <http://www.dolby.com/tech/parametr.html>

4.3. Les filtres auditifs

Le test de Fletcher et les Courbes d'Accord Psychophysiques

La mesure des « courbes d'accord psychophysiques » (CAP, ou *PTC* pour *Psychophysical Tuning Curves*) propose par un test subjectif de masquage une alternative non invasive à la mesure des courbes d'accord neurales.

Dans ce test, le signal est fixé à la fréquence étudiée et à un niveau habituellement bas, par exemple 10 dB SL. On mesure alors le niveau nécessaire du masque, suivant sa position spectrale, pour qu'il masque effectivement le signal.

Afin de limiter la présence de battements qui constituent un indice de détection parasite, si le signal est un son pur on utilise pour le masque, plutôt qu'un son pur, un bruit à bande étroite :

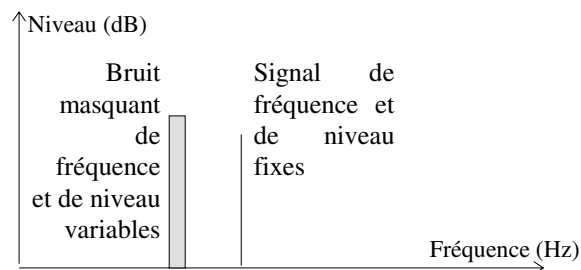


Figure : principe de la mesure des courbes d'accord psychophysiques

Les courbes obtenues ont l'allure ci-dessous :

² Le nombre et la répartition des bandes critiques par juxtaposition sont, rappelons-le, très arbitraires.

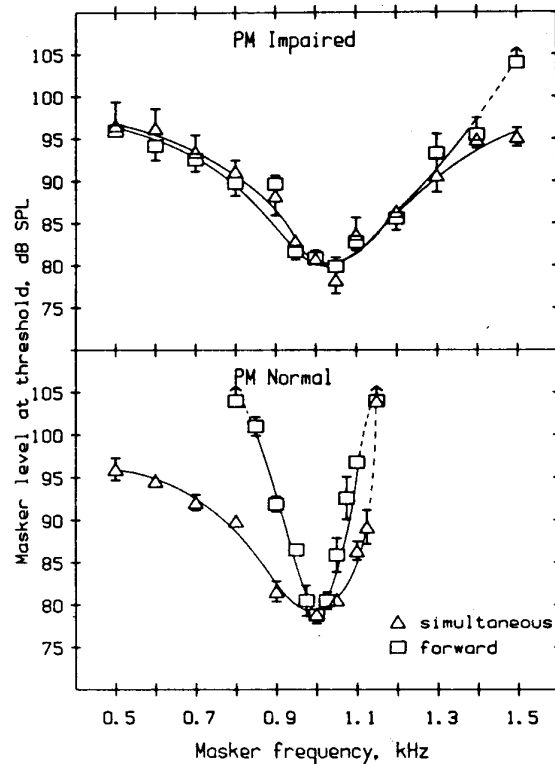


Figure : exemple de courbes d'accord psychophysiques, pour une oreille normale et une oreille malentendante, mesurées en masquage simultané et en masquage proactif.

Le modèle du spectre de puissance et les filtres auditifs

L'expérience de Fletcher³ (1940) a mené à élaborer un modèle du masquage (dit « du spectre de puissance »), basé sur les hypothèses suivantes :

- le système auditif périphérique peut être considéré comme un ensemble de filtres linéaires recouvrants, les « filtres auditifs »,
- la détection d'un signal dans un bruit fait appel à un (seulement) de ces filtres, celui de rapport signal sur bruit (RSB) maximum,
- seules les composantes du bruit qui passent dans ce filtre ont une incidence sur le masquage,
- le seuil de détection de la sinusoïde est ~ égal au niveau du bruit dans le filtre.

Ce modèle permet d'estimer le seuil de détection d'un son pur dans le bruit, par la formule :

$$P_s = K \int_0^{+\infty} W(f)N(f)df$$

où P_s est la puissance minimale du signal pour la détection, K est le rapport des puissances du signal et du masque à la sortie du filtre au seuil de détection, $W(f)$ est une fonction de pondération qui donne la forme du filtre auditif, et $N(f)$ est le spectre de puissance à long terme du masque.

On sait aujourd'hui qu'aucune des hypothèses n'est rigoureusement juste, mais malgré ses imperfections le modèle des filtres auditifs reste largement utile et utilisé.

³ Celle-ci consistait à mesurer le seuil de détection d'une sinusoïde dans un bruit passe-bande centré sur la fréquence de la sinusoïde, et dont la largeur de bande augmentait avec une densité spectrale de puissance constante.

Sous ces hypothèses, les CAP sont reliées de manière très directe aux filtres auditifs :

Si l'on suppose qu'un seul filtre auditif est utilisé, que ce filtre est linéaire, et qu'au seuil de perception du signal le masque produit une sortie constante sur ce filtre, alors la forme du filtre auditif est obtenue en inversant la courbe d'accord psychophysique.

La connaissance de la forme des filtres auditifs en chaque point de la membrane basilaire, permet d'estimer la forme du pattern d'excitation en réponse à un son donné, par juxtaposition des excitations produites dans chaque filtre.

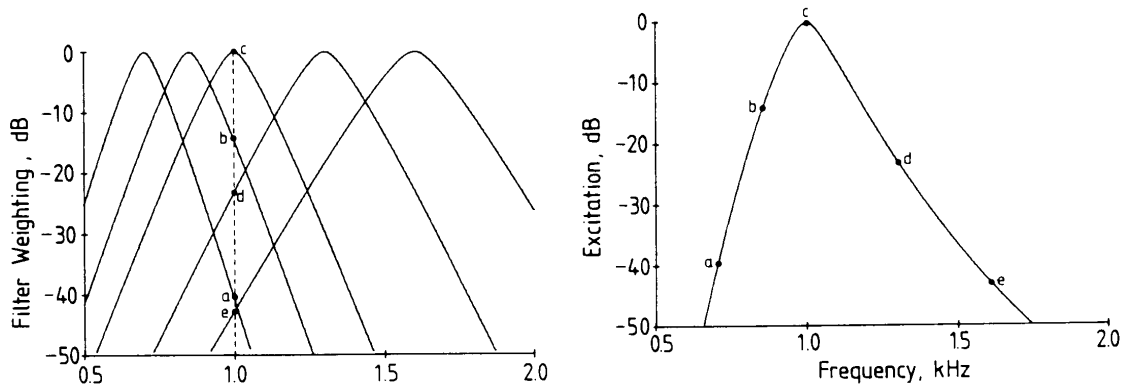


Figure : illustration du calcul du pattern d'excitation en fonction de la forme des filtres auditifs

Les imperfections du modèle se traduisent cependant par différentes limitations et biais possibles de ce mode de mesure :

Limitation 1 : le filtre de plus fort rapport RSB n'est pas forcément celui centré sur la fréquence de la sinusoïde. Le fait de percevoir le signal dans un filtre adjacent où le RSB est meilleur, s'appelle « écoute hors fréquence » (EHF), ou « off-frequency listening ». L'EHF fait que les courbes obtenues sont plus pointues que si un seul filtre était impliqué.

On contre ce biais en limitant l'EHF par un bruit coupe-bande centré sur la fréquence du signal :

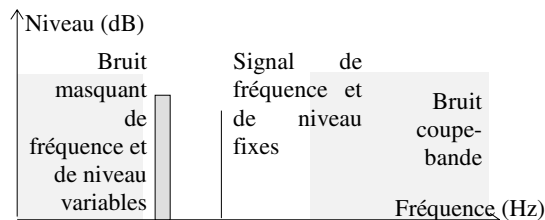


Figure : mesure des courbes d'accord psychophysiques avec bruit coupe-bande limitatif

Limitation 2 : les seuils réels sont abaissés par la possibilité de détection de produits de distorsion entre le masque et le signal.

Ce problème est réduit par l'utilisation d'un masque à bande étroite, et le bruit coupe-bande centré sur la fréquence du signal a également pour effet de masquer les éventuels produits de distorsion.

Limitation 3 : la non-linéarité des filtres réels, et leur dissymétrie, font que la mesure est déformée. Cependant pour des bruits de niveau modéré, la forme du filtre est à peu près symétrique sur une échelle linéaire et la déformation peut être négligée.

Un modèle des filtres auditifs : les fonctions « roex »

Sur la base de résultats de mesure, en 1982, Patterson et coll. ont proposé pour rendre compte de la forme des filtres auditifs une fonction appelée « rounded exponential » ou *roex*. Sous les hypothèses simplificatrices que les filtres auditifs sont symétriques sur une échelle logarithmique, et que leur étalement loin de la fréquence centrale f_c est négligé, cette fonction s'écrit :

$$W(g) = (1 + pg)e^{-pg},$$

où g est l'écart relatif de la fréquence f de calcul à la fréquence f_c du filtre, soit :
 $g = |f - f_c| / f_c$, et p est un paramètre déterminant la largeur du filtre⁴.

Une nouvelle échelle de fréquences : les ERB

Les filtres auditifs n'étant pas rectangulaires, on pourrait définir leur largeur de bande comme la bande passante à -3 dB ; Patterson et coll. ont préféré considérer la largeur du filtre rectangulaire ayant la même surface. On obtient la « bande rectangulaire équivalente » *BRE* du filtre auditif, qui peut être estimée par (Moore et Glasberg, 1983) :

$$BRE = 6,23 f_c^2 + 93,39 f_c + 28,52$$

Cette nouvelle échelle fréquentielle est donc une échelle concurrente de celle des Barks. Elles sont comparées sur la figure ci-dessous :

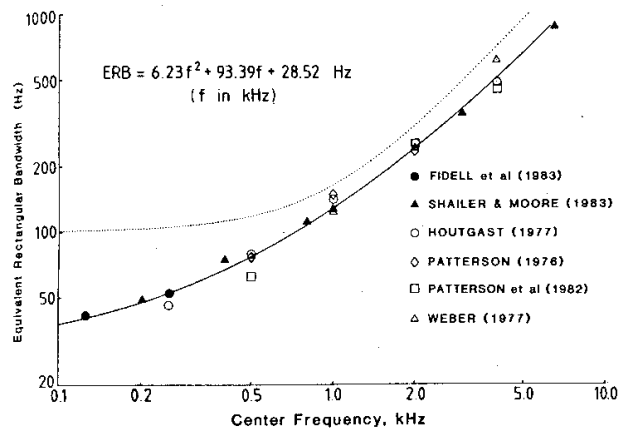


Figure : comparaison des largeurs de bandes critiques, selon les échelles ERB (en trait plein) et Bark (en pointillés).

5. Perception d'intensité et perte auditive

On peut distinguer trois grands types de surdité, ou plutôt de pertes auditives ou hypoacusies (la surdité étant une perte totale de l'audition) :

- les pertes auditives de conduction ou de **transmission**, sont celles qui sont liées à une défaillance des mécanismes de transmission mécanique de la vibration, essentiellement dans l'oreille moyenne.

Par exemple, dans l'otospongiose, une calcification anormale bloque peu à peu la vibration de l'étrier. Lorsque ces pertes auditives ne peuvent être corrigées par une intervention chirurgicale, une réhabilitation peut souvent être apportée par une aide auditive à vibreur osseux type *B.A.H.A.*

- les pertes auditives neurosensorielles ou de **perception** sont celles qui affectent la partie transduction et transmission du signal nerveux vers le cerveau. Ces pertes sont dites :
 - *endocochléaires*, lorsqu'elles ont leur origine dans la cochlée, ou
 - *rétracochléaires*, lorsqu'elles ont leur origine dans le nerf auditif.
- les pertes auditives **centrales**, enfin, sont celles liées à un dysfonctionnement du système nerveux central.

⁴ qui vaut en pratique : $p = 4 f_c / BRE$, en fonction de la bande rectangulaire équivalente *BRE* du filtre auditif

Les pertes auditives les plus courantes, telles que la presbycusie, associent généralement une composante « transmission » et une composante « perception ».

Le degré de perte auditive est évalué et classé suivant une norme du Bureau International d'Audio Phonologie, en fonction de la perte auditive moyenne aux sons purs, (*Pure Tone Average Hearing Loss* ou *PTA*) en moyennant les seuils en dB *HL* aux fréquences 500, 1000 et 2000 Hz :

PTA	Déficiance auditive...
0 à 20 (ou 25) dB	aucune (audition normale)
20 (ou 25) à 40 dB	légère
40 à 55 dB	moyenne / modérée
55 à 70 dB	moyennement / modérément sévère
70 à 90 dB	sévère
Plus de 90 dB	profonde

Tableau : degrés de perte auditive

Alors que les pertes auditives de transmission se traduisent par une pure atténuation, les pertes auditives de perception ont différentes conséquences sur la perception de l'intensité sonore :

5.1. Le recrutement ou rattrapage de la sonie

Dans la plupart des pertes auditives de perception, alors que les seuils de perception sont relevés, les seuils de douleur restent dans la même zone du plan dynamique, autour de 100 dB.

Par conséquent la sonie croît donc plus vite entre eux. Le niveau sonore passe rapidement de « pas assez fort » à « trop fort ».

Cela explique la plainte fréquente de nos grands-parents malentendants : « ne criez pas, j'entends assez fort, mais je ne comprends pas » : alors qu'ils n'entendent pas les sons faibles, les autres restent audibles mais inintelligibles, pour des raisons liées à l'altération des filtres auditifs (voir ci-dessous).

Puisque la même variation de sonie correspond chez les malentendants à une gamme dynamique de sons plus faible, il est logique de supposer que les seuils différentiels d'intensité (SDI) sont plus faibles ou « meilleurs » que ceux des normo-entendants. C'est sur cette hypothèse que se fondent certains tests cliniques de détection du recrutement. Malheureusement ces tests semblent biaisés du fait que l'écart statistique moyen des réponses est également plus important chez les malentendants, et donc que la mesure des SDI est moins précise.

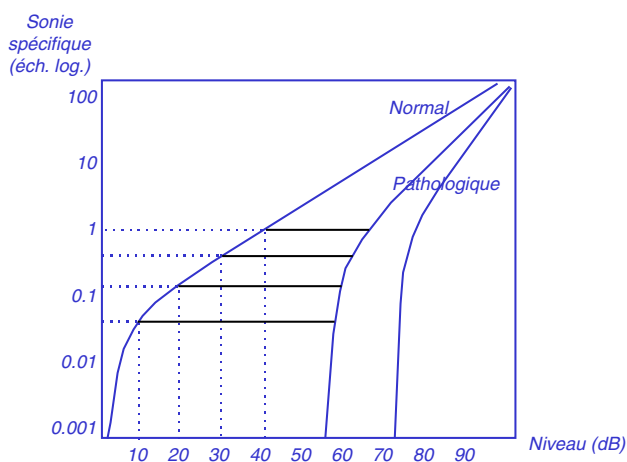


Figure : courbes de sonie normale et pathologiques, avec recrutement. Les traits horizontaux représentent le gain nécessaire pour restituer au malentendant la même sensation d'intensité que celle du normo-entendant ; on voit que ce gain varie avec le niveau d'entrée, d'où la nécessité d'une amplification avec compression dans les aides auditives.

5.2. L'adaptation pathologique

L'adaptation pathologique, observée chez certains malentendants, est une chute rapide de la perception d'un stimulus qui se prolonge dans le temps. Elle peut se manifester même alors que pour ce sujet et pour le son considéré le seuil auditif était quasiment normal.

L'adaptation pathologique est mesurée comme l'adaptation simple.

Elle apparaît à tous les niveaux sonores, et pas seulement aux niveaux faibles comme c'était le cas de l'adaptation simple.

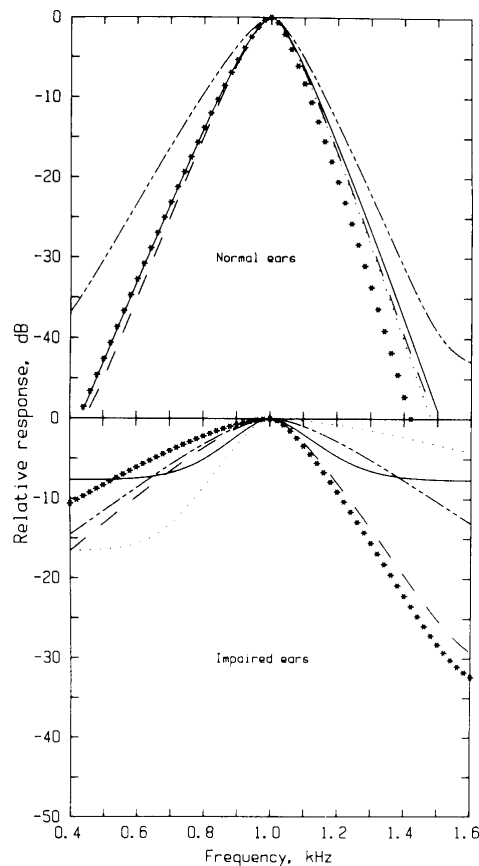
Elle s'explique par une plus grande fatigabilité des mécanismes cochléaires ou du nerf auditif chez le malentendant. Cette fatigabilité peut être liée à son tour à une sollicitation plus soutenue de certains mécanismes pour compenser une déficience.

Chez certains sujets, on peut observer un écart jusqu'à 20 ou 30 dB entre le seuil de perception d'un son pur et le niveau où il reste audible en continu.

5.3. L'altération des filtres auditifs

Chez les malentendants, les courbes d'accord psychophysiques (et donc les filtres auditifs) sont déformées :

- essentiellement, les CAP sont plus larges que chez les normo-entendants.
- elles changent parfois de forme ; elles peuvent prendre une forme en « W », avec deux minima, au lieu de la forme normale en « V ».
- elles ont parfois un sommet décalé par rapport à la fréquence du signal (cela est rare cependant, et correspond à des pertes auditives qui varient brutalement avec la fréquence).



Auditory filter shapes at a centre frequency of 1 kHz for the normal and impaired ears.

Figure : élargissement des filtres auditifs avec la perte auditive

Les différents types de pertes auditives de perception semblent s'expliquer en fonction de l'atteinte des cellules ciliées :

- si les CCE seulement sont altérées : la perte ou la déficience des mécanismes actifs (amplification à bas niveaux et affinement de la sélectivité) entraînent une élévation des seuils et un élargissement des filtres auditifs.
- si les CCE et les CCI sont altérées : les filtres auditifs sont élargis comme précédemment, les seuils absolus sont plus élevés.
- si les CCI seulement sont altérées (cas rare) : les seuils absolus très élevés, il n'y a qu'une faible perte de sélectivité fréquentielle.

Les conséquences sur le masquage d'un élargissement des filtres auditifs se font sentir différemment suivant les situations :

1) lorsque le spectre du masque recouvre celui du signal, les normo-entendants sont également gênés, et les performances des malentendants sont seulement un peu moins bonnes.

2) lorsque les spectres du masque et du signal utile sont différents, les effets du masquage sont bien plus importants et handicapants chez les malentendants ; deux exemples l'illustrent :

- le ventilateur (de spectre grave) qui masque la sirène (aiguë) : différence de spectre moyen, effet de masquage simultané.
- le masquage d'un locuteur par un autre : différence de spectre à court terme, effet de masquage séquentiel ; les malentendants sont moins à même que les normo-entendants de mettre à profit les « creux » temporels et fréquentiels du masque pour capter des informations sur le signal.

Les effets du masquage interne entre les différentes composantes d'un son de parole (ici, la voyelle / i /), sont illustrés par la figure ci-dessous, où l'on peut comparer l'information reçue par un normo-entendant et par un malentendant :

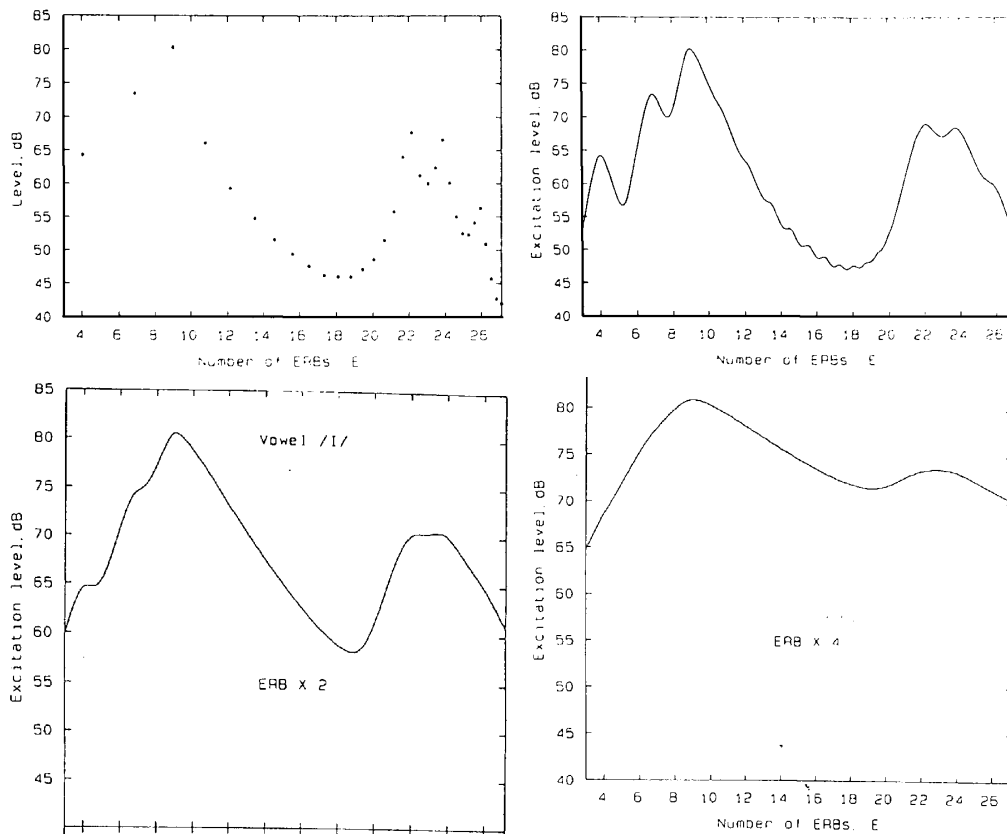


Figure : Spectrogramme et patterns d'excitation (calculés par un modèle) de la voyelle / i / ; en haut : spectrogramme de la voyelle (synthétisée par addition d'harmoniques), et pattern d'excitation produit chez un normo-entendant ; en bas : patterns d'excitation produits chez un malentendant si l'on suppose un élargissement des filtres auditifs d'un facteur 2 (perte modérée) ou 4 (perte sévère). L'échelle des fréquences est une échelle de nombre d'ERBs.

Dans le cas d'un normo-entendant, on voit que les 3 premiers formants (les pics de l'enveloppe fréquentielle) et même les premières harmoniques sont bien discriminés.

Lorsque les filtres auditifs s'élargissent, en revanche, la résolution des premières harmoniques disparaît, et la localisation des formants devient plus imprécise, allant jusqu'à confondre les formants 2 et 3 dans un seul pic assez plat. C'est donc d'une information réduite que le malentendant dispose pour comprendre la parole.

Perception de la hauteur des sons

1. Définitions

Zwicker distingue quatre "hauteurs", suivant qu'on s'intéresse à une grandeur physique ou perceptive, et qu'on adopte une échelle linéaire ou logarithmique. Elles sont regroupées dans le tableau suivant :

	Echelle linéaire	Echelle logarithmique
grandeurs physiques	fréquence (en Hz)	hauteur harmonique
grandeurs subjectives	hauteur / tonie	hauteur mélodique

Cependant le terme hauteur, que nous réservons dans ce cours à des grandeurs **subjectives**, (indépendamment de l'échelle, linéaire ou logarithmique) recouvre différentes choses :

- la hauteur brute :

C'est la position du son sur une échelle subjective "grave - aigu", indépendante de tout sens musical. Elle est reliée à la concentration d'énergie sur l'axe des fréquences :

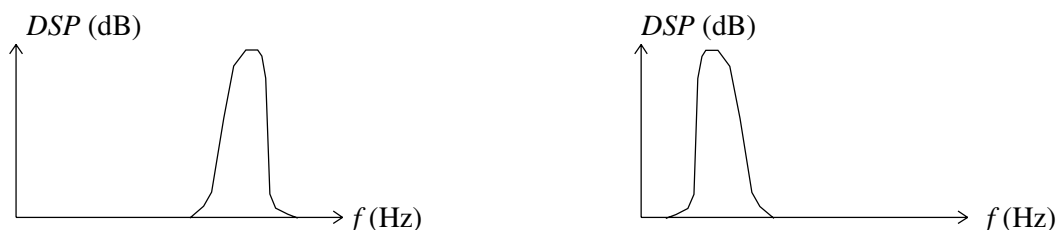


Figure 1 : densités spectrales de puissance de deux bruits respectivement aigu (à gauche) et grave (à droite)

- la hauteur tonale :

Celle-ci exprime la situation du son dans une organisation en intervalles. Une suite d'intervalles formera une mélodie.



Figure 2 : représentation musicale classique d'une succession de hauteurs tonales

Cette hauteur tonale ne concerne que les sons purs ou périodiques, voire *certain*s sons complexes non périodiques mais qui s'en rapprochent suffisamment du point de vue perceptif.

- les hauteurs spectrales :

Ce sont les différentes hauteurs qu'on peut distinguer dans un son complexe, si l'on pratique une écoute « analytique », i.e. qui tend à séparer les composantes du son.

Un ensemble de hauteurs spectrales perçues simultanément peut constituer un "accord".

La perception peut en être analytique (perception distincte des sons simultanés) ou au contraire synthétique (fusion perceptive des différents sons en un seul percept global).

- la hauteur virtuelle :

C'est la hauteur perçue dans un son complexe que l'oreille réduit sur une échelle monodimensionnelle (il s'agit donc forcément d'une écoute synthétique). Nous expliquons plus loin comment quelques idées « intuitives » sont fausses :

- cette hauteur virtuelle ne correspond pas forcément à une raie du spectre !
- l'oreille a la faculté d'entendre une ou des hauteurs, même dans des sons non harmoniques.
- il peut y avoir plusieurs hauteurs virtuelles dans un son (son de cloche).

2. Hauteur tonale des sons purs

2.1. L'échelle de tonie : le Mel

On parle dans ce paragraphe de hauteur tonale ; rappelons-le, c'est la hauteur du son au sens des notes et des intervalles de musique. Cette hauteur tonale est définie de manière univoque pour les sons purs. Les expériences suivantes sont donc d'abord réalisées avec des sons purs.

On définit l'échelle de hauteur tonale, ou tonie, de la manière suivante : par rapport à un son de référence (1 kHz par exemple), on dit que la tonie est double si un autre son est perçu comme deux fois plus aigu, etc.

L'unité de la hauteur tonale, ou tonie, est le Mel.

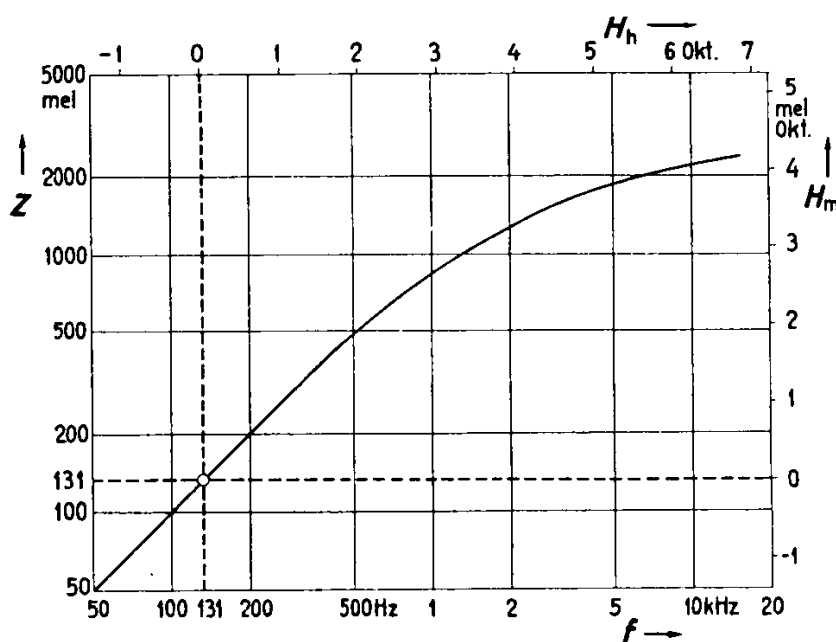


Figure 3 : fonction de variation de la tonie avec la fréquence d'un son pur

On constate expérimentalement que :

- dans les fréquences basses, la tonie est proportionnelle à la fréquence. Dans cette partie du spectre on identifie donc les échelles (c'est un choix arbitraire !) : ainsi, de 0 à ~500 Hz, on a : 1 Mel = 1 Hz.
- ensuite, la tonie croît moins vite que la fréquence, comme le montre la Figure 3.

Ainsi, tandis que la hauteur harmonique peut atteindre 7 octaves, la hauteur mélodique (i.e. la hauteur tonale, exprimée sur une échelle logarithmique) est limitée à environ 4,3 octaves. La Figure 4 permet de comparer ces deux grandeurs avec différentes grandeurs physiques ou physiologiques.

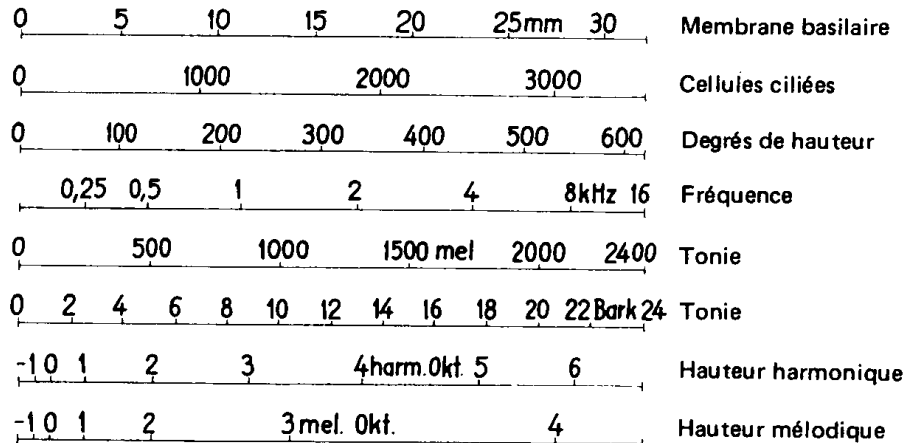


Figure (35,1) — Les échelles naturelles de la membrane basilaire.

Figure 4 : comparaison de différentes échelles de fréquence rapportées à la tonotopie de la membrane basilaire, par Zwicker.

Notons qu'en dehors de l'intervalle [60, 5000] Hz nous ne sommes plus vraiment capables d'associer à un son une hauteur tonale bien définie.

Relation entre le Mel et le Bark

Curieusement, Zwicker a constaté que la courbe donnant la position en Barks de ses 24 B.C. (sur une échelle logarithmique) en fonction de la fréquence (sur une échelle logarithmique), était superposable à celle de la Tonie en Mels.

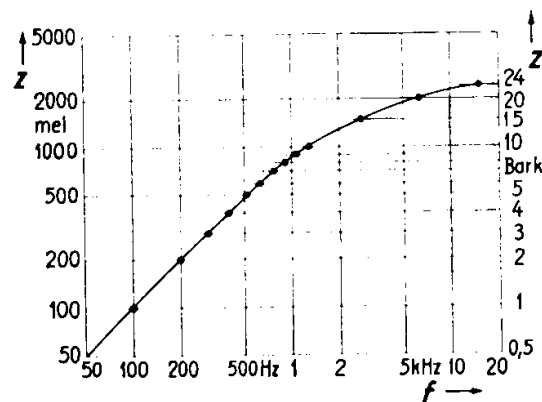


Figure 5 : relation de proportionnalité (empirique) des Mels et des Barks (d'après Zwicker).

Cependant il semble que Zwicker se soit un peu trompé sur ce point. En fait on trouve que la relation Tonie = f (Fréquence) n'est nulle part linéaire.

2.2. Facteurs de variation de la hauteur tonale

De même que la sonie ne dépend pas que du niveau d'intensité acoustique, la hauteur tonale est déterminée essentiellement par la fréquence, mais pas seulement par celle-ci.

Avec l'intensité

L'expérience montre que la hauteur des sons:

- décroît avec l'intensité pour des sons de moins de 2000 Hz,
- croît avec l'intensité pour des sons de plus de 4000 Hz.

Ces variations peuvent atteindre au maximum 5%, soit presque un demi-ton musical !

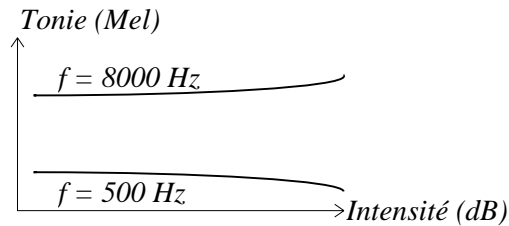


Figure 6 : sens des variations de la hauteur tonale avec l'intensité, pour différentes fréquences.

On peut noter que cela semble incompatible avec la théorie de la détermination temporelle de la hauteur (rappel : avec le « verrouillage de phase », les informations sont transmises au cerveau par des décharges nerveuses plus ou moins synchrones avec l'excitation) ; mais le codage temporel doit être décodé aux niveaux supérieurs de la chaîne auditive, et ce décodage peut être affecté par la position et le taux de décharge des neurones actifs.

Du reste, la théorie spatiale (i.e. celle basée sur la tonotopie ou localisation de l'excitation le long de la membrane basilaire, en fonction de la fréquence) est également en défaut : on observe bien un déplacement du pattern de stimulation le long de la membrane avec le niveau, mais pas toujours dans le sens prévisible.

Avec la durée

Moore a mesuré le seuil différentiel de hauteur (voir plus loin) pour des sons purs pulsés de durée variable.

Lorsque cette durée est relativement courte, le spectre du son pulsé contient de l'énergie à d'autres fréquences que celles du son pur. Ces fréquences altèrent donc le pattern d'excitation produit, comme l'illustre la Figure 7.

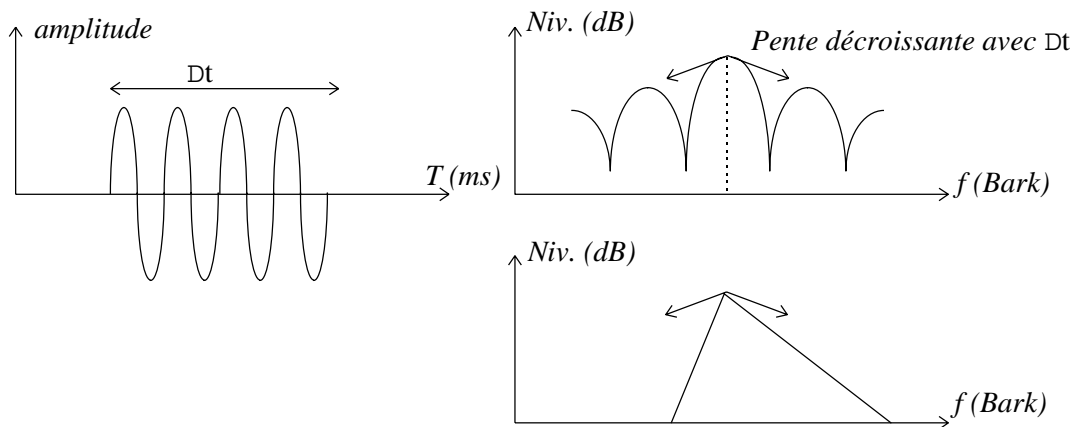


Figure 7 : déformation du spectre d'un son pur pour une durée courte, et incidence sur le pattern d'excitation produit dans le système auditif.

En-dessous d'une certaine durée du son pur, la pente du spectre est inférieure à celle du pattern d'excitation donc un pattern plus large sera excité: la sensation s'écarte de celle du son pur, puis la sensation de hauteur même s'estompe.

On conçoit donc que le seuil différentiel de hauteur est lié à la durée du son. En effet, on a montré que ce seuil Df (voir plus loin : *Difference Limen in Frequency*) est lié à la durée d par une relation d'incertitude du type :

$$Df \cdot d \approx K$$

Pour Df en Hz et d en s., la constante K est de l'ordre de 0,24. En fait sa valeur exacte dépend de l'intensité et de l'enveloppe d'amplitude.

En présence d'un bruit masquant

La hauteur perçue d'un son partiellement masqué est affectée par le bruit masquant:

- pour un bruit masquant plus aigu, la hauteur se déplace vers les graves
- pour un bruit masquant plus grave, la hauteur se déplace vers les aigus

Autrement dit, le déplacement est toujours dans le sens où la hauteur perçue s'éloigne de celle du masque.

2.3. Seuils différentiels de hauteur

Comme pour les autres grandeurs comme l'intensité, la hauteur perçue transmet souvent plus d'information par ses variations que par sa valeur absolue. Ce sont bien les variations de hauteur qui constituent une mélodie. En outre, l'évolution temporelle de la hauteur d'un son est importante :

- comme signal d'alerte : produite par effet Doppler, elle nous informe sur la vitesse de la source.
- du point de vue de l'agrément : une vitesse de rotation instable d'un disque vinyl produit l'effet très désagréable de « pleurage ».

On s'intéresse donc à notre capacité de discrimination des sons en fréquence, par les « Seuils Différentiels en Fréquence » (ou *DLF*, *Difference Limens in Frequency*) qui se mesurent par diverses méthodes, notamment :

- le choix forcé : entre 2 sons A et B de hauteurs légèrement différentes présentés successivement dans un ordre aléatoire, le sujet doit choisir lequel est le plus haut. Le seuil est choisi au point de 75% de réponses justes sur la courbe des réponses.
- par modulation : on mesure alors les « *Frequency Modulation Difference Limens* » comme la plus petite amplitude de modulation de fréquence perceptible (à un rythme de modulation donné, en général autour de 4 Hz).

On observe qu'au mieux, dans des conditions de laboratoire :

le DLF Df est de l'ordre de 2 ‰ (2 pour 1000), soit 1/20^{ie} de demi-ton musical !

(par exemple, on peut tout juste distinguer les fréquences 1000 et 1002 Hz).

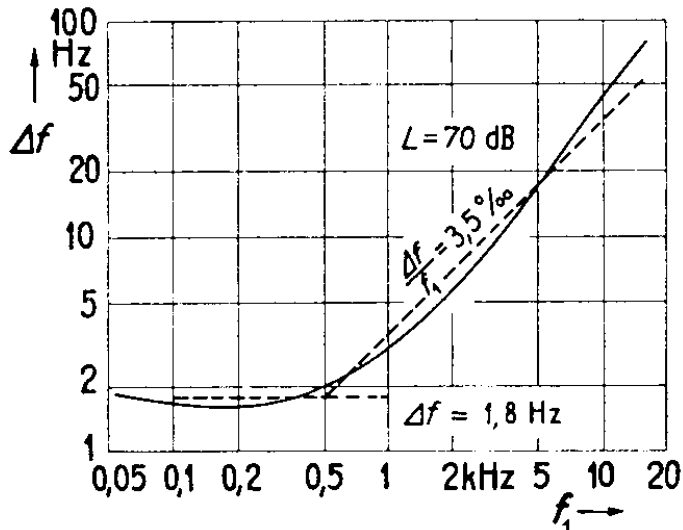


Figure 8 : variation des Seuils Différentiels de Fréquence avec la fréquence de mesure

Zwicker a tenté de fournir une explication des DLF par la perception de changements d'intensité en certains points du pattern d'excitation.

Sachant que les seuils différentiels d'intensité sont de l'ordre de 1 dB, et que la pente maximale du pattern d'excitation d'un son pur est (du côté grave) de 27 dB / Bark, à 1 kHz, deux hauteurs seraient discriminées si leur différence vaut au moins $1/27^{\text{e}}$ de Bark, soit $160/27 = 5$ Hz. On retrouve l'ordre de grandeur des DLFs mesurés.

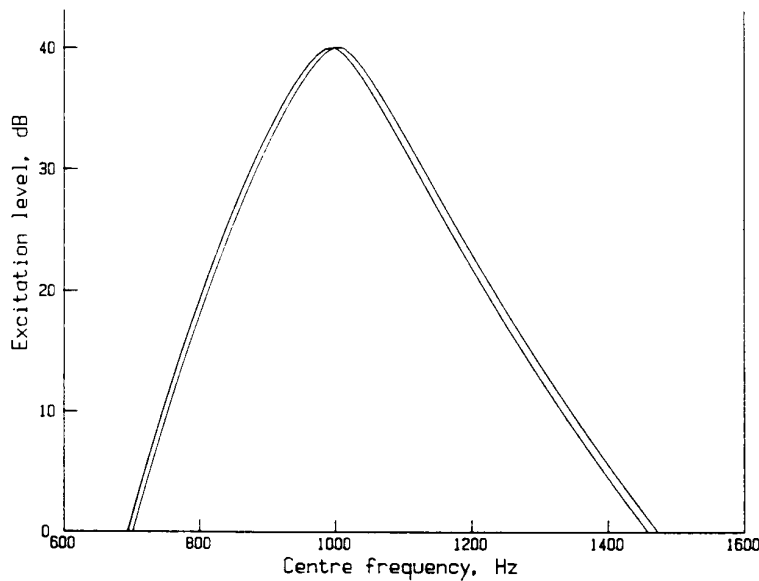


Figure 9 : explication des DLF par la perception de variations d'intensité : on a représenté les patterns d'excitation produits par deux sons purs de fréquences proches.

Mais en réalité le DLF ne varie pas exactement comme les bandes critiques, ce qui suggère que d'autres mécanismes sont aussi impliqués. En fait, le DLF serait déterminé essentiellement par des mécanismes temporels (liés au verrouillage de phase) aux basses fréquences, et tonotopiques aux hautes fréquences.

3. Hauteur des sons complexes

Rappelons avant tout quelques définitions :

- un son dans le cadre présent est un objet sonore qui comporte une ou plusieurs hauteurs - ce en quoi il se différencie d'un bruit⁵,
- dès qu'il ne s'agit pas d'un son pur (i.e. sinusoïdal), on dira que c'est un son complexe,
- une raie est une fréquence particulière à laquelle un son contient de l'énergie,
- lorsqu'un son est constitué de raies, celles-ci sont appelées les partiels du son complexe,
- si ces partiels sont multiples d'une même fréquence, ils seront alors appelés les harmoniques et l'on dira que le son lui-même est un son harmonique,
- dans ce dernier cas l'harmonique n° 1 (et non pas 0 ! ! !), est appelée la fondamentale du son.

Lorsque les sons complexes sont composés de raies, harmoniques ou non, l'oreille synthétise souvent la perception de ces raies pour entendre une ou plusieurs hauteurs.

C'est évidemment le cas dans sons harmoniques, mais pas uniquement.

Contrairement à quelques idées fausses :

1. la hauteur ne correspond pas forcément à une raie présente dans le spectre, qui serait par exemple celle de fréquence la plus basse - même pour des sons harmoniques,
2. elle ne correspond pas non plus à la zone de plus grande amplitude du pattern d'excitation nerveuse,
3. d'ailleurs, pour les sons complexes, la perception de la hauteur n'est jamais indépendante de celle du timbre.

3.1. Hauteurs virtuelles de sons harmoniques

Expérience du "fondamental absent"

Dans un son complexe harmonique si on enlève le fondamental on continue dans beaucoup de cas à « l'entendre » c'est-à-dire à entendre la même hauteur.

Ainsi si l'on enlève successivement, à partir des plus graves, les 10 harmoniques d'un son de fondamental 200 Hz : le timbre du son varie, mais pendant assez longtemps la hauteur perçue reste la même, correspondant à celle d'un son pur de 200 Hz.

Dans une autre expérience, on masque la zone fréquentielle dans laquelle se trouvent les fréquences fondamentales d'une mélodie. On entend toujours les mêmes hauteurs, et la mélodie est inchangée.

Nous avons donc mis en évidence l'existence de hauteurs qui ne correspondent pas à des raies ni même à la présence d'énergie dans le spectre.

Ce phénomène suffirait, notons-le, à remettre en question une théorie de la perception fréquentielle par tonotopie pure.

Une autre expérience plus spectaculaire montre que notre perception de hauteur peut être attachée à un mode d'écoute, analytique ou synthétique, qui est très influencé par le mode de présentation des composantes du son : dans cette expérience, on présente 2 raies de 1200 et 1400 Hz, qui donnent lieu à la perception d'une hauteur de 200 Hz ; puis on diminue progressivement le 1400 Hz jusqu'à le noyer dans un bruit masquant : le sujet peut continuer pourtant à « entendre » la même hauteur de 200 Hz !

Cette hauteur perçue est appelée « hauteur virtuelle » ou « résidu » (Schouten, 1940) (parfois la hauteur perçue est bien présente et non virtuelle ; le terme « résidu » est alors plus approprié ; il rend simplement compte de l'écoute synthétique vers la perception d'une seule hauteur).

⁵ Notons qu'en acoustique de l'environnement et du confort sonore, la définition du bruit est beaucoup plus floue et dépendante de la situation

Dans un premier modèle, on peut penser que la hauteur virtuelle perçue est le « fondamental absent » des raies présentes. Autrement dit, ce serait la (plus grande) sous-harmonique de celles-ci, soit leur PGCD.

Ambiguïté d'octave

Par modulation de fréquence à 200 Hz d'une porteuse de 2200 Hz, on obtient un son composé des trois fréquences 2000, 2200 et 2400 Hz. Dans ce son, l'oreille entend le « fondamental absent », soit 200 Hz.

Si maintenant le niveau du 2000 Hz décroît progressivement, à partir d'un certain point le « poids spectral » de cette composante devient faible, et l'on entendra plutôt 400 Hz. Au point limite, on entend simultanément les deux hauteurs concurrentes : il y a ambiguïté d'octave.

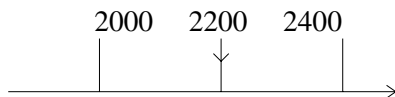


Figure : expérience de l'ambiguïté d'octave

Au-dessous de 500 Hz, notre perception de hauteur de sons purs est très approximative ; il apparaît que les hauteurs perçues dans le bas du registre du piano (dont la note la plus grave est d'environ 27,5 Hz), sont déterminées par les harmoniques. Le fondamental est d'ailleurs quasi-inexistant pour les notes les plus graves. Le piano est accordé sur les battements produits entre les harmoniques et le fondamental d'octaves supérieures.

3.2. Sons non harmoniques

Expérience et premier modèle

Tout se complique lorsque les hauteurs présentées sont un peu modifiées.

Ainsi, si l'on modifie à 2040 Hz la porteuse de l'exemple précédent, on obtient trois raies de 1840, 2040 et 2240 Hz : y a-t-il perception du « fondamental absent » 40 Hz ? Non.

On entendra en fait :

- une hauteur de ~204 Hz (3 sous-harmoniques de chaque raie, 204,4 + 204 + 203,6 sont en quasi-coïncidence, l'écart étant inférieur au seuil différentiel de l'oreille),
- 2 hauteur concurrentes de 185 et 227 Hz.

Il y a donc ambiguïté de hauteur.

Cette expérience suggère un algorithme pour déterminer la perception de la hauteur de sons complexes (Walliser, 1968) :

1. la hauteur correspondant à la différence de fréquence entre 2 partiels adjacents est approximativement déterminée (ici : 200 Hz),
2. Une sous-harmonique du *partiel le plus bas* est déterminée, la plus proche de cette différence.

Ex. précédent: $1840 / 9 = 204,4$ Hz.

Tehardt (1972) a proposé une variante affinée de cette méthode :

2. On choisit plutôt une sous-harmonique d'un partiel « dominant » ou « résolvable », c'est à dire un partiel qu'on pourrait entendre individuellement émerger du tout.

Poids spectral et "sillance" ("salience")

Ainsi, dans le cas de sons non harmoniques, la ou les hauteurs perçues (là encore il peut y avoir ambiguïté) sont déterminées par :

- l'espacement entre les raies,
- la présence d'une sous-harmonique d'un partiel « dominant », de fréquence proche de cet espacement.

2040	2240	2440
1020	1120	1220
680,0	746,7	813,3
510,0	560,0	610,0
408,0	448,0	488,0
340,0	373,3	406,7
291,4	320,0	348,6
255,0	280,0	305,0
226,7	248,9	271,1
204,0	224,0	244,0
185,5	203,6	221,8
	186,7	203,3
		187,7

Figure 10 : coïncidences de sous-harmoniques des partiels et hauteurs virtuelles concurrentes dans un son non harmonique.

Tehardt suggère que ces partiels dominants sont plutôt dans la zone 200-2000 ou 500-1500 Hz. L'amplitude des différentes raies présentes est prise en compte affectée d'un facteur de pondération ou « poids spectral » qui est le plus important dans cette zone de fréquences.

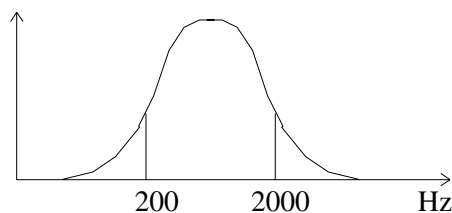


Figure : poids spectral ou « saillance » des composantes spectrales pour la perception de hauteur

De plus Tehardt a supposé que l'importance des fréquences moyennes est probablement liée à un apprentissage, qui favorise les sons qui rentrent dans notre tessiture vocale :

- très jeune nous sommes exposés à des sons harmoniques (parole),
- nous y apprenons à extraire d'une fréquence donnée des indices de présence d'un son complexe dont cette fréquence est une harmonique.

3.3. Ecoute analytique et écoute synthétique

Suivant le cas l'oreille perçoit différemment un ensemble de composantes sonores :

- soit comme un ensemble de composantes distinctes (perception « analytique »),
- soit comme un tout, avec une ou plusieurs "hauteurs" et un timbre associé (perception « synthétique »).

Le mode de perception dépend notamment des stimuli, de la volonté de l'auditeur, et du mode de présentation des stimuli (ordre d'arrivée ou de suppression des partiels). Il est ainsi remarquable de noter que deux sons de même hauteur virtuelle ne sont pas nécessairement fusionnés : supposons que durant l'écoute d'un ensemble formé des harmoniques 5, 6, 7 et 8 de 200 Hz, qui produit une

hauteur virtuelle de 200 Hz, on ajoute subitement un son pur à 200 Hz. Ce dernier son sera perçu comme un second objet sonore, distinct du premier, bien que leurs hauteurs soient identiques. Le mode de présentation des stimuli influe donc sur leur groupement possible, lequel est par ailleurs étroitement liée à leur éventuelle consonance ou dissonance, notion que nous explicitons ci-dessous.

4. Modèles pour la perception de la hauteur

4.1. La tonotopie et les modèles de reconnaissance de formes

Le principe de ce premier modèle est que la hauteur est corrélée à une position bien définie sur la membrane basilaire.

Il repose sur deux postulats :

- que l' « analyse fréquentielle » du stimulus dans la cochlée est de nature tonotopique : deux fréquences pures différentes produisent deux patterns d'excitation différents, et l'excitation de fibres nerveuses différentes, et
- que la hauteur du stimulus est directement liée au pattern d'excitation produit.

La première hypothèse est assez bien admise et corroborée par l'expérience.

La seconde cependant reste controversée. En effet les sons complexes produisent un pattern d'excitation à maxima multiples, dont le plus fort ne correspond pas au fondamental, comme le montrent :

- l'expérience du fondamental absent, ou
- l'expérience du fondamental masqué par un bruit passe-bas.

Cette théorie ne rend donc pas complètement compte de la perception de hauteur des sons complexes, à moins qu'on n'y ajoute l'idée d'un traitement postérieur au niveau central.

C'est ce que suppose Terhardt lorsqu'il propose (en 1969/70) l'algorithme suivant :

Traitement de la cochlée :

1. Analyse spectrale du signal (qu'on peut modéliser par exemple par une sur FFT 800 points, qui donne l'énergie dans 400 canaux fréquentiels),
2. Extraction des composantes du son (en ne retenant que les points d'analyse comportant de l'énergie, on identifie par exemple 50 partiels),
3. Evaluation de l'effet de masque (connaissant les courbes d'effet de masque des sons purs, on élimine ceux qui ne sont pas perçus ; il reste par exemple 10 partiels, qui sont autant de « hauteurs spectrales » possibles).

Traitement du système nerveux central :

1. Pondération des composantes (par la courbe empirique du « poids spectral » vue plus haut),
2. Extraction des hauteurs virtuelles (soit les sous-harmoniques d'un partiel « dominant », de fréquence proche de l'espacement entre les hauteurs spectrales ; on obtient par exemple 4 hauteurs virtuelles).

Rappelons que suivant le cas, les composantes spectrales solitaires peuvent être entendues individuellement, tandis que les composantes spectrales en relations harmoniques seront entendues groupées.

4.2. Le codage temporel

Selon cette autre hypothèse la perception de la hauteur dépendrait du rythme et du pattern de décharges des fibres nerveuses. Elle se base sur la propriété de "verrouillage de phase" qui est observée (seulement jusqu'à 5 kHz ; mais justement, la perception de hauteur est très déréglée au-delà). La Figure 11 illustre cet aspect temporel du codage des fréquences dans le système auditif.

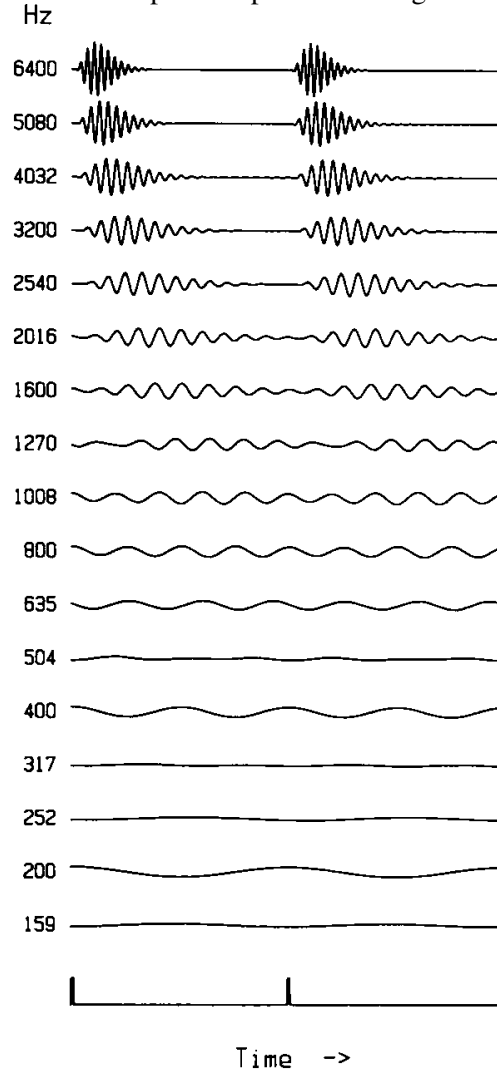


Figure 11 : codage temporel de l'information fréquentielle : patterns des réponses des fibres du nerf auditif de différentes fréquences caractéristiques, à un train d'impulsions à 200 Hz.

4.3. Nécessité d'un modèle mixte

Aucun des deux modèles n'étant individuellement suffisant pour expliquer tous les aspects de la perception de hauteur. On peut noter divers arguments en faveur de chacun :

Arguments en faveur de la "reconnaissance de formes":

Seul ce modèle permet de proposer une explication à divers phénomènes :

- l'expérience du fondamental absent,
- la prédominance des harmoniques basses ou « résolubles », dans la perception de hauteur,
- la diplacousie, pathologie dans laquelle la perception de hauteur diffère sur les deux oreilles.

Arguments en faveur du codage temporel

Ce modèle permet d'expliquer que des hauteurs (faibles) peuvent être entendues :

- alors que les deux harmoniques présentes sont de fréquences trop proches pour être résolubles,
- ou quand les stimuli n'ont pas de structure spectrale bien définie (comme les bruits modulés, qui produisent une perception de hauteur correspondant au rythme de la modulation).

Un modèle mixte

Pour tenter d'expliquer l'ensemble des propriétés de notre perception de la hauteur, Moore a proposé une combinaison de des deux modèles, qu'on peut représenter schématiquement ainsi :

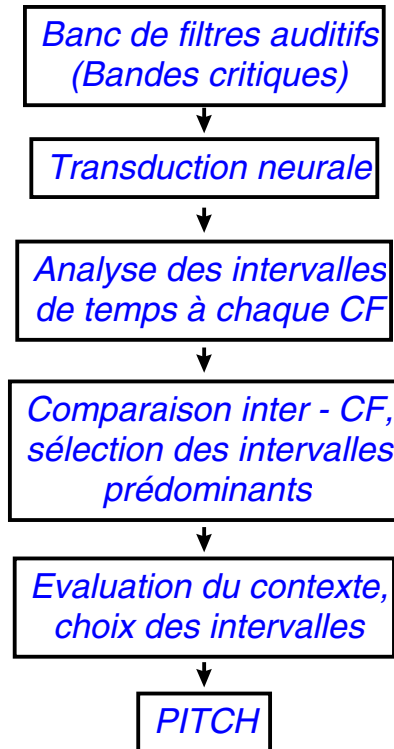


Figure 12 : combinaison des modèles de codage tonotopique et temporel de la hauteur (Moore)

5. Intervalles et gammes

5.1. Construction des gammes

Un intervalle est le rapport entre deux hauteurs.

Les grecs (Pythagore, Erasthoxène, Archimède) ont les premiers remarqué les similitudes de perception qui existent entre certains intervalles, principalement l'octave (rapport 2 entre les hauteurs) qui a la particularité que toutes les harmoniques du son aigu sont également harmoniques du son grave. Ils ont ensuite cherché à découper cet intervalle essentiel en intervalles plus petits pour former des gammes. Diverses approches sont possibles, suivant que l'on procède d'une manière plus inspirée de la physique, des mathématiques ou simplement de l'écoute.

La gamme naturelle

La gamme naturelle est basée sur des coïncidences d'harmoniques. Ainsi la gamme de do majeur contient :

- la Quinte (3/2) de do : sol,
- la Tierce majeure (5/4) de do : mi,
- etc...

La gamme naturelle contient 12 intervalles. Elle est juste *harmoniquement* puisque tous ses sons sont harmoniques d'un même fondamental, propriété qui détermine la consonance, mais se révèle fausse *mélodiquement*.

La gamme de Pythagore

La gamme de Pythagore est basée sur une succession d'intervalles de quinte ($[3/2]^n$). elle est "juste" *mélodiquement*, mais ne l'est pas *harmoniquement* (les harmoniques des sons ne coïncident plus, les accords sont faux...)

La gamme tempérée

La gamme tempérée, apparue au XVIII^e siècle, réalise un compromis entre les deux précédentes, en confondant les dièses et les bémols. Les intervalles réguliers entre les notes permettent de jouer dans n'importe quelle tonalité ; elle est constituée de 12 intervalles (logarithmiquement) égaux donc égaux à $2^{1/12}$.

Autres gammes

On peut citer encore d'autres gammes, par exemple la gamme d'Aristoxène Zarlin, la gamme de Mercator-Holder ou la gamme de Delezenne... Ces gammes « occidentales » comportent 12 demi-tons espacées plus ou moins logarithmiquement ; mais il en existe d'autres :

- la gamme pentatonique (à 5 tons) est à la base de la musique chinoise, mais se retrouve aussi dans les musiques celtiques ou des Indiens d'Amérique....
- les indiens Nasca du Pérou ont basé leur musique sur des échelles linéaires (et non logarithmiques) mais cela est assez exceptionnel.

Nous donnons les intervalles dans les principales gammes (par rapport à la tonique do):

Gamme	do	ré	mi	fa	sol	la	si	do
Naturelle	1	9/8	5/4	21/16	3/2	13/8	15/8	2
Pythagore	1	9/8	81/64	4/3	3/2	27/16	243/128	2
Tempérée	1	$2^{2/12}$	$2^{4/12}$	$2^{5/12}$	$2^{7/12}$	$2^{9/12}$	$2^{11/12}$	2

Notons qu'en principe, un piano (accordé selon la gamme tempérée) et un instrument à vent (dont certaines notes aiguës sont produites comme des harmoniques de notes plus graves) ne peuvent jamais être accordés. En fait la possibilité de modifier la hauteur -- en jouant sur les doigtés ou sur le souffle pour l'instrument à vent, et sur la durée et l'intensité du son pour le piano -- et la tolérance de l'oreille, font que ce désagrément reste du domaine théorique.

D'ailleurs, contrairement à une idée fréquente, un piano moderne n'est jamais rigoureusement accordé selon la gamme tempérée. L'accord est en fait un compromis entre les différentes gammes que nous avons évoquées. De plus, dans les extrêmes du registre, les octaves mêmes du piano sont « étirées » par rapport aux octaves « physiquement justes » de rapport 2, pour compenser le biais de l'appareil auditif.

5.2. L'oreille absolue

C'est la capacité de reconnaître et de nommer la hauteur d'un son musical sans référence à un son de comparaison.

Cette faculté est rare (elle concerne moins de 1% de la population).

Elle peut être entraînée dans une certaine mesure (mais les résultats sont rarement aussi bons que lorsque la faculté est "naturelle", i.e., en fait, acquise dans l'enfance).

5.3. Préférences

Les rapports de fréquence les plus agréables dans une mélodie sont-ils les plus rigoureusement harmoniques ? L'expérience suivante montre que non :

Tehardt a présenté à un large public trois versions d'une mélodie aiguë accompagnée d'un accompagnement grave :

- a) version « compressée » d'un demi-ton : la basse est en do, la mélodie en si,
- b) version « dilatée » d'un demi-ton : la basse est en do, la mélodie en do#,
- c) version « mathématiquement correcte » : la basse comme la mélodie sont en do.

Il a observé que si la version « dilatée » b) est jugée acceptable, et même préférée par 40% des sujets, en revanche la version « compressée » a) est toujours jugée désagréable.

Perception du timbre des sons

1. Consonance et dissonance

Les notions opposées de consonance et de dissonance sont une caractéristique très importante de la perception des sons, étroitement liée à la perception de hauteur. Depuis Pythagore, la consonance est une clé de la musique. C'est de fait un trait robuste du timbre des sons.

La consonance exprime la préférence de l'oreille pour certaines associations de fréquences entendues simultanément. L'étude de la consonance de sons composés de 5 premières harmoniques d'amplitudes égales montre que la consonance atteint des maxima locaux quand le rapport des fondamentales est rationnel.

Les intervalles les plus consonants sont ainsi :

- l'octave (rapport 2),
- la quinte juste (rapport 3 ramené entre 1 et 2, soit $3/2$),
- la quarte (rapport $4/3$),
- la tierce majeure (rapport 3 ramené entre 1 et 2, soit $5/4$),
- la tierce mineure (rapport $6/5$), la sixte (rapport $5/3$), etc.

L'écoute simultanée de sons consonants produit généralement un accord. Par exemple, la hauteur de base étant appelée « tonique », l'ensemble « tonique + quinte + tierce majeure » constitue un accord parfait majeur. Notons que cet ensemble est constitué des 5 premières harmoniques de la tonique, ramenées à une seule octave.

Le groupement des composantes du son se fait donc selon plusieurs étages de complexité : d'abord le groupement éventuel des partiels en un son - groupement naturel s'il s'agit des partiels d'une source unique -, puis groupement de sons émis par différentes sources en un objet sonore qui peut être un accord...

Notons que certains auteurs distinguent la consonance tonale ou consonance sensorielle, de la consonance au sens musical qui dépend aussi de certaines règles culturelles.

1.1. Consonance des sons purs

Si l'on augmente progressivement la différence de fréquence df entre deux sons purs écoutés simultanément, on observe que :

- tant que df / f est faible, sons sont consonants, en même temps qu'on perçoit des battements,
- puis la dissonance apparaît, croît et devient maximale vers un quart de bande critique (soit près de 20 Hz en basses fréquences, puis 4% de la fréquence soit un peu moins d'un demi-ton aux fréquences plus élevées)

- les sons purs redeviennent consonants dès que la différence de fréquence dépasse la largeur d'une bande critique. Notons que ce n'est pas le cas des intervalles de tierce (voire de quinte) en basses fréquences ; de fait, ces intervalles sont dissonants et on les évite dans les basses.

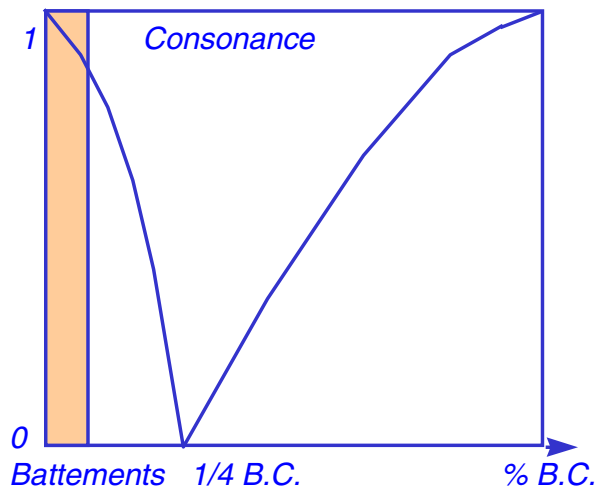


Figure 13 : consonance des sons purs en fonction du rapport de leurs fréquences.

1.2. Consonance des sons complexes

On observe que les sons de rapport de fréquence non rationnels sont généralement dissonants. La Figure 14 représente schématiquement l'évolution de la consonance de sons harmoniques en fonction du rapport de leurs fréquences fondamentales.

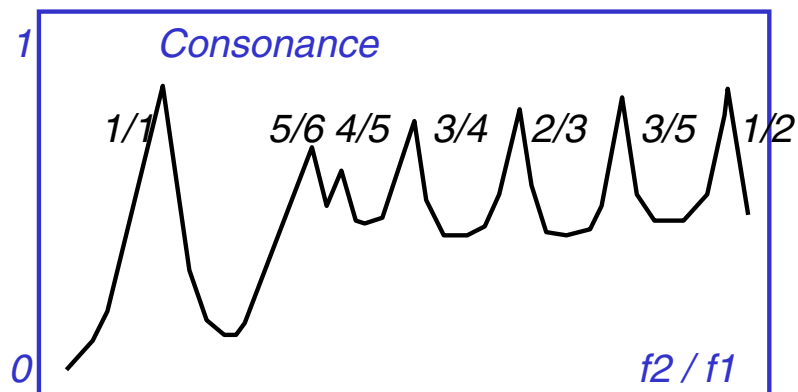


Figure 14 : Consonance de sons harmoniques en fonction du rapport de leurs fréquences fondamentales.

Cette consonance s'explique par la coïncidence plus ou moins forte d'harmoniques entre les deux sons, comme l'illustrent la Figure 15 et la Figure 16.

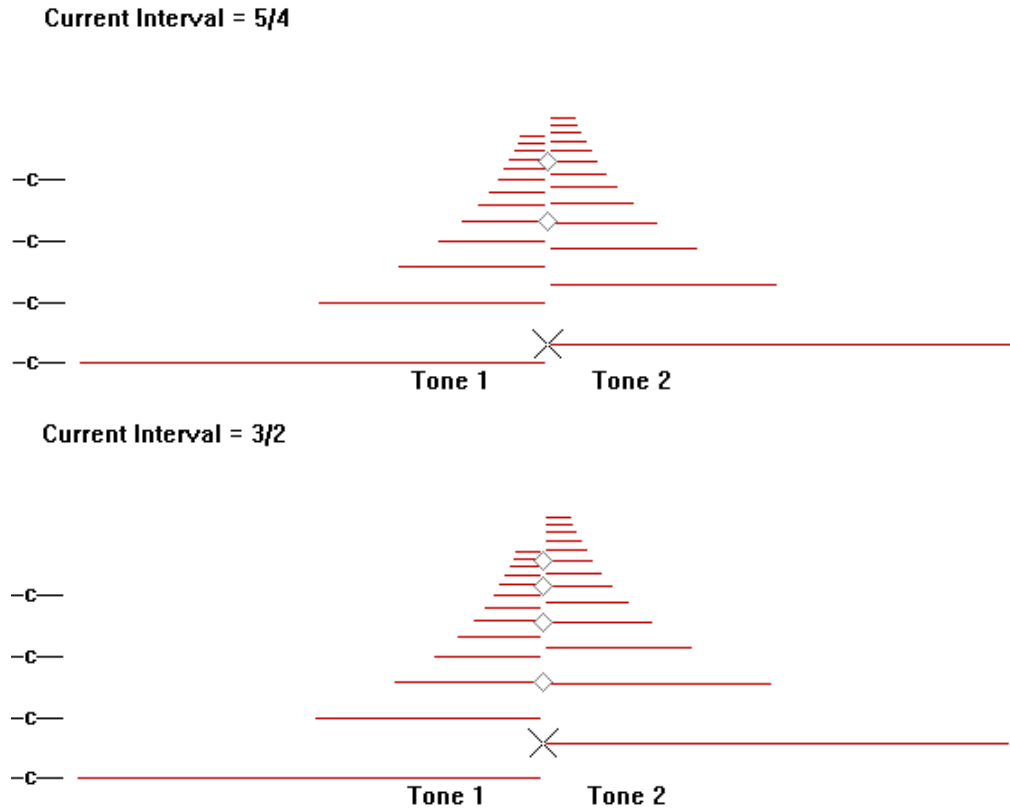


Figure 15 : coïncidence d'harmoniques entre sons séparés par les intervalles de tierce (en haut), et de quinte juste (en bas)

Une théorie suggère que cette préférence pour les rapports de fréquence simples est acquise très jeune par l'exposition aux sons harmoniques tels que ceux de la voix.

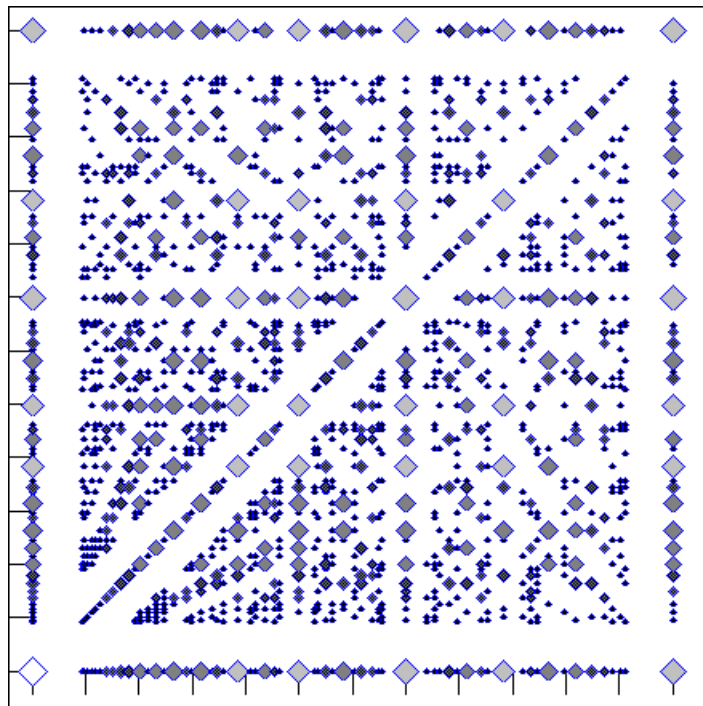


Figure 16 : carte de consonance entre sons harmoniques dans des rapports de fréquences fondamentales de 0 à 1 octave

2. La notion de timbre

La notion de timbre prend différents sens dans le langage courant, suivant qu'on l'applique par exemple à une voix, ou à une voyelle isolée. A l'échelle d'un son, il est défini selon l'ANSI comme « l'attribut de la sensation auditive qui permet de différencier deux sons de même hauteur et de même intensité » ; c'est donc le fourre-tout des propriétés perceptives du son non encore étudiées ; comme il fallait s'y attendre c'est tout sauf une grandeur monodimensionnelle.

Selon une conception « classique » due à Helmholtz, le timbre résulte principalement de la composition spectrale de la partie quasi-stationnaire du son. Cependant cette conception résiste mal aux observations suivantes :

- le timbre résiste aux altérations de la transmission
- l'attaque joue souvent dans le timbre un rôle fondamental

On étudie donc l'influence sur le timbre, notamment, de la composition spectrale à un instant donné et de son évolution temporelle.

2.1. Timbre et composition spectrale

La composition spectrale d'un son est complexe ; elle est souvent décomposable en enveloppe et structure fine (les formants et les harmoniques dans le cas de la voix). La manière dont l'énergie est concentrée selon l'axe des fréquences peut être mesurée à travers :

- le **centre de gravité spectral** (CGS) qui influe sur le caractère « rond » ou « brillant » du son ;

le CGS est défini par :

$$CGS = \frac{\sum_{k=1}^N k \cdot A_k}{\sum_{k=1}^N A_k} \quad (Hz)$$

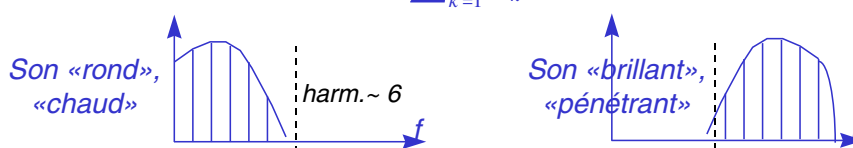


Figure : illustration de la notion de centre de gravité spectral

- le **rayon de giration spectral** (RGS) qui complète cette mesure par une notion de « compacité » du son ; comme en mécanique, il représente la largeur de l'objet (l'enveloppe spectrale) autour du centre de gravité :

$$RGS = \sqrt{\frac{\sum_{k=1}^N (k - CGS)^2 \cdot A_k}{\sum_{k=1}^N A_k}} \quad (Hz)$$

En ce qui concerne la structure fine, on observe que l'harmonicit  n'est pas une r gle absolue : une l g re inharmonicit  rend le son plus « vivant », « color  ».

2.2. Timbre et enveloppe temporelle

Mais l'enveloppe temporelle du son ou de ses composantes a aussi une grande importance ; ainsi :

- l'enveloppe de l'attaque est caract ristique du type de son: corde frott e, frapp e, gratt e, etc... (retourn , le piano ressemble   un son d'harmonium).
- les sons « cuivr s » sont caract ris s notamment par un niveau d'harmoniques  lev es qui cro t dans le temps.

Du reste, de mani re g n rale l'oreille n'aime pas la monotonie et se focalise sur tout ce qui est non-stationnaire : vibrato, effet choral, transitoires, variations du spectre dans le temps.

L'identification d'un timbre est ainsi fonction du caractère périodique (i.e. tonal) ou non du son, des évolutions (de l'enveloppe, du spectre, de la périodicité) et de la nature des sons précédents et suivants.

Enfin l'homogénéité des timbres d'instruments est difficilement explicable par un modèle simple ; en effet :

- transposer des hauteurs d'harmoniques ne transpose pas le timbre (notamment des instruments « à formants »),
- le son du basson ou du piano est homogène malgré un « rapport d'harmoniques » (une enveloppe spectrale) non constant(e) dans les différents registres de l'instrument.

3. Vers un espace des timbres

Quels sont les paramètres les plus importants du timbre ? Puisqu'ils ne correspondent pas à des paramètres physiques simples on commence par essayer de les caractériser en définissant des attributs verbaux : le son peut être «mat», «velouté», «doux», etc.

Par des tests de similarité (par paires ou par triplets) suivis d'analyses factorielles (Plomp, 1970; Wessel, 1979; McAdams, 1992...) on obtient comme composantes principales :

- la brillance, caractérisée principalement par le centre de gravité spectral
- le mordant, caractérisée par la durée de l'attaque : de l'ordre de 1 à 2 ms pour le piano, 50 ms pour le violon

Brillance et mordant rendent compte de 85% de l'information caractéristique des timbres.

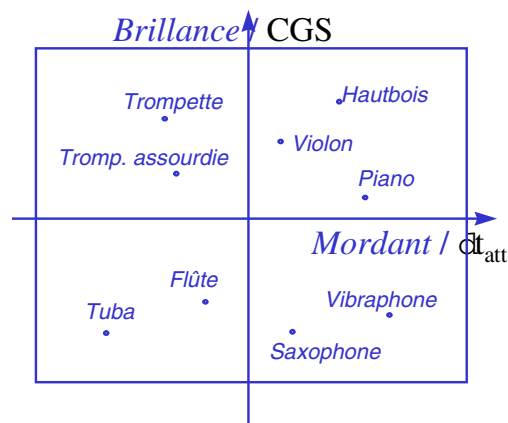


Figure : un espace des timbres à deux dimensions, la brillance et le mordant.

Ecoute binaurale et perception de l'espace sonore

La perception de l'espace sonore peut être appréhendée selon deux attitudes particulières du système auditif, auxquelles nous nous intéressons successivement :

- la localisation des sources,
- l'étendue subjective des sources.

Ces capacités utilisent essentiellement, mais pas exclusivement, des indices binauraux.

1. La localisation des sons

D'abord il faut souligner la différence entre la position réelle de la source physique du son, et la localisation qui intéresse un événement auditif subjectif. Cet événement auditif peut :

- exister indépendamment de tout événement visuel,
- ne pas être fixe : la stéréo nous permet de reproduire un événement auditif localisé entre deux enceintes, qui éventuellement se déplace,
- être très peu localisé - et c'est là que la notion d'étendue subjective prendra son importance.

On s'intéresse aux capacités de localisation en direction et en distance. La direction d'incidence du son est identifiée selon les conventions suivantes :

- l'azimut est l'angle que fait la direction incidente avec le demi-plan vertical situé dans l'axe du visage du sujet.
- l'élévation est l'angle que fait cette direction incidente avec un plan horizontal.

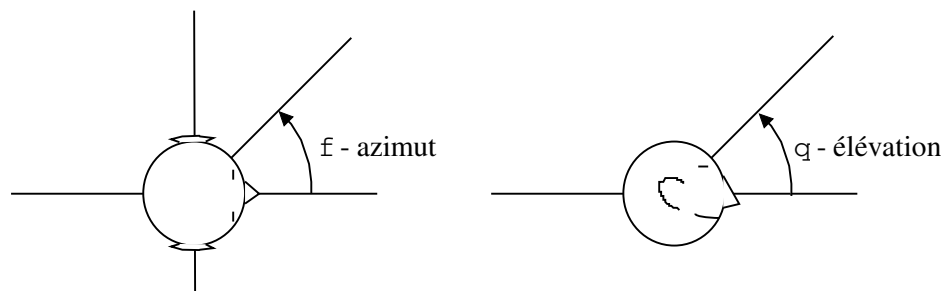


Figure : l'azimut et l'élévation

La localisation en azimut

On peut mesurer la capacité de localisation en azimut, en utilisant comme source soit un bruit à bande étroite soit un bruit blanc pulsé de durée 100 ms. On observe que cette capacité est bonne vers l'avant ou vers l'arrière (de l'ordre de quelques degrés), un peu moins bonne sur les côtés.

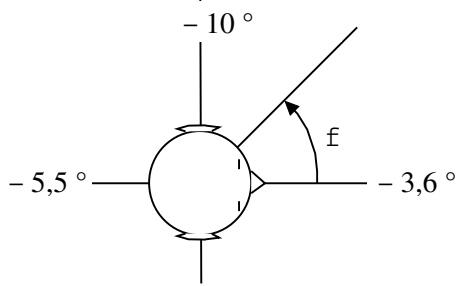


Figure : les capacités de localisation **binaurales** dans un plan horizontal

Comme l'ont proposé Stevens et Newman en 1936, il semble que deux indices binauraux soient impliqués essentiellement dans la localisation :

- l'un, efficace surtout en **Basses Fréquences**, est la **Différence Interaurale de Temps** ("Interaural Time Difference" ou *ITD*).

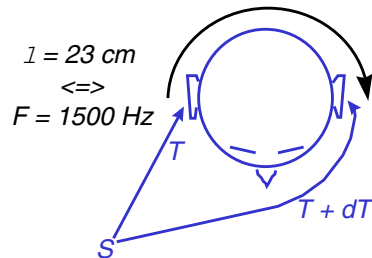


Figure : la différence interaurale de temps.

Notons que pour des sons périodiques, la distance interaurale étant de 23 cm environ, si la longueur d'onde est inférieure ou égale à 23 cm cette information est ambiguë : différentes directions incidentes produisent une même différence de phase. La différence de phase n'est en fait utile que pour les sons de fréquence inférieure à 1500 Hz.

- l'autre indice, efficace surtout en **Hautes Fréquences**, est la **Différence Interaurale d'Intensité** ("Interaural Intensity Difference" ou *IID*).

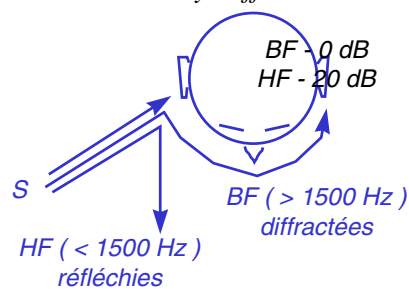


Figure : la différence interaurale d'intensité.

On peut modéliser la tête comme un obstacle sphérique dans le champ acoustique. On sait que son effet dépend de la forme de l'obstacle, et de la fréquence ; essentiellement :

- aux hautes fréquences - si la longueur d'onde est très inférieure aux dimensions de l'obstacle, le son est presque entièrement réfléchi par celui-ci (si l'on néglige les phénomènes éventuels de transmission),
- aux basses fréquences - caractérisées par une longueur d'onde très supérieure aux dimensions de l'obstacle, l'effet de ce dernier est quasiment négligeable.

- entre ces situations extrêmes les effets de la diffraction sont plus ou moins marqués se traduisent par une atténuation plus ou moins importante du son dans le contournement de la tête.

L'atténuation ainsi produite varie en pratique de 0 à 20 dB environ.

On constate que les capacités de localisation sont les moins bonnes autour de 1500 Hz, région fréquentielle dans laquelle ni les indices temporels ni les indices d'intensité ne sont suffisamment porteurs d'information.

La localisation en élévation

L'étude des capacités de localisation en élévation donne des résultats très différents suivant le signal utilisé :

- pour des sons de parole, nous avons une capacité de localisation raisonnablement bonne : de l'ordre de 10° pour des sons provenant de l'avant, un peu moins bonne pour des sons provenant de l'arrière.

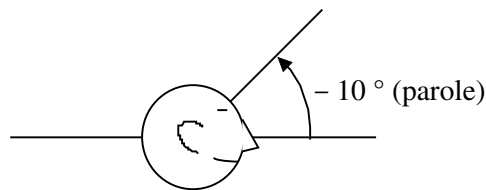


Figure : capacité de localisation en élévation des sons de parole

- pour des sons purs en revanche, les résultats sont tout à fait différents : l'élévation perçue d'un son pur n'est quasiment pas fonction de sa direction d'incidence réelle, mais uniquement de la fréquence ! Ainsi, un son de basse fréquence semble provenir d'une source « basse », un son de haute fréquence d'une source « haute ».

En fait on constate que c'est la largeur de bande du signal qui est déterminante. Une largeur de bande d'au moins deux tiers d'octave est nécessaire pour permettre une capacité de localisation en élévation.

Dans le cas de sons purs, Blauert s'est intéressé au trajet de l'événement auditif, en fonction de la fréquence. Lorsque celle-ci augmente, la provenance apparente du son contourne par deux fois la tête, selon la figure suivante :

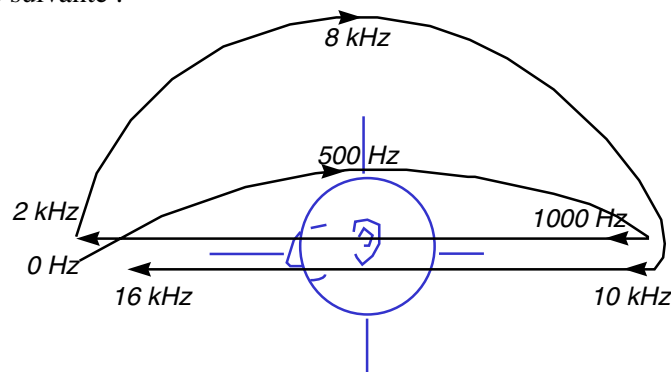


Figure : trajet vertical de l'événement auditif en fonction de la fréquence d'un son pur présenté dans l'axe de la tête

Blauert a ainsi mis en évidence l'existence de « bandes directionnelles », ou cônes de l'espace dans lesquels nous entendons préférentiellement une ou plusieurs fréquences de sons quasi-purs.

La localisation en distance

L'étude de la distance subjective de la source montre que pour un son pur, celle-ci est en premier lieu liée à l'intensité perçue : paradoxalement, un son pur semble provenir d'autant plus loin que son intensité est élevée.

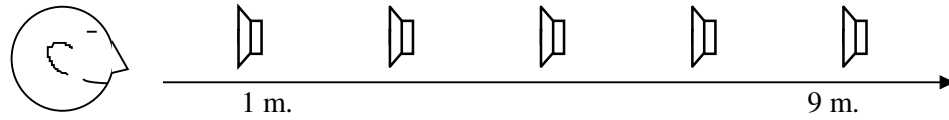


Figure : étude de la distance subjective de la source

Dans le cas de sons complexes la perception de la distance utilise de multiples indices :

- le niveau sonore absolu (en champ libre : $W = 4\pi r^2 I \dots$), notamment pour des sources multiples,
- les variations de spectre : celui-ci change sur des distances modérées, avec l'absorption de l'air, les hautes fréquences étant plus rapidement atténuées que les basses fréquences.
- les rapports d'intensité et les délais temporels son direct / sons réfléchis en espace clos.

Notre capacité d'exploitation de ces indices varie notablement selon notre familiarité avec la source. La précision dans tous les cas reste faible, avec des erreurs courantes de l'ordre de 20%.

La localisation des transitoires

Dans la localisation des transitoires les différences interaurales de temps prennent une plus grande importance que pour les sons purs.

On peut réaliser un test de détection d'un changement de localisation associé à l'*ITD* :

- sur des "bursts" de bruit ou de son pur, l'*ITD* seuil pour la détection est de 10 ms (soit environ 1° de latéralisation). L'acuité de localisation croît avec la durée du *burst* jusqu'à 700 ms.
- pour des clicks, l'*ITD* seuil est de l'ordre de 30 ms.

L'étude de la relation entre nos capacités de localisation des transitoires et leur **contenu fréquentiel et temporel**, montre que celles-ci :

- utilisent plus les basses fréquences que les hautes fréquences : un filtrage passe-bas altère peu les performances, alors qu'elles sont nettement dégradées par un filtrage passe-haut.
- utilisent plus l'enveloppe que la structure temporelle fine du signal. Ainsi, on détecte aussi bien une différence dans la phase de modulation à 300 Hz d'une porteuse à 4000 Hz, qu'un délai temporel d'un son pur à 300 Hz. C'est donc le délai temporel entre les enveloppes qui détermine la latéralisation (Henning, 1974).
- utilisent plus les différences interaurales dans la partie durable son, que les différences interaurales de l'attaque. Ainsi, si deux *bursts* de bruit sont présentés à chaque oreille avec une *ITD* sur les temps d'attaque, la latéralisation disparaît dès que la durée des *bursts* dépasse 10 ms. environ.

L'étude des capacités de localisation des transitoires met encore en évidence le phénomène de **l'adaptation binaurale** :

Hafter et coll. (1983, 1988) ont étudié la localisation d'un train de clicks filtrés HF, en mesurant les seuils de détection d'*ITDs* et d'*IIDs* en fonction du nombre n et de l'espacement I des clicks. Ils ont observé que :

pour $I \neq 10$ ms : le seuil décroît avec n (en $\frac{1}{\sqrt{n}}$) ; tous les clicks portent donc autant d'information.

pour $I < 1$ ms : (> 1000 Hz) : le seuil ne varie plus avec n ; ainsi le système auditif ne traite plus que l'attaque du train de clicks.

Il y a donc une « adaptation » rapide au rythme élevé des clicks ; on note qu'un nouvel indice temporel distinct ("*trigger*") rétablit alors la capacité à traiter les clicks suivants.

Les capacités monaurales de localisation

Si maintenant le sujet ne dispose plus que d'une oreille, sa capacité de localisation en azimut fait est sérieusement dégradée, comme l'illustre la figure ci-dessous :

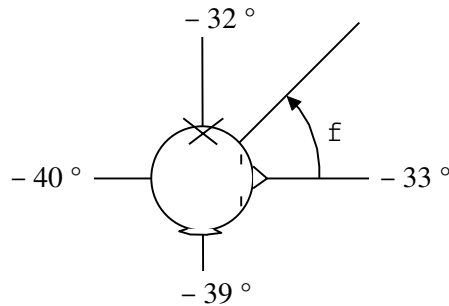


Figure : les capacités de localisation **monaurales** dans un plan horizontal

Dans l'espace, chaque direction d'incidence appartient à un « **cône de confusion** » sur lequel toutes les sources ont la même *ITD* et plus ou moins la même *IID*. La localisation, même binaurale, est donc sujette à des ambiguïtés, qui sont habituellement résolues par des mouvements de la tête.

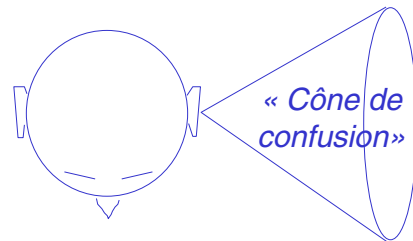


Figure : le cône de confusion

Si ces mouvements de la tête sont possibles, on constate que la localisation monaurale est presque aussi bonne que la localisation binaurale. Ainsi d'autres indices que les *ITD* et *IID* sont utilisés: en particulier, le changement du pattern spectral suivant la direction d'incidence.

Rôle de la tête et du pavillon

Les *ITDs*, les *IIDs* et les mouvements de tête ne suffisent pas à expliquer complètement de nos capacités de localisation notamment en élévation.

L'ensemble de la tête et du pavillon réalise un filtrage complexe, dont l'effet est notable de 500 à 16 000 Hz. Ce filtrage varie avec la direction d'incidence du son. On en mesure donc les fonctions de transfert, ou *HRTF* ("*head-related transfer functions*") en fonction de la direction d'incidence (azimut et élévation).

On distingue les fonctions de transfert :

en champ libre : il s'agit, pour une position donnée de la source, de la différence de ce que recueille un micro à l'entrée du conduit auditif externe suivant que l'ensemble tête et oreille externe est ou n'est pas présent.

monaurale : on mesure alors, sur une oreille, l'effet d'un déplacement (en azimut et en élévation) de la source, sur ce qui est reçu à l'entrée du conduit auditif.

interaurale : celle-ci s'obtient comme la différence de ce qui est reçu par les oreilles droite et gauche, à l'entrée du conduit auditif.

On observe de grandes variations, avec l'angle d'incidence, du niveau reçu par chaque oreille.

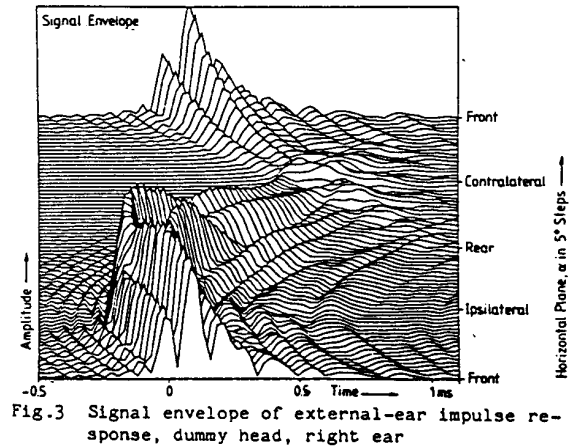
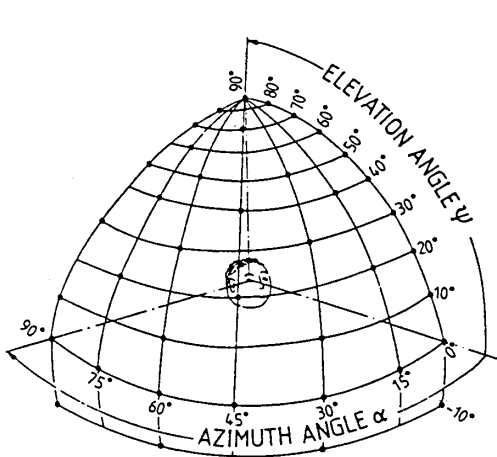


Figure : illustration de la mesure des HRTF ; à droite, on a représenté l'enveloppe temporelle du signal en fonction de la direction d'incidence, pour une oreille (Blauert, 1980).

Le pavillon intervient essentiellement sur les hautes fréquences, supérieures à 6 kHz. On peut ainsi montrer que nos capacités de localisation de sons HF (8 à 10 kHz) diminuent si on obstrue les reliefs du pavillon. Ces capacités restent quasi-inchangées vers 3 kHz.

Son importance est mise en évidence par deux expériences amusantes :

- une expérience due à Batteau (1967), qui montre que le son enregistré par deux micros à travers des pavillons artificiels, permet à la restitution binaurale, au casque, de bons jugements de localisation (i.e. à l'extérieur de la tête, et pas simplement de latéralisation) en azimut et en élévation. Sans les pavillons artificiels, ces performances étaient complètement perdues.



Figure : expérience de Batteau (1967)

- une autre expérience, celle de Freedman et Fisher (1968), qui a montré que la capacité de localisation pouvait être conservée si l'on prolonge les conduits auditifs par des tubes cylindriques ; cependant cette capacité ne reste bonne qu'à condition d'adapter des pavillons artificiels à l'extrémité des tubes :

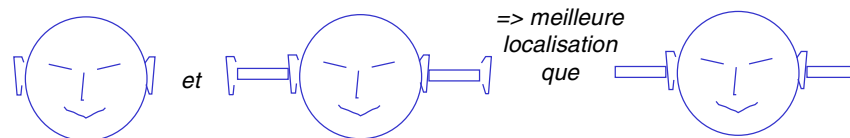


Figure : expérience de Freedman et Fisher (1968)

Notons enfin que pour utiliser les variations spectrales dues aux changements de direction incidente du signal il faut être en mesure de les distinguer de celles inhérentes au signal. C'est ce qui fait l'importance de la connaissance a priori de la source et de l'espace environnant.

On a montré en effet que la localisation exploite des connaissances sur la source et sur le local, dont l'acquisition se fait très rapidement à partir d'indices auditifs et visuels.

L'effet de précedence

Deux sons qui arrivent aux oreilles dans un temps assez bref (soit de 5 ms pour des clics à 40 ms pour des sons complexes) sont entendus comme un seul. C'est alors le premier qui détermine la perception spatiale.

C'est ce que l'on appelle l'effet de « précedence » ou « **loi du premier front d'onde** ».

Une conséquence de ce phénomène est notre capacité réduite à localiser des échos. Mais inversement, la présence d'échos multiples, dont l'énergie totale peut être supérieure à celle du son direct, affecte peu notre capacité à juger de direction incidente de celui-ci.

L'effet de précedence n'existe que pour les sons discontinus ou transitoires. La présence du deuxième son n'est cependant pas entièrement sans effet : elle peut introduire un biais sur la localisation du premier, déplaçant sa direction d'incidence apparente de jusqu'à 7°.

L'effet de précedence disparaît dans les cas suivants :

- si l'*ITD* entre les deux sons est inférieure à 1 ms ; on obtient alors une localisation « moyenne »,
- si l'intensité du deuxième son est relativement trop grande (de 10 à 15 dB supérieure),
- si les sons ne sont pas suffisamment similaires, tant du point de vue de l'enveloppe temporelle que du contenu spectral.

L'effet de précedence est un phénomène binaural, important pour la parole : la fusion d'un son et de ses échos aide à la **compréhension**. Pour s'en convaincre, on peut vérifier que dans une pièce réverbérante, se boucher une oreille est très handicapant : on entend moins bien, on localise moins bien, et la réverbération semble croître.

On peut également rapprocher de ces données l'article de Greenberg et Arai (1998), au titre explicite : Speech Intelligibility is Highly Tolerant of Cross-Channel Spectral Asynchrony.

L'effet de précedence a des conséquences essentielles en **HI-FI**, avec la **stéréo**. Dans celle-ci, la direction est codée surtout par différences d'intensité ; si on se rapproche trop d'un haut-parleur, (au point que l'*ITD* dépasse 1 ms), le son semble ne plus venir que de ce haut-parleur et l'effet de stéréophonie est perdu. Dans une pièce de dimensions moyennes, la conservation de la stéréo laisse une latitude de mouvements de l'ordre de 60 cm. Cela explique la notion de « siège stéréo ».

La latéralisation et l'équivalence temps-intensité

Une expérience simple, illustrée par les figures ci-dessous, permet de mesurer l'influence respective des *ITDs* et des *IIDs* sur la latéralisation (au casque) du son. Notons que dès que l'*ITD* dépasse 1 ms environ - resp. dès que l'*IID* dépasse 15-20 dB - le son est perçu exclusivement du côté où il arrive le plus tôt - resp. où il est le plus fort.



Figure : dispositif expérimental pour mesurer l'influence des *ITDs* et des *IIDs* sur la latéralisation

Dans les deux cas (retards créant une *ITD*, ou atténuateurs créant une *IID*) on observe des courbes similaires :

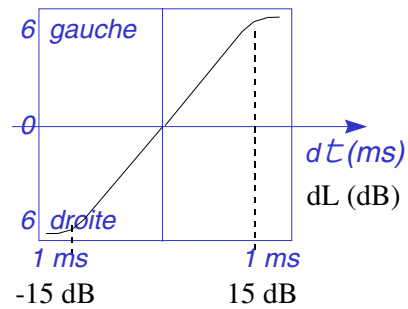


Figure : latéralisation fonction des ITDs et des IIDs

Par ailleurs il est possible de compenser un effet par l'autre. Par exemple, si une *ITD* entre deux hauts-parleurs crée la localisation subjective d'une source dans une direction donnée, une certaine *IID* affectant une intensité plus forte au son retardé peut permettre de contrebalancer ce premier effet, et de recentrer la source subjective.

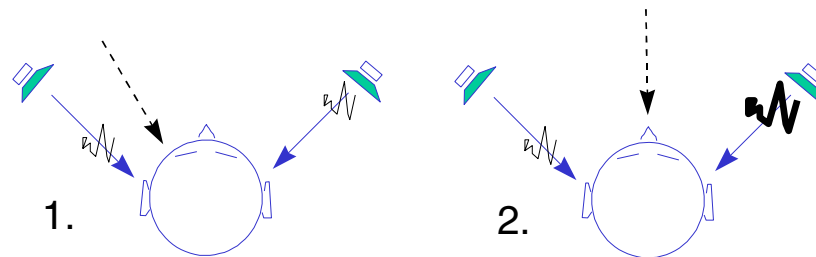


Figure : compensation d'une ITD par une IID

Cette expérience suggère que les différences interaurales de temps et d'intensité pourraient être codées de manière similaire dans le système nerveux central. Cela pourrait être expliqué par le fait que les temps de réponse aux sons intenses seraient plus courts ; l'information sur l'intensité serait de cette manière transformée en information temporelle.

Les mesures de cette transformation s'expriment en $\mu\text{s}/\text{dB}$, elles vont de $1,7 \mu\text{s}/\text{dB}$ (pour des sons purs) à $100 \mu\text{s}/\text{dB}$ (pour des trains d'impulsions).

Cependant l'équivalence temps-intensité n'est pas parfaite ; en effet, lorsqu'on équilibre ainsi artificiellement la localisation d'un son en compensant une *ITD* par une *IID* :

- la variabilité des jugements d'azimut est supérieure à celle observée avec des sons véritablement centrés,
- dans certains cas (sons purs basse fréquence, clicks) on obtient deux images sonores séparées : une « **image temporelle** » (peu affectée par l'intensité), et une « **image d'intensité** ».
- enfin, on reste souvent capables de distinguer les « vrais » sons **diotiques** (où les informations binaurales sont réellement issues d'une même source) des sons **dichotiques** (i.e. dont la localisation est ainsi équilibrée artificiellement).

Conclusions sur la localisation

Dans la localisation des sources le système nerveux emploie de nombreux indices, principalement :

- les *ITDs* et les *IIDs*,
- les changements de spectre dus à la tête et au pavillon,
- les changements des indices précédents dus aux mouvements de la tête ou de la source.

Dans les situations courantes, tous ces indices se confortent pour donner une image auditive cohérente.

2. Autres (in-)capacités binaurales

Quelques notions d'organisation perceptuelle

Comme dans tous les domaines perceptifs, la perception auditive ne se réduit pas à l'écoute de sons isolés mais met en jeu des structures syntaxiques. Il est utile d'avoir en tête les principaux mécanismes de groupement perceptifs, énoncés par les théoriciens de la Gestalt Psychologie comme Bregman (voir son ouvrage de référence : Auditory Scene Analysis : Perceptual organisation of sound, Bradford Books, 1994). Nous illustrons ces mécanismes par des exemples visuels :

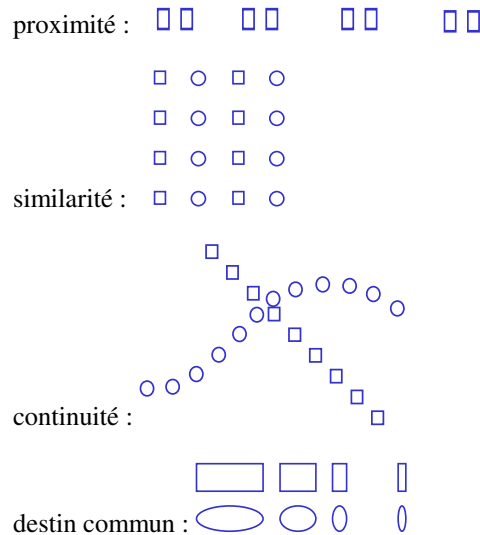


Figure : illustration des mécanismes de groupement perceptifs

Les battements binauraux

Lorsqu'on présente aux deux oreilles des fréquences légèrement différentes, on perçoit des battements binauraux qui sont une fluctuation du son, au rythme de la différence des fréquences binaurales.

Il ne faut pas confondre ces battements binauraux avec les battements habituellement entendus entre deux instruments de musique mal accordés, qui résultent d'une fluctuation purement physique de l'enveloppe d'amplitude d'une somme de sons purs de fréquences proches.

Les battements binauraux ne sont jamais aussi distincts que ces battements physiques. Ils résultent d'une interaction binaurale dans le système nerveux central. Ils sont une démonstration concrète de l'existence du « **verrouillage de phase** », ou synchronisation partielle des décharges du nerf auditif avec le rythme de l'excitation sonore.

Ces battements binauraux ne se produisent qu'en basses fréquences : les plus marqués entre 300 et 600 Hz, ils disparaissent au-dessus de 1000 Hz.

Lorsque la différence des fréquences présentées aux deux oreilles s'accroît, l'effet subjectif produit évolue ; on entend successivement :

- un changement de la localisation apparente du son,
- un changement de son intensité,
- un changement de timbre du son qui semble devenir « rugueux »,
- et enfin, deux sons subjectivement différents.

Différences binaurales de niveau de masquage (MLDs)

Les différences binaurales de niveau de masquage ou "**Masking Level Differences**" (MLDs) sont une autre illustration remarquable du bénéfice que nous retirons de l'audition binaurale. Deux expériences la mettent en évidence :

- la mesure des niveaux de masquage d'un son pur par un bruit, lorsque la phase relative du son pur par rapport au bruit varie sur une oreille, tandis qu'elle reste fixée sur l'autre oreille. La phase d'un signal, on le sait, est habituellement peu porteuse d'information pour l'audition ; cependant on constate que si les phases relatives d'un signal et d'un bruit masquant sont différentes entre les deux oreilles, le seuil de détection du signal est amélioré de jusqu'à 10-15 dB, par rapport à la condition où les déphasages binauraux sont identiques.

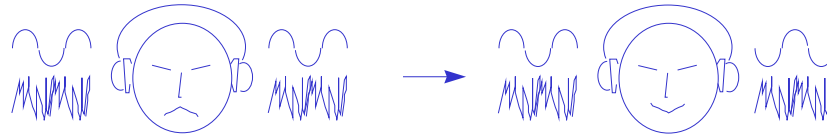


Figure : mesure des MLDs dues aux différences de déphasage signal/bruit entre les deux oreilles

- une autre expérience spectaculaire consiste à mesurer le seuil de détection d'un son pur dans un bruit masquant, dans deux conditions : (1) le son pur et le bruit sont présentés seulement à une oreille, et (2) le bruit est également présenté, seul, à l'autre oreille.

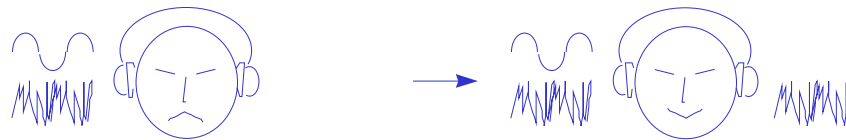


Figure : expérience sur les MLDs : signal plus bruit vs bruit seul

On constate que dans le second cas le seuil de détection du son pur est abaissé (i.e. meilleur), ce qui montre que le système nerveux central a su exploiter la forme connue du bruit pour extraire plus facilement le signal recherché, dans le canal où ils étaient mélangés.

Une conséquence pratique est par exemple qu'il ne vaut mieux pas se boucher l'oreille opposée au combiné lorsque l'on téléphone en milieu bruyant : dans la mesure où le bruit est présent, mélangé au signal, côté combiné, l'extraction du signal peut être plus efficace si l'information « bruit » reste disponible sur l'autre oreille.

De manière générale, notre perception dans le bruit est améliorée lorsque la différence de phase entre le signal et le bruit est différente entre les deux oreilles. C'est ce qui se passe en pratique chaque fois que les directions d'incidence du signal et du bruit sont différentes ; l'amélioration est maximale dans le cas d'une opposition de phase.

Ce phénomène joue ainsi un rôle important dans notre capacité à comprendre la parole en milieu bruyant, comme dans la situation de brouhaha couramment appelée « cocktail party ».

La paresse du système binaural

On constate que l'audition ne suit les mouvements d'une source que si ces mouvements sont assez lents : ainsi, l'angle de mouvement minimum audible ou "*Minimum Audible Movement Angle*" (*MAMA*) est de :

- 5° pour une vitesse de déplacement angulaire de 15°/s
- 21° pour une vitesse de déplacement angulaire de 90°/s.

Blauert (1972) a également mesuré notre capacité à suivre des mouvements d'*ITD* ou d'*IID* sur des trains d'impulsions. Pour un rythme de 80 impulsions par seconde, l'*ITD* ou l'*IID* subissant une modulation sinusoïdale, nous sommes capables de « suivre » auditivement le détail de la modulation pour :

un mouvement d'*ITD*, jusqu'à 2,4 Hz

un mouvement d'*IID*, jusqu'à 3,1 Hz

Au-delà, notre perception se brouille en une localisation moyenne, moins précise.

Ces expériences démontrent donc que le système binaural est très peu sensible aux mouvements rapides de la source.

L'influence de la vision sur la perception auditive

Les indices visuels peuvent déplacer (d'une manière qui se prolonge dans le temps, après la présentation de ces indices) la localisation des sources sonores.

Deux expériences amusantes illustrent la manière dont la vision peut ainsi influencer sur l'audition :

- le « pseudophone » expérimenté par Young en 1928, permettait de faire entendre à chaque oreille les sons provenant réellement du côté opposé ou d'une direction différente.

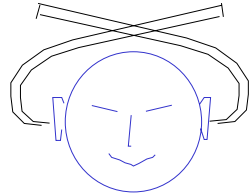


Figure : le pseudophone de Young

Les expérimentateurs notèrent que durant le port de l'appareil, le réapprentissage auditif était limité ; il n'y avait pas de réelle réorientation auditive.

Cependant, après avoir le port d'un pseudophone qui déplaçait en azimuth l'image auditive, on observait parfois un dédoublement de l'image auditive d'une source visible : une image auditive coïncidant avec la position réelle de la source, une autre étant légèrement décalée dans le sens opposé à celui créé auparavant par le pseudophone.

- dans une autre expérience due à Wallach (1940) le sujet est placé au centre d'un décor tournant, et il entend des sons provenant d'un haut-parleur fixe, placé derrière le décor à un azimuth variable.

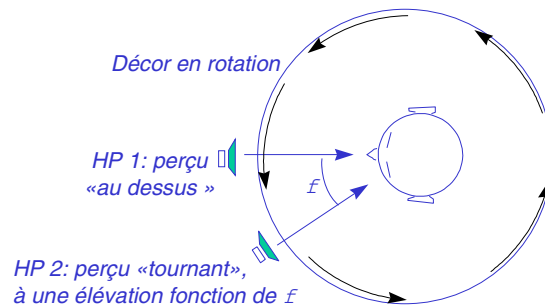


Figure : expérience de Wallach (1940)

Dans le cas où le haut-parleur était face au sujet, l'image auditive produite était fixe, à la verticale, comme si la source avait été au-dessus ou au-dessous du sujet. Dans le cas d'un azimuth non nul, la source était perçue comme tournante, à une élévation qui dépendait de cet azimuth.

Ces expériences montrent que la vision joue un rôle important dans la localisation, et participe à notre réorganisation de l'espace auditif.

Les capacités particulières des aveugles

On observe que les aveugles montrent une capacité remarquable à détecter des objets et juger de leur distance. Il a été avancé que des sensations cutanées (courants d'air...) jouaient un rôle dans cette capacité, mais il semble que ce rôle soit secondaire vis à vis de celui joué par l'audition.

Les aveugles se repèrent à partir de sons qu'ils produisent, par exemple en frappant le sol de leur canne, ou des sons de l'environnement.

Wilson (1967) a remarqué que la présence simultanée d'une source et de l'image de cette source créée par réflexion sur un obstacle, donne lieu à des interférences qui créent une série de minima et de maxima dans le spectre global ; si le spectre du son original est assez large, cette variation du spectre peut être perçue comme un son, le "*reflection tone*" dont la hauteur croît lorsque l'obstacle se rapproche.

Les aveugles ont ainsi la capacité de détecter la présence d'objets occupant un angle de seulement 3,5°, et de détecter des changements de distance de l'ordre de 20%.

Des expériences ont également mis en évidence une capacité à discriminer des matériaux (bois, velours...) et même des formes, par des aveugles qui utilisaient pour cela des sons qu'ils produisaient : sons vocaux, clics, sifflements...

3. L'étendue subjective des sources

Un événement auditif peut être très peu localisé. C'est le cas par exemple dans les églises où les réflexions multiples du son nous parviennent simultanément de toutes les directions.

Nous nous intéressons au cas simple de la stéréophonie, où un sujet reçoit simultanément les sons provenant de deux sources. Ce cas permet de restituer artificiellement une impression d'étendue.

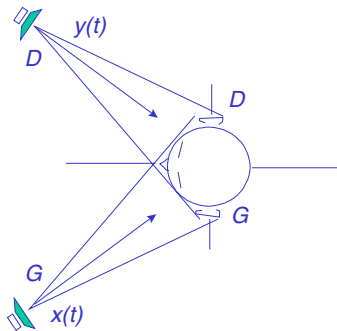


Figure : illustration d'un dispositif stéréophonique

Si $x(t)$ et $y(t)$ sont les signaux reçus par le sujet, en provenance des deux sources, on définit la fonction d'intercorrélation normalisée F par :

$$F_{xy}(t) = \frac{\int_{-T}^T x(t)y(t+t)dt}{\int_{-T}^T x^2(t)dt \int_{-T}^T y^2(t)dt}$$

On dit que $x(t)$ et $y(t)$ sont cohérents si : $(k =) \max_t |F_{xy}(t)| = 1$

C'est le cas évidemment lorsque $x(t) = y(t)$, le maximum étant alors atteint en $t = 0$. Dans le cas général, k varie entre 0 et 1, et gouverne l'étendue subjective de la source :

- pour $k = 1$, les signaux sont parfaitement corrélés, on entend alors un seul événement auditif dont la localisation varie avec t
- pour $k = 0$, à l'inverse, les signaux sont décorrélés, on entend alors deux événements auditifs distincts, localisés chacun sur un des haut-parleurs.
- on obtient une sensation équilibrée d'étendue pour $k \approx 0,4$.

Latéralisation

Par l'application judicieuse de filtrages (passe-haut, passe-bas...) aux deux signaux $x(t)$ et $y(t)$, on peut créer la valeur de k désirée. La figure ci-dessous illustre l'effet de différentes valeurs de k sur la latéralisation obtenue :

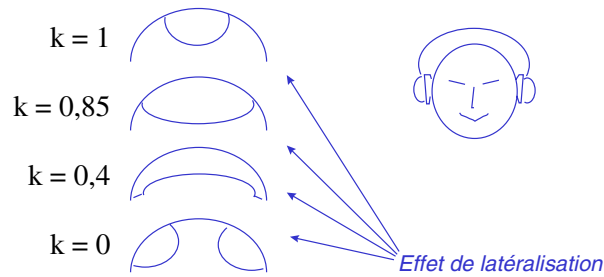


Figure : influence du maximum d'intercorrélacion normalisée sur la latéralisation

Echos et réverbération

Rappelons que les réflexions d'une source sonore sur les obstacles environnants, lorsqu'elles sont perçues distinctement, sont appelées des échos tandis que les réflexions suivantes se fusionnent habituellement pour former la réverbération.

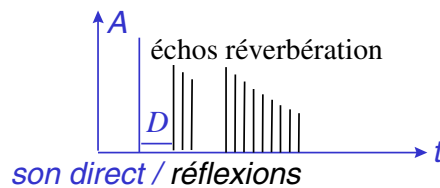


Figure : échos et réverbération

On observe que l'effet perceptif des premières réflexions est fonction du délai D qui sépare leur arrivée de celle du son direct :

- si D est inférieur à quelques ms : les réflexions contribuent à donner une impression d'espace,
- si D dépasse quelques ms: les réflexions sont perceptibles séparément et donnent une impression désagréable d'échos.

De manière générale, les réflexions individuellement marquées nuisent au confort acoustique. Un certain taux de réverbération est en revanche nécessaire à la création d'une sensation d'étendue sonore.

Ainsi, dans la création d'une salle de concert ou de spectacle, on pourra préférer une forme qui favorise les réflexions multiples - type « boîte à chaussures » - à une forme d'amphithéâtre qui favorise la diffusion du son dans une seule direction, et la perception uniquement du son direct, voire d'échos isolés. Cela est illustré par la figure ci-dessous.

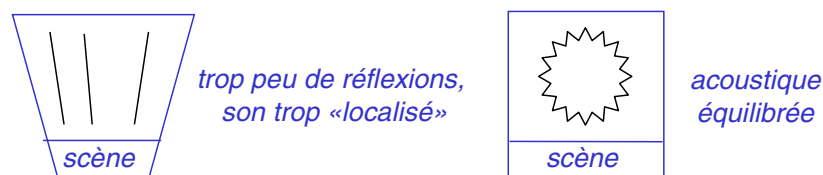


Figure : contrôle de la réverbération dans une salle de concert.

Application : simulation d'espace et stéréophonie

A partir de ce que nous avons vu ci-dessus, il est possible de créer artificiellement une impression d'espace à partir d'un enregistrement mono ou stéréo. Il convient de distinguer :

- l'effet de stéréophonie qui peut être créé à partir d'un signal monophonique : à l'aide de filtres et d'une réverbération artificielle (qui s'ajoute à la réverbération naturelle de la salle), on crée une sensation d'étendue de la source. Cette sensation est totalement artificielle.

- la « vraie » stéréophonie obtenue à partir d'un enregistrement à deux voies, qui peuvent éventuellement être artificiellement mixées et filtrées a posteriori. L'expérience montre que les microphones ne doivent être ni trop proches - ils captent alors la même chose - ni trop éloignés - la prise de son présente alors un « trou » au centre. La meilleure disposition est évidemment fonction de la salle, de la taille et de la position de l'orchestre... La disposition relative des deux micros fait aujourd'hui l'objet d'une norme, selon laquelle les micros doivent être espacés de 20 cm, leurs directions d'orientation formant entre elle un angle de 110°.

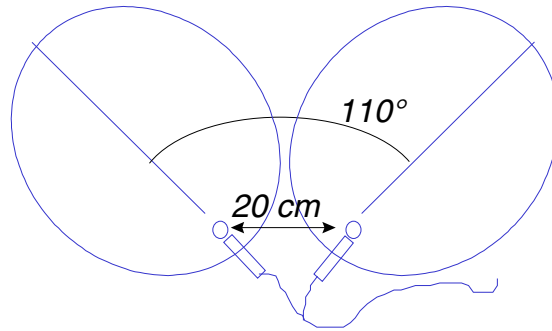


Figure : disposition normalisée des microphones pour la prise de son stéréophonique.

Annexe : Méthodes pour l'étude des sensations auditives

Exemple d'expérience

Mesure du seuil de détection d'un son pur dans le bruit par la méthode du choix forcé.

Dans cette expérience on utilise 2 types de stimuli :

- un bruit seul,
- le bruit additionné d'un son pur de fréquence et de niveau déterminés.

Les stimuli sont présentés alternativement au sujet, dans un ordre aléatoire; on fait varier l'intensité du son pur (et éventuellement sa fréquence).

A chaque présentation, on demande au sujet si le son pur est ou non présent. La réponse est obligatoire (oui ou non).

On construit la matrice stimulus / réponse:

	Réponse : s	b
Stimulus : s	Fausse alarme	Rejet correct
b	Détection correcte	Détection manquée

Dans le cas où le niveau du son pur est très faible, on a des réponses équiprobables : 50% de réponses justes.

A l'inverse s'il est très fort, on a 100% de bonnes réponses.

On trace la « courbe psychométrique » suivante:

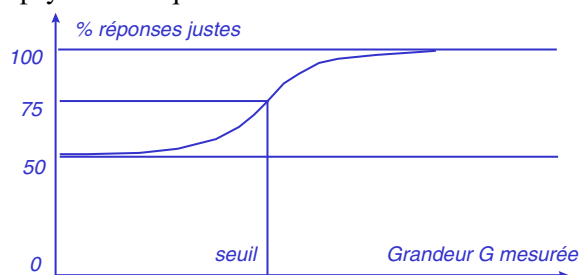


Figure : courbe psychométrique dans un test du type « deux alternatives à choix forcé »

On estime que le seuil de détection est le point correspondant à 75% de réponses justes.

(On vérifie le plus souvent que la courbe est monotone, sans quoi la définition du seuil peut être caduque. Un exemple typique de courbe non monotone est la courbe d'intelligibilité de la parole en fonction du niveau sonore chez certains malentendants : après avoir atteint son maximum, pour les niveaux forts, l'intelligibilité décroît).

La mesure de la pente de la courbe donne une indication sur la sélectivité du seuil.

Cette méthode est sûre, et permet dans certaines conditions de tester plusieurs sujets en même temps. Mais elle est longue et fournit une quantité de points inutiles. On lui préfère souvent en pratique des méthodes adaptatives, qui permettent une convergence plus rapide et sont bien adaptées au traitement informatique.

Méthodes de test

Celles-ci sont présentées plus ou moins par ordre croissant de difficulté :

- **Choix forcé** : cas précédent

- **Ajustement**

Dans ces méthodes le sujet règle lui-même le stimulus.

C'est applicable à la mesure de la perception de l'intensité (cas de l'audiomètre de Békésy) et de la hauteur.

- **Stimuli constants**, appréciation graduée

Exemple du test AAAB:

- on présente 2 paires de stimuli dont l'une peut comporter (ou non) un stimulus différent.

- on demande : 1. Y a-t-il un son différent?

2. Si oui, dans quelle paire?

3. Quel est le degré de différence? A apprécier entre 0 et 6 (c'est à dire 5 niveaux utiles, nombre optimal pour éviter la dilution de l'information tout en permettant le raccordement à des lois normales).

Ce type de test sert beaucoup en hi-fi.

Types de relations étudiées

(par ordre de difficulté de réponse):

Test	Type de réponse dans le cas du choix forcé
valeurs limites; seuils / seuils différentiels	présence du son / différence entre deux sons
valeurs de comparaison	le 2ie son est / n'est pas moins fort que le 1er
valeurs relatives	le 2ie son est / n'est pas <u>moitié</u> moins fort que le 1er

Traitement des résultats

Les indicateurs statistiques habituels (moyenne, variance, écart-type) ne sont pas invariants par transformation non-linéaire. Or dans beaucoup d'échelles de grandeurs subjectives apparaissent des transformations non-linéaires (comme le log dans le niveau *SPL*).

On préfère donc utiliser les indicateurs suivants, qui sont invariants:

- la **valeur centrale** : c'est le point frontière entre les moitiés supérieure et inférieure de l'ensemble des valeurs de mesure.

- l'**écart statistique**: zone dans laquelle se trouve un quart des valeurs mesurées.

Ces indicateurs sont déterminés par simple comptage.

Difficultés et contraintes

1. Dans tous ces tests, on doit faire attention au choix de la population qui peut biaiser énormément les résultats (par exemple, les musiciens s'attachent plus au contenu musical qu'à la qualité du son ; certains préfèrent le son synthétisé ; les collègues de laboratoire ont l'oreille plus entraînée ; etc.). On choisit souvent la population (contrairement aux sondages) en fonction de la question posée.
2. Il faut toujours tenir compte de la fatigabilité du sujet. Les performances sont altérées au cours d'un test long et difficile.