

Chapitre 6

La perception de la hauteur

Il existe une relation évidente à première vue entre fréquence et hauteur, mais la façon dont la fréquence est interprétée par le système auditif n'est pas si simple. Jusqu'aux années 60, les chercheurs ont défendu l'une ou l'autre des deux principales théories de la perception de la hauteur (localisation versus périodicité). Aujourd'hui, on reconnaît le rôle que jouent les processus de haut niveau (cognitifs) au niveau du cortex cérébral. Les deux théories n'en demeurent pas moins vraies, en ce sens que des mécanismes périphériques doivent être fonctionnels. Un premier problème, qui semble avoir préoccupé davantage les musiciens que les psychoacousticiens et les physiologues, apparaît au niveau de la définition même de la sensation de hauteur.

6.1 Hauteur spectrale et hauteur musicale

Depuis longtemps, les musiciens et psychologues de la musique ont essayé de définir les deux composantes de la sensation désignée par le nom général de hauteur. En fait, lorsqu'on fait évoluer la grandeur d'excitation « fréquence du stimulus », deux sensations évoluent parallèlement, qui ont reçu des noms très différents selon les chercheurs. En anglais, on désigne ces sensations par les termes *height* et *pitch*.

La première (*height*), en français **hauteur spectrale** (ou hauteur brute), correspond en fait à une composante du timbre et est reliée aux sensations de clarté, de densité et de volume¹ du son. Sur le plan acoustique, la hauteur spectrale est reliée à la perception du maximum d'énergie spectrale et est corrélée au **centre de gravité spectral**², une mesure proportionnelle à la somme des fréquences des partiels du spectre pondérées par leurs amplitudes.

La seconde (*pitch*), que l'on peut dénommer en français par **hauteur musicale**, serait la « véritable » sensation de hauteur. On sait que celle-ci n'a pas la même importance chez tous les auditeurs, et que certaines personnes y sont sourdes (*tone-deafness* ou atonie). Sur le plan acoustique, la hauteur musicale est corrélée à la **fréquence fondamentale** du son exprimée en Hertz.

Hauteur dans le grave → basses fréquences

Hauteur dans l'aigu → hautes fréquences

¹Le « volume » désigne ici la « grosseur » du son, et non l'intensité.

²Voir le chapitre portant sur le timbre pour plus de détail sur cette notion de centre de gravité spectral.

Certains chercheurs considèrent la hauteur musicale essentiellement apprise, alors que d'autres la considèrent aussi universelle que la sensation de sonie.

La hauteur musicale et la hauteur spectrale sont deux sensations qui cohabitent toujours, même dans le cas des sons purs. En effet, un son pur de basse fréquence est perçu comme étant grave, sombre et volumineux, alors qu'un son pur de haute fréquence est perçu comme étant aigu, mince et brillant. De plus, on peut généralement associer une hauteur musicale à un son pur (à 440 Hz, un son pur sera perçu sur la note La). Bien que ces deux sensations soient étroitement reliées, quatre critères permettent de distinguer la hauteur musicale de la hauteur spectrale :

1. Ce qui permet de juger la relation musicale entre deux sons, c'est-à-dire l'établissement de corrélations d'intervalles, alors que la hauteur spectrale ne permet que des comparaisons relatives. Ceci exclut les sons très graves (< 30 Hz), et très aigus (< 5000 Hz) pour lesquels la sensation de hauteur musicale se perd. Par ailleurs, les bruits ont des hauteurs relatives qu'on peut comparer, mais pas de hauteur musicale clairement définie (on parle parfois de hauteur indéterminée, pour les percussions par exemple).
2. Ce qui permet l'intonation vocale d'un son. Ce critère est intéressant et a été illustré par plusieurs expériences. Par exemple, lorsqu'on demande à un sujet de penser une note, des électrodes captant les vibrations de ses cordes vocales montrent que le sujet fait effectivement vibrer ses cordes vocales, même en dehors de toute forme d'émission sonore vocale.
3. Ce qui fait la ressemblance des octaves. La plupart des modèles qui veulent schématiser les deux aspects de la sensation de hauteur se fondent sur cette ressemblance, sur l'impression de recommencement qui apparaît lorsqu'un stimulus arrive à l'octave d'un autre. Ceci implique cependant le critère 4. Notons qu'un glissando continu du grave à l'aigu ne génère pas cette impression de recommencement. Même dans une gamme chromatique très régulière, il est fort possible que l'auditeur ne soit pas sensible au passage de l'octave. La ressemblance de deux sons à distance d'octave en audition simultanée semble cependant universelle.
4. Ce qui permet de relier un son à une note, c'est-à-dire à une des zones que constituent les 12 demi-tons de l'octave (dans la musique occidentale), même si la valeur physique (la fréquence) peut être très variable.

6.2 Tonie et chroma

D'autres auteurs proposent de raffiner davantage la définition de la hauteur, prise comme sensation distincte du timbre. On parlera dans ce cas de **hauteur fondamentale**. Celle-ci est généralement associée à la fréquence la plus basse d'un son harmonique, ou au maximum de déplacement de la membrane basilaire, mais ce n'est pas toujours le cas. En effet, on peut faire



couramment l'expérience du **phénomène de la fondamentale absente**, au téléphone par exemple, qui ne transmet les fréquences qu'entre 300 et 3400 Hz, ce qui ne nous empêche pourtant pas de percevoir la fréquence fondamentale de la voix. Nous reviendrons sur ce phénomène dans une section ultérieure de ce chapitre.

La hauteur fondamentale aurait ainsi elle-même deux composantes ou dimensions :

- la **tonie**, reliée à la fréquence et à la *différenciation grave-aigu*. Pour caractériser cette dimension, Stevens a établi l'échelle des Mels.
- la **chroma** (du grec, *couleur*), une dimension liée à la perception de la hauteur qui tendrait à souligner que les *différences significatives se font d'abord à l'intérieur de l'octave* et non sur une échelle étalée sur l'ensemble du spectre fréquentiel audible. La chroma serait une qualité selon laquelle deux sons purs dont le rapport de fréquence vaut deux (et qui forment donc un intervalle d'octave) sont similaires ou identiques. La chroma est donc une dimension circulaire.

Exemples : Do (C), Ré (D), Mi (E), Fa (F), Sol (G), La (A), Si (B)

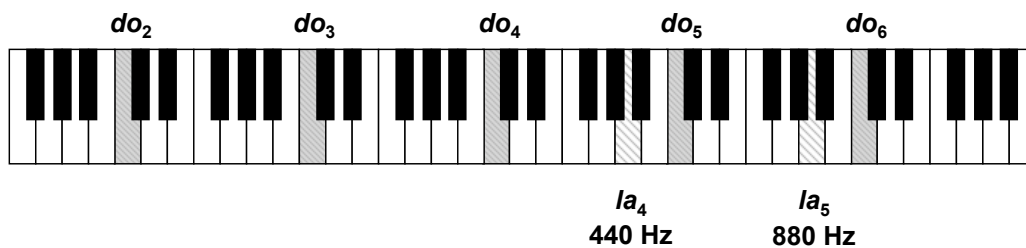
↑ ↑
 notation notation
 latine anglo-saxonne

En conclusion, on peut dire que la hauteur est cet attribut de la sensation auditive qui permet d'ordonner les sons en termes *de grave à aigu* (tonie), et/ou de les ordonner sur une *échelle musicale* (chroma).

6.3 Les échelles musicales

6.3.1 L'octave

L'intervalle de base des échelles musicales est l'*octave*. C'est le cas dans toutes les cultures musicales du monde. Deux notes séparées d'une octave ont entre elles un rapport de fréquences de 2:1. Dans la gamme tempérée occidentale, par exemple, le *la* au-dessus du *do* moyen au piano et le *la* suivant ont des fréquences respectives de 440 Hz et 880 Hz.



On distingue 12 classes de hauteurs (ou chromas), chaque classe correspondant au nom d'une note : *do, do#, ré, ré#, mi, fa, fa#, sol, sol#, la, la#, si*. Les notes séparées d'une octave portent le même nom et sont d'ailleurs toujours jugées très similaires (même par les non-musiciens). Certains vont même les confondre et juger que c'est le même son. C'est ce qu'on appelle la *confusion d'octave*.

6.3.2 Les types d'échelles musicales

Une échelle musicale est donc une succession de notes arrangées en ordre ascendant ou descendant. Voici quelques exemples d'échelles musicales :

- *pentatonique* (5 tons)
- *chromatique* (12 notes correspondant aux 12 demi-tons)
- *heptatonique diatonique majeure* appelé aussi mode *ionien*, constitué de 7 notes séparées par les intervalles ton/ton/demi-ton/ton/ton/ton/demi-ton. Pour la gamme de *do* majeur, ce sont toutes les touches blanches du piano. Les autres modes heptatoniques sont le mode *dorien* (mode de *ré*), le mode *phrygien* (mode de *mi*), le mode *lydien* (mode de *fa*), le mode *mixolydien* (mode de *sol*), etc.

Les gammes *diatoniques* sont les plus couramment utilisées dans les cultures occidentales. La gamme *pentatonique*, par exemple, est la base de la musique chinoise et on la rencontre aussi dans les musiques celtique et amérindienne.

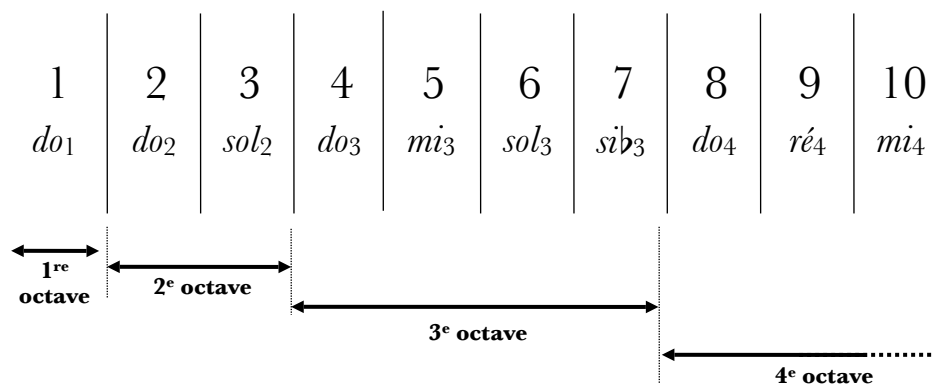
La plupart des cultures musicales utilisent des gammes logarithmiques, définies sur la base de rapports de fréquences égaux ($\Delta f/f = \text{constante}$). Mais quelques rares cultures (les indiens Nasca du Pérou par exemple) utilisent des gammes linéaires définies sur la base de différences de fréquences égales ($\Delta f = \text{constante}$).

Dans la plupart des échelles musicales, on retrouve l’octave (rapport 2:1) et très souvent aussi, la quinte juste (rapport 3:2). Ces intervalles sont qualifiés d’universaux par les ethnomusicologues. Notons que des bébés de quelques mois peuvent effectuer des tâches complexes de jugements de hauteur et de transposition. Il s’agit ici d’un argument fort pour rejeter l’idée que la hauteur musicale ne serait que le produit de l’acculturation.

Précisons également qu’il n’est pas nécessaire d’avoir une formation musicale pour percevoir la hauteur musicale. Même les non-musiciens peuvent faire des tâches musicales sophistiquées, surtout si on leur donne comme base des airs musicaux familiers, lesquels sont très bien encodés en mémoire.

6.3.3 L’échelle des physiciens

L’échelle des physiciens, aussi appelée l’échelle de Zarlino (1517-1590), est basée sur la série des harmoniques.



Rappelons qu’un son complexe, périodique et entretenu peut être analysé comme une *superposition de sons sinusoïdaux* dont les fréquences sont des *multiples entiers* de la fondamentale.

Les intervalles entre deux harmoniques successifs correspondent donc à des intervalles musicaux de plus en plus petits, depuis l’octave (2:1) jusqu’à la seconde mineure (16:15).

	tontondemi-tontontontondemi-ton							
Degré	do	ré	mi	fa	sol	la	si	do
Rapport à la tonique	1	9/8	5/4	4/3	3/2	5/3	15/8	2
	↑ tonique		↑ médiane		↑ dominante		↑ sensible	

La série des harmoniques a pu servir à expliquer l’origine “naturelle” de la gamme, des tonalités majeures et mineures, des sensations de consonance et de dissonance, des degrés de parenté entre les accords, etc. Par exemple, l’accord parfait majeur (*do-mi-sol*) est dérivé de la juxtaposition des harmoniques d’ordre 4, 5 et 6.

L’intérêt de cette échelle est qu’elle comprend des intervalles justes, et en particulier une tierce juste, définie par le rapport (5:4), ce qui n’est pas le cas de l’échelle de Pythagore (basée sur le cycle des quintes).

Mais l’échelle des physiciens présente tout de même un inconvénient : les intervalles *do-ré* et *ré-mi* sont différents.

- *do-ré* : rapport $\frac{9}{8} \Rightarrow$ c’est le ton dit “majeur”
- *ré-mi* : rapport $(\frac{5}{4}) \div (\frac{9}{8}) = \frac{10}{9} \Rightarrow$ c’est le ton dit “mineur”

La différence entre ces deux tons est le *comma zarlinien* (rapport 81/80). Par ailleurs, la quinte *ré-la* est plus petite que la quinte *do-sol* : $(\frac{5}{3}) \div (\frac{9}{8}) = \frac{40}{27} = 1,48$ au lieu de 1,5.

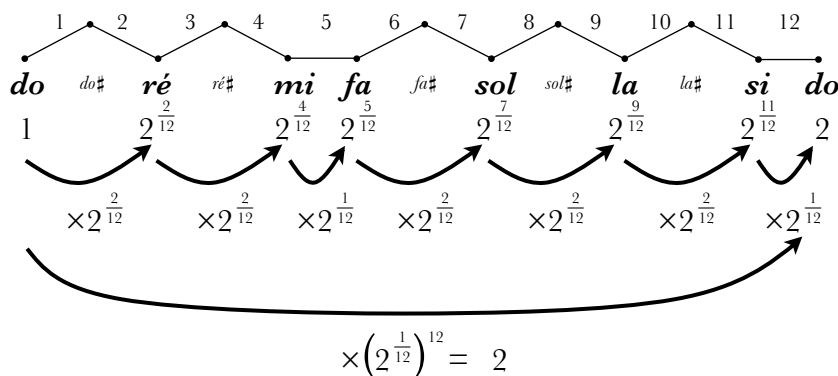
6.3.4 Le tempérament égal

Le compositeur français J.-P. Rameau (1730) a promu l’utilisation du tempérament égal pour les instruments à clavier. Par contre, c’est une erreur courante de penser que J. C. Bach voulait que les préludes et fugues de son *Clavier Bien Tempéré* soient interprétés dans un tempérament égal. Bach connaissait le tempérament égal mais n’y accordait pas de préférence particulière. Les musicologues pensent plutôt que le tempérament utilisé par Bach était un tempérament dit *irrégulier* minimisant les différences d’intervalles lors des transpositions.

Dans le cas du tempérament égal, l’octave est divisée en 12 intervalles logarithmiquement égaux. Chaque demi-ton correspond à la racine douzième de l’octave.

$$\text{demi-ton} = \sqrt[12]{\text{octave}} = \sqrt[12]{2} = 2^{\frac{1}{12}}$$

Pour monter d’un demi-ton dans la gamme, on multiplie la fréquence par $2^{\frac{1}{12}} \approx 1,059$. Par exemple, le *la*♯ qui suit immédiatement le *la* 440 Hz a une fréquence de 466 Hz (car $440 \text{ Hz} \times 1,059 \approx 466 \text{ Hz}$).



Après 12 sauts de demi-tons tempérés (rapport $2^{\frac{1}{12}}$), on retombe nécessairement sur l'octave puisque $(2^{\frac{1}{12}})^{12} = 2$.

Notons qu'avec les instruments à cordes frottées (violons) et les instruments à vent, l'instrumentiste peut varier la fréquence produite et s'écarter du tempérament égal. D'ailleurs, les sensibles (7^e note de la gamme) sont souvent jouées plus aiguës que les notes du clavier, amenant ainsi plus de tension musicale.

6.4 Une échelle de tonie : l'échelle Mel

Stevens a été le premier à établir une échelle de tonie³. À cause de la longue tradition de l'acoustique musicale relative aux échelles musicales, cette échelle a été assez contestée, souvent à cause d'une mauvaise compréhension de l'objet même de la mesure faite par cette échelle. Certains ont cru que Stevens cherchait à remplacer les rapports de fréquences traditionnellement acceptés pour la définition des intervalles musicaux, par de nouveaux rapports, ceux de l'échelle Mel⁴. Pour Stevens, *une tonie doublée ne correspond pas nécessairement à l'octave*, qui correspond pourtant à un doublement de fréquence ! Mais quand Stevens parle de tonie double, il ne prétend pas définir l'octave car en réalité, il restreint cette échelle à la première dimension de la hauteur fondamentale, relative à la différenciation grave-aigu, indépendamment de la seconde dimension qu'est la chroma. L'échelle Mel a donc pour but de mettre en évidence la « position » des perceptions à l'intérieur du continuum qui va de l'extrême grave à l'extrême aigu, en dehors de toute référence avec les intervalles musicaux. Cette échelle est d'ailleurs tout à fait en accord avec ce qui a été découvert sur la subdivision de ce continuum en bandes critiques, et sur le mode de vibration de la membrane basilaire en fonction de la fréquence. Précisons que l'unité « mel » vient de « mélodie », terme qui après coup est plutôt mal choisi vu les résultats de l'étude.

L'échelle Mel n'a pas fait l'objet de consensus international comme l'échelle de sonie. Stevens utilisait comme première référence 1000 mels pour 1000 Hz. Zwicker utilise une référence de 310 mels à 310 Hz. Dans la mesure où l'échelle n'a pas beaucoup d'utilisations pratiques et sert surtout à démontrer une relation subjective, ces différences ne sont pas importantes.

³Stevens, S.S. & Volkman, J. (1940) The relation of pitch to frequency : a revised scale. *American Journal of Psychology*, 53(3), 329-353.

⁴Pederson, P. (1965) The mel scale. *Journal of Music Theory*, 9(2), 295-308.

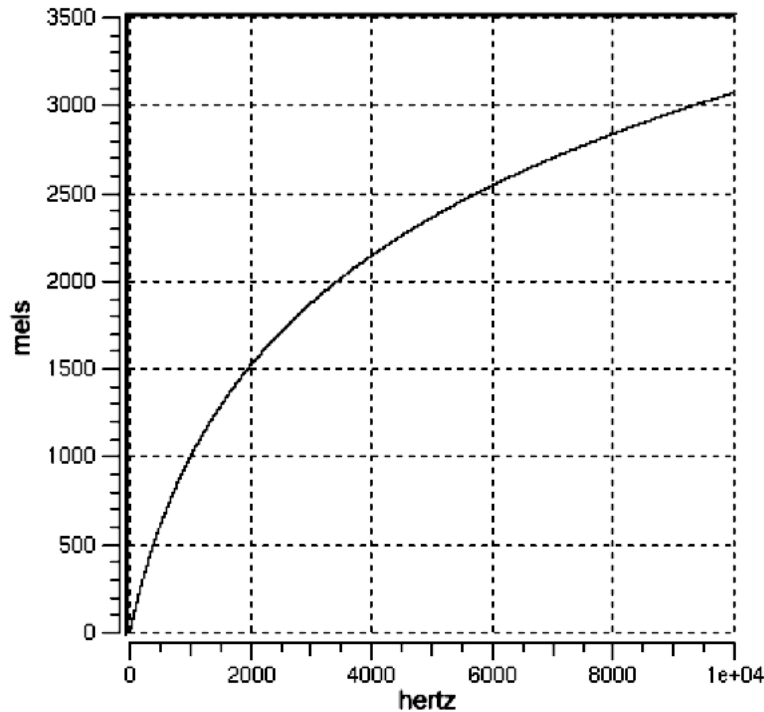
La plupart des échelles obtenues avec des sons purs indiquent une perception linéaire de la fréquence jusque vers 500 Hz, en fonction de la fréquence : un rapport de fréquences de 2, (l'octave) produit approximativement une sensation de tonie doublée. Mais au-dessus de 500 Hz, la relation fréquence-tonie perd sa linéarité : pour que la tonie soit doublée, le rapport des fréquences doit être plus grand que l'octave.

L'échelle Mel est construite de la même façon que l'échelle des sones, par mesurage direct : on demande à des participants d'ajuster un son de façon à ce qu'il soit, par exemple, deux fois moins aigu qu'un son de référence (1000 mels), ou deux fois plus aigu, etc.

Suivant la convention de Stevens (à 1000 Hz, un son pur d'intensité 40 dB a une tonie de 1000 mels), une formule proposée par O'Shaughnessy⁵ établit la relation entre l'échelle de tonie en mels et la fréquence en Hz :

$$n_{\text{Mel}} = 2595 \times \log \left(\frac{f}{700} + 1 \right)$$

Fréquence	Tonie
100 Hz	150 mels
200 Hz	283 mels
1000 Hz	1000 mels
2000 Hz	1521 mels
3000 Hz	1876 mels



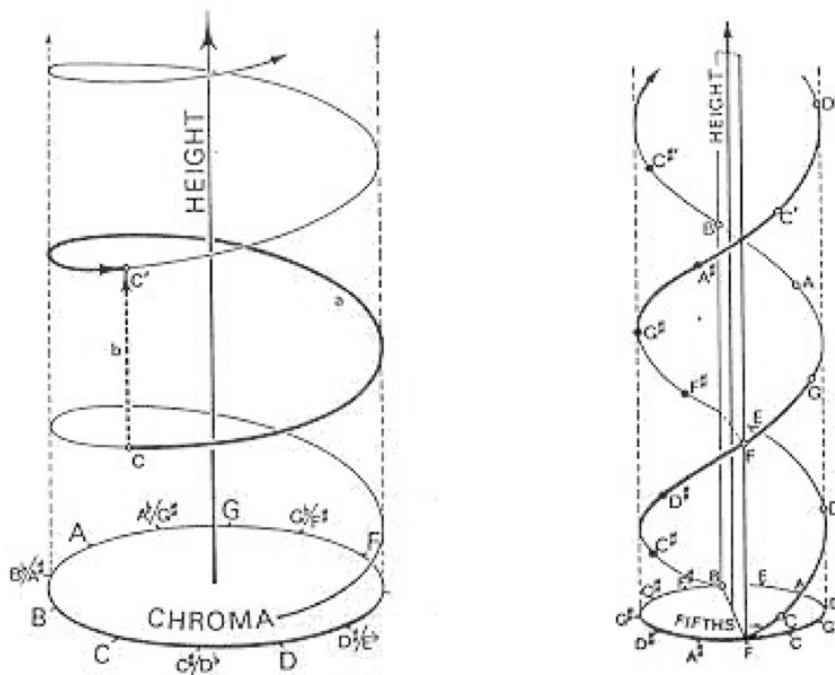
⁵O'Shaughnessy, D. (1978) Speech communication : Human and machine. Addison-Wesley, New York.

L'échelle de tonie rend compte du fait que hauteur perçue croît moins vite que la fréquence. Quand on passe de 1000 Hz à 2000 Hz, on double la fréquence, mais la tonie passe de 1000 mels à 1521 mels seulement (rapport 3:2). Et autour de 3000 Hz, un rapport de 2:1 en termes d'unités psychologiques (mels) correspond à rapport d'environ 3:1 en termes d'unités physiques (Hz). On comprend donc qu'une octave dans les hautes fréquences (autour de 2000 Hz par exemple) va être perçue plus grande qu'une octave dans les basses fréquences (autour de 250 Hz par exemple).

Il n'y a pas de contradiction entre les deux types de perceptions – tonie et chroma – qui cohabitent sans difficulté, l'intervalle musical relevant fort probablement davantage d'une perception de forme que d'une perception de dimension. Comme un carré peut être grand ou petit, l'intervalle musical peut être grand ou petit, selon le registre où il est utilisé, sans perdre sa qualité propre.

6.5 La circularité de la perception de la hauteur

Comme nous l'avons vu précédemment, la hauteur tonale (ou fondamentale) possède la dimension « différenciation grave-aigu » – la tonie – et la dimension « position de la note dans une octave » – la chroma. Le psychologue expérimental *Roger Shepard* a proposé que l'on peut déformer l'échelle habituellement rectiligne des hauteurs en une simple hélice faisant un tour complet sur une octave. La dimension « grave-aigu » est représentée par la dimension verticale, et les notes de la gamme sont représentées par des positions distinctes sur le *cercle des chromas*. Toutes les fréquences correspondant à la même chroma (tous les *la* par exemple) seront projetés sur le même point d'un cercle.



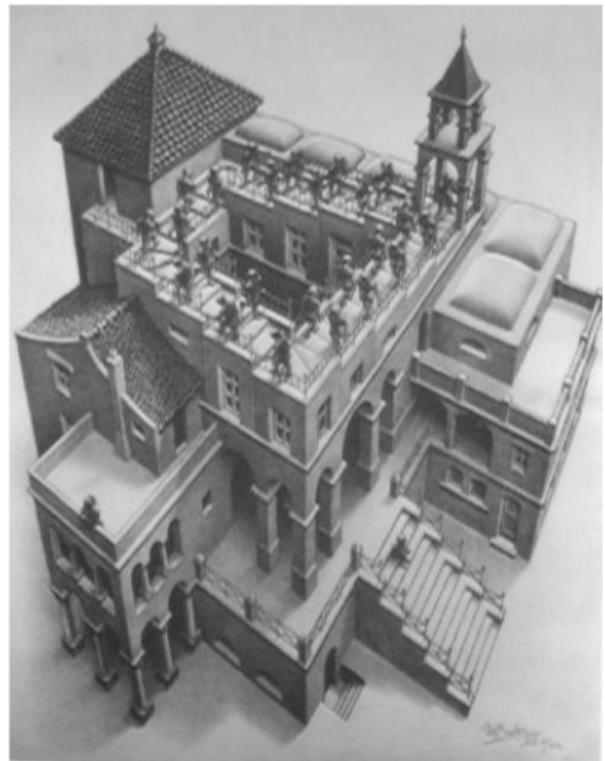
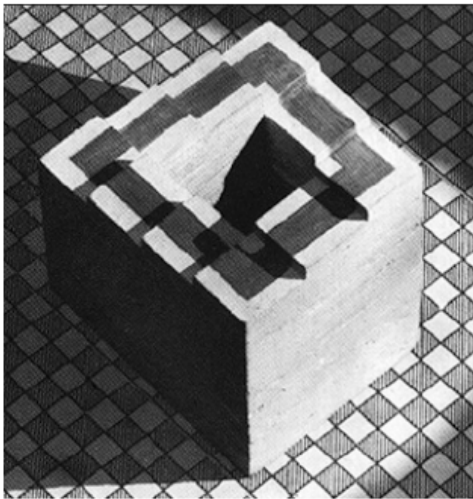
Sur cette structure en hélice, un accroissement de similarité est représenté par une distance de séparation réduite (par exemple, une octaviation, de $C \rightarrow C'$ sur la figure).

Roger Shepard propose une seconde structure tridimensionnelle en *double hélice* combinant l'échelle rectiligne des hauteurs et le cycle des quintes (*cf.* échelle de Pythagore). Cette structure a les propriétés suivantes :

- les tons inclus dans une tonalité majeure particulière sont séparés des tons qui ne le sont pas par un plan vertical passant par l'axe de la double hélice. Par exemple, le plan passant par B-F (*si-fa*) sépare les tons présents dans *do* majeur et ceux qui ne le sont pas.
- la transposition dans les tonalités les plus proches sont obtenues par les plus petites rotations du plan vertical. Par exemple, les plans passant par F (*fa*) et G (*sol*) sont proches du plan passant par C (*do*).

Une illusion auditive de Roger Shepard : les *gammes sans fin*

Produit de la synthèse additive, ces sons donnent l'illusion d'être constamment “montant” ou “descendant”. Ce sont les équivalents auditifs des peintures d'escaliers de M.C. Escher.

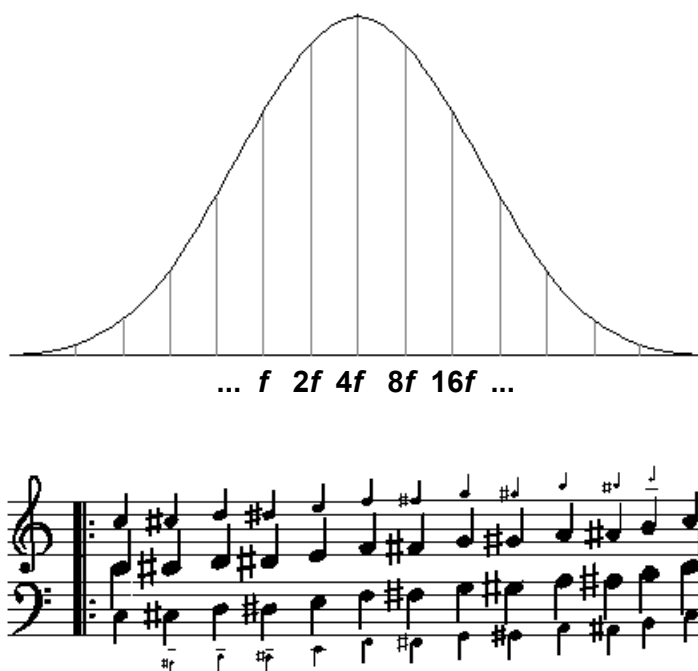


Le psychologue expérimental *Roger Shepard* (Stanford University) les a mis au point dans les années 60, avec les compositeurs James Tenney et Jean-Claude Risset (qui a aussi conçu et réalisé la version continue de cette illusion, sous la forme de *glissandi sans fin*).

L'illusion auditive de Roger Shepard illustre la dimension indépendante de chroma dans la perception des hauteurs et confirme la *circularité de jugement de hauteur relative* (impression de recommencement lorsqu'un stimulus arrive à l'octave d'un autre). Il faut noter que l'effet de

circularité n'est pas perçue dans le cas du glissando sans fin où les changements de hauteur sont continus.

Pour réaliser ces sons, il faut composer un son complexe par synthèse additive à partir de *partiels séparés par des octaves*. Les fréquences composant le spectre de ces sons sont donc proportionnelles à $f, 2f, 4f, 8f, 16f$, etc. C'est comme si on enlevait d'un son complexe harmonique toutes les fréquences harmoniques qui ne sont pas des puissances de 2. Comme illustré sur la figure ci-dessous, l'amplitude des composantes spectrales doit suivre une *enveloppe spectrale en forme de cloche*. Pour produire un son qui semble "monter" indéfiniment, toutes les composantes spectrales montent de demi-tons en demi-tons, simultanément. Elles s'atténuent progressivement et disparaissent dans les hautes fréquences et réapparaissent progressivement dans les basses fréquences.



Cette démonstration sonore permet de faire *varier la chroma sans faire varier la tonie* : la dimension de tonie est dégénérée grâce à la combinaison de toutes les octaves et la fréquence fondamentale apparente augmente sans fin, alors que les tons décrivent le *cercle des chromas*.

Démo sur le web : <http://www.cs.ubc.ca/nest/imager/contributions/flinn/illusions/illusions.html>

Notons qu'il est possible de produire l'effet inverse, en augmentant la hauteur spectrale (de grave à aigu) tout en maintenant la chroma constante (on reste sur une même note, d'octave à octave).

Par ailleurs, bien que la notion de chroma implique que les différentes notes de même nom (tous les Do par exemple) sont perceptuellement similaires, la transposition des notes d'une mélodie dans des octaves aléatoires peut altérer le contour mélodique à ce point que la mélodie devient difficilement reconnaissable.

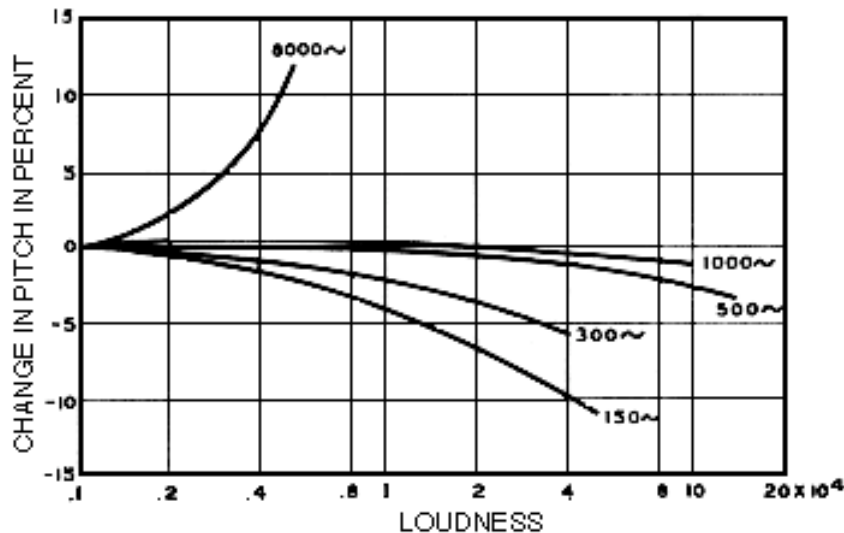
6.6 L'oreille absolue

L'oreille absolue est cette capacité que possèdent certaines personnes d'identifier une note musicale sans note de référence. La majorité d'entre nous avons l'oreille dite relative, c'est-à-dire que nous devons entendre une note de référence afin de chanter un intervalle. Les personnes avec l'oreille absolue ont encodé la fréquence des notes (par exemple, le *la* 440 Hz), et peuvent donc reconnaître les notes de manière directe. Fait à souligner : il semblerait que les possesseurs d'oreille absolue font plus souvent des erreurs de confusion d'octaves que les possesseurs d'oreille relative, comme si la sensation de chroma s'était développée au détriment de la sensation de tonie.

6.7 Facteurs affectant la tonie des sons purs

6.7.1 Influence de l'intensité sur la hauteur perçue

La hauteur est perçue à partir de 2 ou 3 dB au-dessus du seuil d'audition. Elle varie aussi en fonction du niveau. Les premiers expérimentateurs ont rapporté une influence substantielle de l'intensité sur la hauteur perçue à 90 dB. Il semble maintenant que l'effet est petit et varie d'un sujet à l'autre. En utilisant des sons de longue durée, Stevens (1935) a trouvé que les sons en-dessous de 1000 Hz décroissent en hauteur apparente avec une augmentation d'intensité alors que les sons au-dessus de 2000 Hz augmentent en hauteur apparente avec une augmentation d'intensité. Le graphique ci-dessous représente le changement de tonie en pourcentage de la fréquence présentée en fonction de l'intensité, et ce pour différentes fréquences.



6.7.2 Influence de la durée sur la hauteur perçue

Pour percevoir la hauteur, le son pur doit avoir une durée minimale. Cette durée varie d'une fréquence à l'autre. Von Békésy a trouvé qu'un son avec une fréquence de moins de 1000 Hz devait avoir une durée minimale équivalente à environ 3 à 9 périodes pour avoir une hauteur

définie. Par exemple, à 300 Hz, une période correspond à $1/300 \text{ Hz} = 0,0033 \text{ s} = 3,3 \text{ ms}$ et il faudra entre 16,5 et 29,7 ms pour percevoir la hauteur de façon définie. Au-dessus de 1000 Hz, cette valeur critique de durée est de 10 ms indépendamment de la fréquence du son.

6.7.3 Influence du masquage

Un son masquant sur la même oreille peut changer la hauteur d'un son pur. Si le son masquant est un bruit blanc, la hauteur va augmenter. Si le son masquant est un son pur de fréquence inférieure, la hauteur va augmenter également.

6.8 La mécanismes de perception de la hauteur

6.8.1 La perception de la hauteur des sons complexes

Comme nous l'avons vu précédemment, lorsqu'un son ne contient qu'une seule fréquence (son pur), sa hauteur est déterminée en grande partie par sa fréquence. Mais dans la nature, il n'existe pas de sons purs. Les sons vocaux, instrumentaux et environnementaux sont composés de plusieurs fréquences, ayant chacune leur amplitude propre. Or, dans le cas des sons harmoniques, nous ne percevons pas plusieurs, mais une seule hauteur, comme s'il n'y avait qu'une seule fréquence. Généralement, la hauteur d'un son complexe correspond à la hauteur de sa fréquence fondamentale. Ainsi, pour un son de violon composé des fréquences harmoniques 440 Hz, 880 Hz, 1320 Hz, 1760 Hz, etc., la hauteur de ce son correspondrait à 440 Hz. Autrement dit, un son pur jugé comme ayant la même hauteur que ce son complexe aurait une fréquence de 440 Hz.

Mais que se passe-t-il si la fondamentale est absente ? Imaginons que l'on présente un son complexe harmonique de fréquence fondamentale égale à 200 Hz et constitué de 10 harmoniques d'amplitudes égales. Dans un premier temps, toutes les harmoniques sont présentes. Ensuite, le son est présenté sans la fondamentale, puis sans les deux premières harmoniques, etc. À l'audition, on remarquera un changement de timbre, mais pas de changement de hauteur fondamentale même si une bande de bruit à basse fréquence est prévue pour masquer un son de 200 Hz qui pourrait être généré par distorsion.

Ce phénomène nous pousse à nous interroger sur les mécanismes d'encodage de la hauteur par le système auditif.

Jusque dans les années 1960, deux théories antagonistes ont tenté d'expliquer la perception de la hauteur. Ce sont la **théorie de la localisation** (*place theory*), et la **théorie de la périodicité** appelée également la **théorie temporelle** (*temporal theory*). Ces théories postulent l'intervention d'un mécanisme au niveau de l'oreille interne.

Aujourd'hui, on fait appel à des *mécanismes* plus *centraux*, c'est-à-dire faisant intervenir le cerveau (cortex cérébral) pour expliquer notre perception de la hauteur, et donc faisant appel à des processus de haut niveau (cognitifs). Les deux théories n'en demeurent pas moins vraies, en ce sens que des mécanismes périphériques (dans l'oreille) doivent être fonctionnels.

Rappelons que le système auditif est organisé tonotopiquement depuis l'oreille interne jusqu'au *cortex auditif primaire* (gyri de Heschl), c'est-à-dire que l'information pour chaque fréquence (nombre de cycles de l'onde par seconde) est représentée de façon systématique tout le long des voies auditives. Ainsi, au cortex auditif primaire, la représentation pour les fréquences allant de

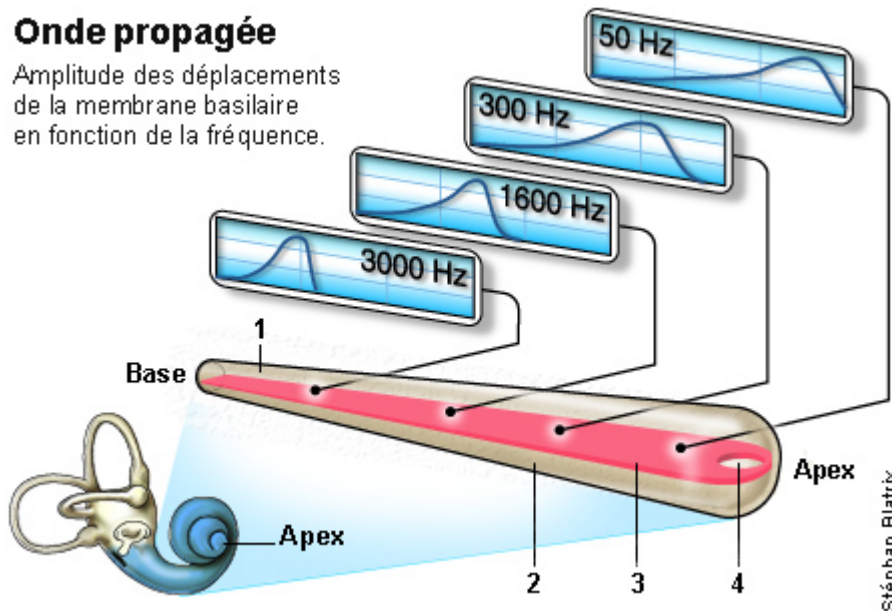
basses à aiguës suit une trajectoire ordonnée allant de la face externe à la face interne du cortex primaire.

6.8.2 La théorie de la localisation (*Place theory*)

L'idée vient d'abord de *Helmholtz* (en 1863) qui se représentait la membrane basilaire comme formée de cordes tendues, accordées aux différentes fréquences, du grave à l'aigu et entrant en résonance avec le signal acoustique.

À partir de 1930, *von Békésy* développa la *théorie de l'onde propagée* qui lui valut le prix Nobel en 1961 : une onde progressive remonte le long de la cochlée et entraîne la membrane basilaire qui subit une déformation maximale en un point qui varie selon la fréquence.

La théorie de la localisation se baserait donc sur le fait que le son complexe subit une analyse spectrale au niveau de la membrane basilaire, qui est excitée maximale en différents endroits correspondant aux différentes composantes du son. Les neurones dont les terminaisons nerveuses se trouvent au point de vibration maximale transmettraient l'information relative à la hauteur. Békésy explique ainsi la *tonotopie passive*.



Phénomènes qui appuient la théorie de la localisation

- la diplacousie (différence de perception des hauteurs entre l'oreille gauche et l'oreille droite)
- l'influence de l'intensité sur la hauteur

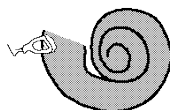
La **diplacousie** est un phénomène qu'on peut observer chez la plupart des auditeurs, mais dans des proportions fort différentes. Il se manifeste par une perception de hauteur différente, pour une même fréquence, selon qu'elle se présente à une ou l'autre des deux oreilles. Les différences anatomiques entre les deux oreilles, et en particulier entre leurs membranes basilaires, pourraient expliquer cette différence. Précisons que lorsque le son est entendu simultanément par les deux oreilles, il est fusionné en une seule hauteur.

Par ailleurs, on peut observer facilement **l'influence de l'intensité sur la hauteur** : lorsqu'un son pur augmente en intensité, sa hauteur change, la plupart du temps vers le grave. Quoiqu'il n'y a pas de raison évidente que le lieu de vibration maximale de la membrane basilaire soit déplacé, la plus grande intensité force une plus grande plage de la membrane basilaire à vibrer. Plus de neurones se trouvent excités. Ceci irait dans le sens de la théorie proposée notamment par Zwicker selon laquelle la fréquence n'est pas déterminée par le lieu de déplacement maximal, mais par la région limite entre la zone qui se déplace et celle qui ne se déplace pas, autrement dit par les flancs, et leurs différentes pentes, de la courbe représentant les «niveaux d'excitation» de chacun des points de la membrane basilaire.

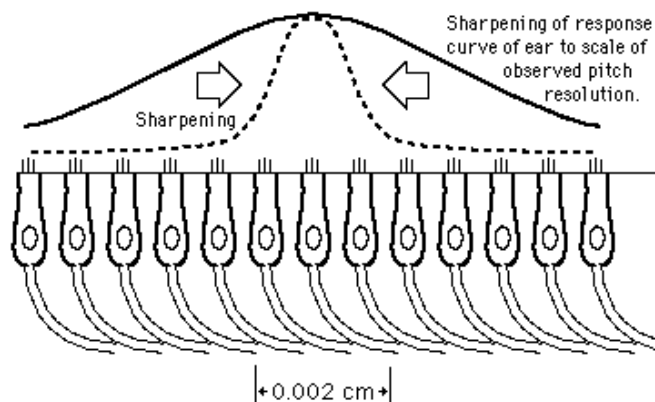
Phénomènes qui vont à l'encontre de la théorie de la localisation

- Le pouvoir différentiel
- Le pouvoir séparateur
- Le phénomène de la fondamentale absente

La taille extrêmement petite de la cochlée laisse à penser que la théorie de la localisation n'est pas suffisante pour expliquer **le pouvoir différentiel**, autrement dit, la résolution extrêmement fine de la perception des hauteurs. La cochlée est un tout petit organe de 3,2 cm de long, sur 2 tours et 3/4, peut distinguer 1 500 hauteurs avec entre 16 000 et 20 000 cellules ciliées. Cela signifie qu'il y aurait une hauteur distincte pour chaque $32 \text{ mm} / 1\,500 = 0,02 \text{ mm}$ ($20 \mu\text{m}$), ce qui n'est physiquement pas possible.



Pour concilier la théorie de la localisation avec la résolution fréquentielle fine du système auditif, les chercheurs ont proposé l'existence d'un **mécanisme d'affinage de la détection des hauteurs** basé sur l'*inhibition latérale de la membrane basilaire*. Les cellules ciliées adjacentes au maximum de vibration seraient inhibées (plus précisément, le système de recharge pourrait être inhibé). L'inhibition pourrait aussi avoir lieu au niveau du ganglion spiral, ne laissant passer que les influx nerveux envoyés à plus hautes fréquences. Ce système serait contrôlé par le cerveau. De tels phénomènes d'inhibition sont fréquemment observés dans différents domaines de perception.



La différence juste audible pour la tonie est à peu près de 5 cents. D'après Rossing, cette "jnd" (*just noticeable difference*) dépend de la *fréquence*, de l'*intensité*, de la *durée* du son, de la *formation* musicale du sujet et de la *méthode* de mesure. Le "jnd" pourrait être évaluée à 1/30^e de la bande critique.

Le **pouvoir séparateur** est la capacité de distinguer deux sons de fréquences différentes présentés *simultanément*. Pour des sons rapprochés mais perceptuellement séparés (pas de battement), la membrane basilaire ne présente qu'un seul lieu de vibration maximale très large. Alors, comment s'explique la séparation des sensations ?

En réalité, le taux de décharge d'un neurone (le nombre de potentiel d'action par seconde) à un niveau d'intensité donné est maximal pour une certaine fréquence, appelée *fréquence caractéristique* du neurone. Il existe donc une relation entre la *localisation d'une fibre nerveuse* et sa *fréquence caractéristique* (FC). Selon leur emplacement, les neurones ont des fréquences caractéristiques différentes. Les neurones qui sont situés près de la base ont des FC élevées, donc sont sensibles aux fréquences élevées, et les neurones situés près de l'apex ont des FC basses, donc sont sensibles maximalement aux fréquences basses. Selon cette théorie, *la hauteur du son est donc reliée au patron d'excitation du son*. Dans le cas d'un son pur, la hauteur correspond à l'endroit où se produit le maximum d'excitation.

Le **phénomène de la fondamentale absente** fut l'argument le plus fort contre la théorie de la localisation. Dans les sons complexes, le maximum de vibration de la membrane basilaire se produit en effet souvent en un point fort éloigné de celui correspondant à la fondamentale. Celle-ci est souvent même totalement absente. Le timbre s'en trouve modifié, mais pas la hauteur perçue. De même, le téléphone ne transmet que les fréquences comprises entre 300 et 3400 Hz alors que la fréquence fondamentale de la voix humaine (fréquence à laquelle les cordes vocales vibrent) est généralement inférieure à la fréquence de coupure du téléphone à 300 Hz.

Homme adulte : entre 85-155 Hz	Enfant (10 ans) : entre 208-259 Hz
Femme adulte : entre 165-255 Hz	Bébé (12 mois) : entre 247-410 Hz

Et pourtant, nous ne percevons pas la voix plus aiguë qu'elle ne l'est.

La hauteur perçue ne correspond donc pas toujours à la fréquence la plus basse du spectre. Quand la fondamentale est absente, on entend un *son résiduel* qui n'est pas présent dans le signal physique. Par *son résiduel*, on désigne ainsi la perception d'un son généré lorsque deux ou plusieurs sons, appelés les *primaires*, ayant entre eux une relation harmonique, c'est-à-dire un diviseur commun, sont présentés simultanément. La fréquence perçue correspond donc au *plus grand diviseur commun* des composantes primaires ayant entre elles une relation harmonique. Autrement dit, c'est la plus haute fréquence pour laquelle les sons pourraient être reliés harmoniquement.

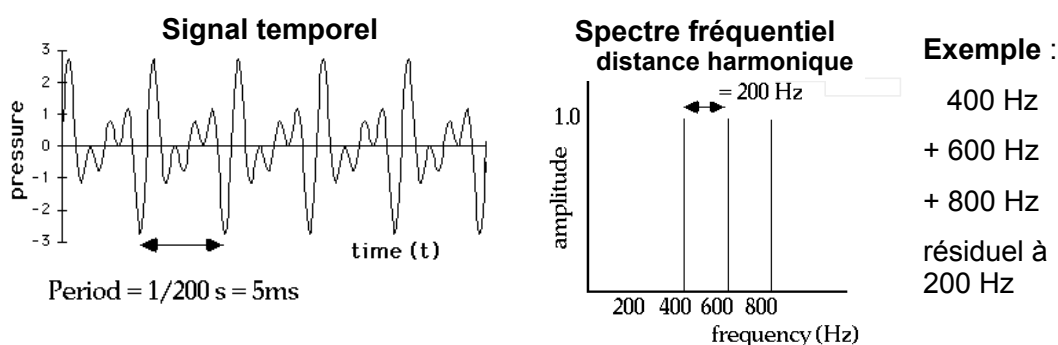
La perception de la fondamentale absente (son résiduel) dépend de certaines conditions :

- Il faut au moins 3 harmoniques consécutives. Plus il y en a, plus il est facile d'extraire la fondamentale.
- Il faut que le rang des harmoniques soit plus petit que 20 (les 5-6 premières étant les plus importantes).
- Il faut au moins deux harmoniques en-dessous de 5000 Hz.

Ainsi, un son composé des fréquences 400, 600 et 800 Hz aura une hauteur perçue de 200 Hz. Même s'il n'y a pas d'énergie physique à 200 Hz, la hauteur perçue correspond à la fréquence qui est la fondamentale théorique du son. Comme l'illustre la figure ci-dessous, le signal temporel est périodique et la période correspond à la fréquence fondamentale absente :

$$5 \text{ ms} = 0,005 \text{ s} = \frac{1}{200 \text{ Hz}}$$

On en déduit que l'information sur la fréquence fondamentale est donc absente dans le domaine fréquentiel mais elle est présente dans le domaine temporel !



Comme autre application à la vie courante de cette particularité de la perception des hauteurs, on peut citer le cas de l'écoute sur de petits haut-parleurs qui se comportent comme des filtres passe-haut, en coupant les basses fréquences du son. Malgré l'information manquante au niveau du spectre, le système auditif est capable de percevoir la fréquence fondamentale du son.

Quand on a présenté le phénomène de la fondamentale absente comme un argument contre la théorie de la localisation, les partisans de cette théorie se sont défendus en proposant que la fréquence fondamentale est *physiquement recréée* dans la cochlée par des *distorsions non-linéaires*.

Mais il semble que cette explication ne soit pas satisfaisante car, si la fondamentale était recréée, ou le résiduel généré, par distorsion :

- il ne pourrait pas y avoir de telle distorsion lorsque les sons sont inférieurs à 50-60 dB, or le son résiduel apparaît bien pour des primaires de faible intensité ;
- on devrait pouvoir *éliminer* le résiduel par l'envoi simultané d'un son de même fréquence mais en opposition de phase ;
- le son résiduel devrait pouvoir être *masqué* par une bande de bruit centrée sur sa fréquence ;
- enfin, en présence d'un autre son de fréquence très rapprochée, le son résiduel devrait provoquer des *battements*.

Toutes ces expériences ont été tentées et n'eurent les résultats attendus que dans le cas de sons différentiels effectivement provoqués par la distorsion non-linéaire de primaires de forte intensité. Mais elles n'ont aucun effet sur le son résiduel tel que défini plus haut.

Si la fondamentale ne doit pas nécessairement être présente, quelle est la dimension du son qui fait que la hauteur perçue corresponde tout de même à celle qu'aurait la fondamentale si elle y était ?

De nombreuses expériences ont été menées, à l'aide de sons complexes générés par synthèse additive, dans le but de révéler les caractéristiques de la perception du son résiduel.

Le son résiduel n'est pas un son différentiel

La fréquence du son résiduel correspond toujours à celle d'une fondamentale, quelle soit présente ou non, c'est-à-dire que les primaires doivent être en relation harmonique (par exemple : 1100, 1200, 1300, 1400 Hz). S'ils ne le sont qu'imparfaitement, par exemple, si on ajoute un nombre constant de Hertz à chacune des fréquences, ce qui donnerait 1110, 1210, 1310, 1410 Hz, la fréquence perçue augmente légèrement, alors que la différence entre les fréquences reste bien de 100 Hz, comme dans le premier cas. Le son résiduel n'est donc pas un son différentiel (résultant de la différence des fréquences des composantes spectrales en présence).

Voici un autre exemple :

composantes primaires à 1 800, 2 000, 2 200 Hz → son résiduel à 200 Hz

composantes primaires à 1 840, 2 040, 2 240 Hz → son résiduel à 204 Hz

On peut estimer la fréquence du son résiduel en calculant la moyenne des harmoniques présentes pondérées par l'inverse de leur rang le plus probable. Pour l'exemple ci-dessus, cela donnerait :

$$\left(\frac{1\,840\text{ Hz}}{9} + \frac{2\,040\text{ Hz}}{10} + \frac{2\,240\text{ Hz}}{11} \right) \div 3 = 204,03\text{ Hz}$$

À partir de cette fondamentale, on peut calculer les harmoniques correspondant aux rangs 9, 10 et 11, ce qui donne :

$$(9, 10, 11) \times 204\text{ Hz} = 1\,836\text{ Hz}, 2\,040\text{ Hz}, 2\,244\text{ Hz}$$

Il semblerait donc que le système auditif recherche la solution la plus proche qui est plausible physiquement. L'exemple ci-dessus est assez artificiel, mais précisons que l'on pourrait observer ce mécanisme de perception de la hauteur résiduelle dans le cas de certains instruments, comme le piano ou la guitare, qui présentent des partiels qui ne sont pas parfaitement harmoniques et une fréquence fondamentale généralement faible. On peut remarquer que la hauteur perçue pour le son complet semble plus haute que celle perçue pour le fondamental, lorsque celui-ci est isolé par filtrage.

La force perçue du son résiduel dépend de la forme d'onde

La force perçue du son résiduel est très peu dépendante de l'intensité des primaires, mais par contre, elle dépend fortement de l'enveloppe de la forme d'onde du complexe (contour dynamique au cours du temps). Plus la périodicité est marquée, c'est-à-dire plus les changements périodiques de l'enveloppe temporelle sont brusques, plus le son résiduel est intense.

Ceci implique donc une influence de la relation de phase entre les partiels d'un son, puisque selon leurs phases respectives, des partiels identiques quant à la fréquence et à l'amplitude peuvent donner des formes d'ondes différentes.

Ce phénomène sera aussi très important pour comprendre que certains sons musicaux sont plus ou moins efficaces à générer une «hauteur musicale» très précise.

Les sons pseudo-périodiques

Les sons pseudo-périodiques, obtenus par modulation ou par alternance, ont démontré que la sensation de hauteur était davantage due à la récurrence plus ou moins régulière du phénomène physique, plutôt qu'à son contenu fréquentiel (selon une analyse de Fourier par exemple).

Toutes ces observations permettent de conclure que ce sont des *informations temporelles* qui pourraient être à la source de la sensation de hauteur. Ceci nous amène à présenter la seconde théorie permettant d'expliquer la perception de la hauteur : la *théorie de la périodicité*.

6.8.3 La théorie de la périodicité

La théorie de la périodicité propose que ce sont des informations temporelles qui sont à la source de la sensation de hauteur, le rôle de la fondamentale d'un son complexe n'étant donc que de renforcer la périodicité, car sa présence ne serait pas indispensable.

Lorsqu'on soumet le système auditif à un son périodique, la membrane basilaire vibre, emmenant les cellules ciliées dans ce mouvement. La fréquence serait alors codée d'après le patron temporel de la décharge des neurones périphériques, décharge qui se produit à une phase particulière de l'onde.

Il faut noter que les influx nerveux au départ de la cochlée ne peuvent pas se succéder à n'importe quel rythme : après chaque décharge du neurone, une *période de récupération*, de 1 à 10 ms, est nécessaire avant qu'un autre influx puisse être généré, ce qui limiterait la fréquence en théorie à des fréquences inférieures à 100 ou 1000 Hz.

Aux basses fréquences, les cellules nerveuses émettent un influx nerveux périodique synchronisé, au moins statistiquement, avec les crêtes de la forme d'onde du stimulus : c'est le phénomène de **synchronisation sur la période** ou "*phase locking*" en anglais.

Par contre, pour les fréquences plus élevées, la synchronisation n'est plus possible à cause de la période de récupération nécessaire et certaines cellules vont ainsi décharger 1 fois sur 2, d'autres 1 fois sur 3, etc. Les décharges neuronales successives se produisent ainsi à un multiple entier de la période de l'onde. Plusieurs fibres nerveuses réagiraient successivement, l'une répondant alors que la précédente est en période de récupération. C'est la **théorie des salves** ou "*Volley Theory*" qui permet d'expliquer pourquoi la théorie de périodicité est toujours valide pour les fréquences plus élevées, quoique moins efficace.

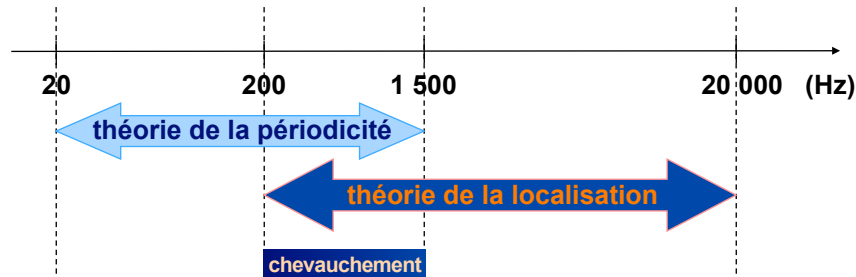
Précisons que ce problème n'est pas spécifique à la sensation de hauteur. Dans le cas de la perception de la localisation d'une source dans l'espace, on a mis en évidence des phénomènes qui impliquent des mesures de différences de temps de l'ordre de quelques micro-secondes ! Mais on réfère alors à des perceptions binaurales, alors que le phénomène du son résiduel se manifeste à l'intérieur d'une seule oreille.

6.8.4 Tentative de synthèse des deux théories sur la perception de la hauteur

Il est important de souligner que la plupart des expériences qui démontrent l'une ou l'autre des deux théories (localisation ou périodicité), ne donnent pas les mêmes résultats selon la fréquence des stimulus. Presque toutes les expériences de la théorie de la localisation sont davantage valables pour les hautes fréquences, celles de la théorie de la périodicité étant davantage

valables pour les basses fréquences. En effet, même en tenant compte de la théorie des salves, la rapidité des influx nerveux n'est pas suffisante pour expliquer la perception de la hauteur dans les hautes fréquences.

De plus, la frontière n'est pas stricte entre les domaines de fréquences pour lesquels s'appliquent les deux mécanismes de la perception de la hauteur. On reconnaît généralement une région de chevauchement entre 200 et 1500 Hz, c'est-à-dire une zone de fréquence où les deux mécanismes peuvent être observés.



Il est intéressant de remarquer que ces fréquences correspondent au registre à l'intérieur duquel se retrouvent toutes les fréquences musicales, c'est-à-dire les fondamentales (présentes ou non) des hauteurs utilisées en musique, au moins sur le plan mélodique. Il est fort possible que la coexistence de deux mécanismes complémentaires renforce la perception de hauteur.

Quoique le lien n'ait pas été systématiquement étudié, on peut aussi penser que chacun des deux mécanismes pourrait être à l'origine de l'une des deux sensations de hauteur, la hauteur spectrale (ou de « timbre ») relevant de la théorie de la localisation, et la hauteur « musicale » relevant de la théorie de la périodicité. Ceci pourrait expliquer plusieurs différences observées entre sujets musiciens et sujets non-musiciens.

tonie (distinction grave-aigu) → théorie de la *localisation*
chroma (*do, ré, mi, fa, sol, la, si*) → théorie de la *périodicité*

6.9 L'origine des acouphènes : la plasticité cérébrale ?

Les acouphènes de type « tonal », c'est-à-dire semblables à un son pur (de haute fréquence) et décrits comme un « sillement », pourraient être dû à la plasticité cérébrale. Grâce à des technologies d'imagerie cérébrale modernes, une étude (Mühnickel et al., 1998) a montré que, chez les personnes souffrant de ce type d'acousphène, la tonotopie au niveau du cortex auditif primaire (gyrus de Heschl controlatéral à l'oreille de l'acouphène) est anormale : il y a une déviation, par rapport aux oreilles de personnes sans acouphène, à l'endroit qui correspond à la fréquence de l'acouphène. Cette déviation est une conséquence de la plasticité cérébrale : le cerveau se réorganise suite à des changements périphériques ou centraux. Ainsi, une perte auditive périphérique (qui est sous-entendue dans les cas d'acouphènes) amènerait des changements au niveau central.