

# Modèles psychoacoustiques et perception de hauteur

Daniel Pressnitzer

Ircam-CNRS, 1 place Igor Stravinsky, 75004 Paris.

email : [pressnitzer@ircam.fr](mailto:pressnitzer@ircam.fr)

## 1 Introduction

Pour la plupart d'entre nous, la perception de hauteur semble être quasiment évidente et immédiate. L'histoire du musicien se promenant dans une forêt avec son ami illustre bien ceci : alors qu'un bruit soudain se fait entendre, et que l'ami demande effrayé "Qu'est-ce que cela peut bien être ?", le musicien répond aussitôt "Un si bémol, je crois". Pourtant, la hauteur d'un son est un exemple typique de grandeur psychologique et non physique dont nous sommes encore loin de comprendre les causes. Au cours de cette conférence, nous aimerions développer l'idée qu'il faut prendre en compte un ensemble de caractéristiques du système auditif pour étudier cet attribut perceptif complexe. L'un des moyens d'intégrer ces caractéristiques est de bâtir des modèles computationnels simulant les étapes précoces de traitement auditif, et nous décrirons brièvement le principe de tels modèles. Enfin, une étude spécifique visant à délimiter la frontière entre rythme et hauteur sera présentée pour illustrer sur un exemple l'objet d'une telle approche.

## 2 Perception de hauteur

Un grand nombre de sons de natures très différentes peuvent donner lieu à la perception de hauteur (voir [4] et [6] pour des revues). Le plus simple d'entre eux est bien sûr le son pur, ou son sinusoïdal. Un tel son est entièrement défini par sa fréquence, son amplitude et sa phase. Superficiellement, il semble qu'un tel ne pose pas de problème particulier pour la compréhension des percepts qu'il évoque : son intensité perçue correspondrait à son amplitude, sa hauteur perçue à sa fréquence. En fait, les choses sont beaucoup plus compliquées que cela. L'intensité perçue pour une amplitude donnée change avec la fréquence, elle est plus faible pour les sons graves. La hauteur pour une même fréquence peut, elle, changer en fonction de l'intensité [17] ou même disparaître pour des fréquences trop graves ou trop aiguës.

Si la situation est plus compliquée qu'il n'y paraît pour le son pur, elle devient encore plus intrigante encore lorsque l'on s'intéresse à d'autres types de signaux. Il est bien connu que les sons harmoniques (composés de sons purs aux fréquences multiples d'une fondamentale) provoquent une impression de hauteur sans équivoque, correspondant justement à celle d'un son pur à la fréquence fondamentale. La majorité des sons de l'orchestre symphonique sont d'ailleurs de ce type. Les chercheurs en psychoacoustique ont développé un certain nombre de sons grâce auxquels ils réduisent les indices physiques disponibles, de manière à décider lesquels sont importants pour la perception de hauteur. Ainsi, il est possible de supprimer par filtrage l'énergie correspondant à la fréquence fondamentale du son, sans pour autant perdre la

sensation de hauteur. La présence d'énergie spectrale à la fréquence fondamentale n'est donc pas nécessaire pour percevoir la hauteur correspondante [15]. Il est aussi possible de brouiller la structure temporelle, en utilisant des sons obtenus par délai-addition de bruit blanc [20]. Ces sons, dit bruits itérés, produisent une perception de hauteur alors même qu'ils ne sont pas parfaitement périodiques. Un bruit blanc modulé en amplitude, donc sans structure spectrale, peut s'avérer suffisant pour reconnaître des mélodies [1]. Il serait possible de multiplier ainsi les exemples : comme on le voit, la perception de hauteur ne peut s'appuyer sur une description physique simple des signaux acoustiques.

### 3 Modélisations auditive

Une solution possible pour expliquer et prédire la perception de hauteur de manière unifiée consiste à bâtir un modèle basé sur les transformations subies par tout son lorsqu'il est traité par notre système auditif périphérique. Une fois que ces étapes sont prises en compte, les caractéristiques physiques sont transformées en indices perceptifs pouvant alors servir de base aux modèles de hauteur.

#### 3.1 Etapes précoces de traitement auditif

L'une des premières étape de traitement auditif que tout son subit lorsqu'il parvient à nos oreilles est une propagation le long d'une membrane située dans l'oreille interne, la membrane basilaire. Du fait de propriétés mécaniques et physiologiques, chaque région de cette membrane répond préférentiellement à une certaine gamme de fréquences [19, 18]. De manière simplifiée, il est possible de modéliser le résultat net de cette étape de traitement par un banc de filtres [10]. Chacun des filtres est appelé canal auditif, et il est défini par sa fréquence centrale et sa largeur de bande passante [5]. Si l'on considère l'activité dans les canaux auditifs au cours du temps, une représentation de type spectrographique dont les paramètres sont adaptés à l'audition est obtenue. L'activité dans chaque canal sera ensuite codée sous forme d'impulsions nerveuses par des cellules situées le long de la membrane, les cellules ciliées internes. Il existe des modèles physiologiques reproduisant l'action des cellules ciliées, notamment leur capacité à se synchroniser avec l'activité qu'elle reçoive jusqu'à certaines fréquences [8]. Les impulsions ainsi produites sont ensuite transmises vers le cerveau (voir [12] ou [13] pour des revues approfondies de toutes ces étapes).

#### 3.2 Exemple du son harmonique

Considérons comme exemple de son provoquant une forte sensation de hauteur un son parfaitement périodique, ici constitué par la superposition des quinze premiers harmoniques d'une fréquence fondamentale 100 Hz. Nous avons choisi tous les harmoniques en phase, et donc l'onde temporelle d'un tel son ressemble à un train d'impulsions (Figure 1).

Lorsque l'on calcule l'activité dans les canaux auditifs provoquée par le son de la Figure 1 en utilisant un modèle simple [11], une représentation de type "auditive" est obtenue (Figure 2). A l'intérieur de cette représentation, il est possible de remarquer plusieurs indices trahissant la période du son harmonique. Si l'on calcule l'énergie présente dans chacun des canaux auditifs et que l'on reporte le résultat en fonction de la fréquence centrale du canal (Figure 3, haut), une représentation de type spectre à long terme est obtenue. Une structure harmonique est clairement visible, signe de la présence de périodicité. Ce premier indice est dit spectral. Si l'on

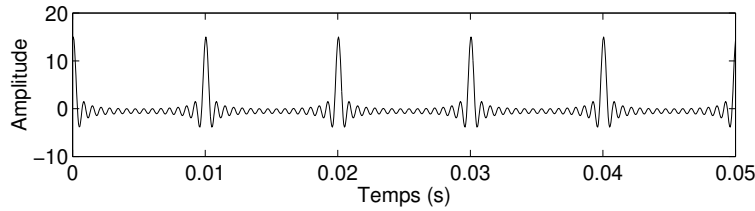


FIG. 1 – *Onde temporelle d’un son harmonique de fréquence fondamentale 100 Hz comportant 15 harmoniques en phase.*

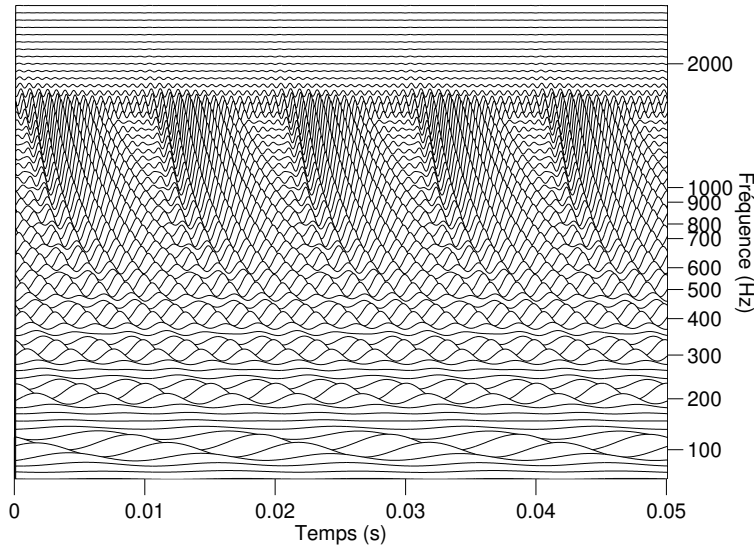


FIG. 2 – *Représentation de l’activité au cours du temps provoquée par le son harmonique de la Figure refnwave dans un ensemble de canaux auditifs centrés de 50 à 2500 Hz. Une forme habituelle d’implémentation des filtres, dite “gammatone”, a été utilisée [11].*

considère l’activité dans certain canaux de fréquence centrale élevée, l’enveloppe du signal est régulière et reflète la périodicité (Figure 3, milieu). Enfin, dans d’autres canaux de fréquence centrale grave, la structure temporelle fine correspond à un son pur à la fréquence fondamentale du son (Figure 3, bas).

### 3.3 Modèles de perception de hauteur

Un certains nombres de modèles spécifiques ont été développés de manière à extraire une estimation de la hauteur perçue à partir de représentations de type de celle de la Figure 2. Une première catégorie de théories s’appuie sur les indices spectraux et un apprentissage de type reconnaissance de patterns [3]. D’autres considèrent que les indices de type temporels à l’intérieur des canaux auditifs sont pertinents et qu’ils peuvent être extraits à partir d’une fonction d’autocorrélation [7, 16] ou à partir d’un décompte d’intervalles entre pics d’activité [9]. Enfin, certains pensent que les deux mécanismes spectraux et temporels coexistent pour donner lieu à la perception de hauteur [2].

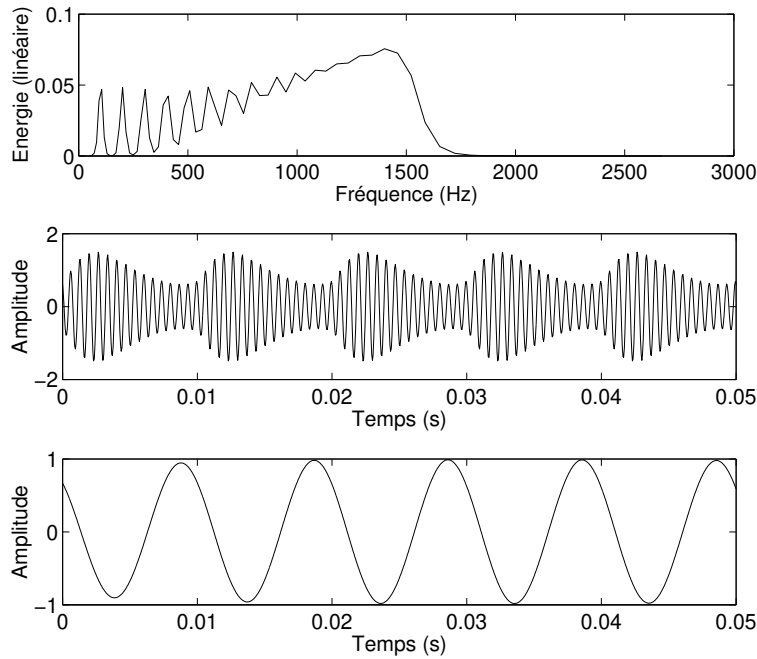


FIG. 3 – A partir d'un signal périodique (fréquence fondamentale 100 Hz, 15 composantes, représenté en Figure 1), le système auditif dispose de plusieurs indices de régularité : (a) la répartition de l'énergie à travers les canaux donne une représentation spectrale ; (b) un canal dans lequel plusieurs composantes interagissent – ici centré à 1.5 kHz – fournit une enveloppe régulière ; (c) un canal centré à proximité de la fréquence fondamentale reproduit la structure temporelle fine d'un son pur à 100 Hz. Ces indices sont directement extraits de la simulation du filtrage auditif représentée en Figure 2. Les unités d'amplitude sont arbitraires.

L'existence de ces modèles concurrents ne doit pas être vue comme un aveu d'échec, mais bien au contraire comme le signe d'une recherche féconde sur les causes même du percept de hauteur. Mieux ces causes seront comprises, plus généraux seront les modèles proposés car ils s'attacheront à reproduire des *mécanismes* auditifs et non pas des *observations* isolées. Dès à présent, le modèle de type autocorrélation est à même de prédire la perception de hauteur de la plupart des sons dans la pratique, incluant la fondamentale absente et le bruit itéré, et ceci à partir d'un principe unique ne réclamant pas à être adapté pour chaque situation. Pour finir, nous aimerions évoquer en guise d'exemple l'une de ces situations particulière.

## 4 Hauteur et rythme

La musique occidentale a choisi, tout au moins traditionnellement, de noter hauteur et rythme de manières fort différentes : la hauteur est représentée par la position verticale de chaque note sur la portée ; le rythme est noté par le type et la position horizontale de ces mêmes notes. Si l'on considérait uniquement les caractéristiques physiques des sons, il existerait en fait un continuum entre hauteur et rythme : il est possible de passer continuellement d'un son harmonique de période 10 ms (100 Hz) à un son de période 1 s (1 Hz). Pourtant, l'un de ces sons provoque une perception de hauteur alors que l'autre est perçu comme un rythme. Ce qui nous permet de distinguer hauteur et rythme, ce sont encore une fois uniquement certaines caractéristiques de la perception auditive humaine.

### 4.1 Limite inférieure de hauteur mélodique

La distinction faite par le langage entre hauteur et rythme repose sur une convention implicite. Pour la définir de manière plus rigoureuse, il convient de développer des méthodes permettant de distinguer expérimentalement entre les deux percepts. L'une des possibilités découle du fait que la hauteur est "la qualité des sons permettant d'établir des mélodies" [4]. On parle dans ce cas de hauteur tonale ou mélodique. Il est alors possible d'étudier les limites de la perception de hauteur en mesurant la capacité ou non d'auditeurs à traiter de courtes mélodies. Cette mesure, objective, est très différente de la méthode qui consisterait à demander qualitativement aux auditeurs s'ils perçoivent une "hauteur" non.

Une expérience a donc été réalisée où une mélodie de 4 notes, tirée au hasard dans la gamme chromatique à l'intérieur d'une tierce majeure au dessus d'une note de base, était présentée aux auditeurs. La mélodie était ensuite répétée, mais une modification de plus ou moins un demi-ton était introduite sur l'une des quatre notes. La position temporelle de la note modifiée était aussi tirée au hasard. La tâche des auditeurs était d'indiquer sur quelle note la modification avait été introduite. Lorsque les notes constituant les mélodies possèdent une hauteur claire, par exemple lorsque la note de base est autour de 250 Hz, la tâche est extrêmement aisée même pour des auditeurs non-musiciens. Lorsque la période de la note de base excède une certaine valeur, la tâche devient impossible. Nous interprétons ceci comme la disparition d'un percept de hauteur assez saillant pour effectuer la tâche, et dénommons cette valeur la Limite Inférieure de Hauteur Mélodique (LIHM).

La mesure de LIHM a été d'abord réalisée avec comme "notes" des trains d'impulsion large-bande. Dans un deuxième temps, les trains d'impulsions ont été limités à une certaine région fréquentielle. Il est ainsi possible de varier la fréquence fondamentale indépendamment de la région fréquentielle dans laquelle de l'énergie spectrale est présente. Nous sommes alors

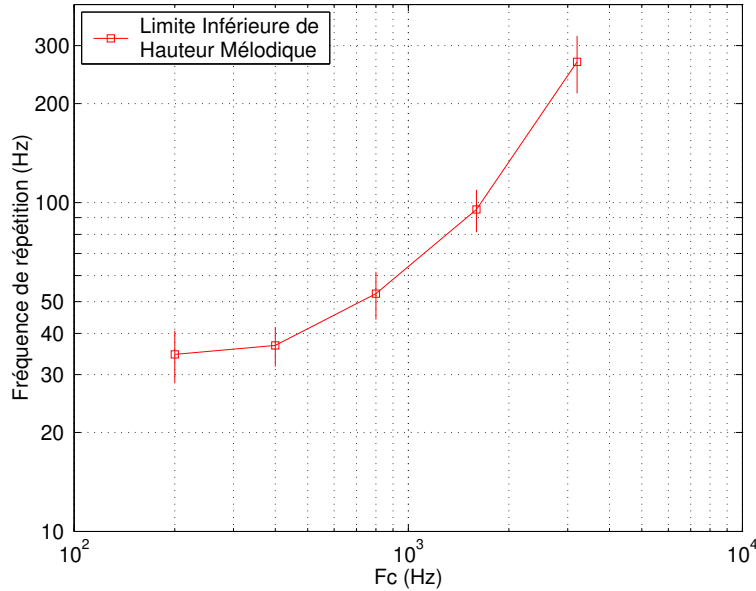


FIG. 4 – Résultats de l’expérience de Limite Inférieure de Hauteur Mélodique. La fréquence de la note de base pour laquelle la tâche mélodique devient impossible est indiquée en fonction de la fréquence de coupure du filtrage (voir texte). Moyennes et écarts-type pour 3 sujets.

dans un cas de fondamentale absente. La fréquence basse de la bande-passante du filtre appliqué, notée  $F_c$ , a été étudiée comme un paramètre expérimental supplémentaire. Un bruit continu était appliqué pour masquer d’éventuels produits de distorsion.

Lorsque des trains d’impulsions large-bande sont utilisés, la LIHM a est atteinte pour une période proche de 33 ms (30 Hz). Pour ce qui est de la perception de courtes mélodies chromatique, cette valeur représente sans doute la meilleure performance possible. Notons que cette valeur est proche de la note la plus basse du clavier du piano (27.5 Hz).

Lorsque la région fréquentielle est manipulée, la LIHM change : la perception de hauteur persiste pour des régularités plus longues dans les régions fréquentielles graves (Figure 4). Ainsi, il n’existe pas une unique valeur seuil pour définir la limite de perception de hauteur, mais cette limite dépend du timbre des sons utilisés. La valeur obtenue est même étonnamment haute pour des sons n’ayant pas d’énergie dans les régions fréquentielles graves. Il s’agit là d’un exemple d’interaction entre caractéristiques physiques (périodicité, composition fréquentielle) et caractéristiques du système auditif. Il est possible de reproduire cet effet avec un modèle de type autocorrélation [14], illustrant de nouveau l’intérêt de posséder un outil fournissant une représentation “auditive” du son, en complément des méthodes d’analyses acoustiques plus traditionnelles.

## Références

- [1] E. M. Burns and N. F. Viemeister. Played again SAM : Further observations on the pitch of amplitude-modulated noise. *J. Acoust. Soc. Am.*, 70 :1655–1660, 1981.
- [2] R. P. Carlyon and T. M. Shackleton. Comparing the fundamental frequencies of resolved and unresolved harmonics : Evidence for two pitch mechanisms? *J. Acoust. Soc. Am.*, 95 :3541–3554, 1994.
- [3] M. A. Cohen, S. Grossberg, and L. L. Wyse. A spectral network model of pitch perception. *J. Acoust. Soc. Am.*, 98(2) :862–879, 1995.
- [4] L. Demany. *Psychoacoustique et Perception Auditive*, chapter Perception de la hauteur tonale, pages 44–81. INSERM/EMI/SFA, Paris, 1989.
- [5] B. R. Glasberg and B. C. J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47 :103–138, 1990.
- [6] A. J. M. Houtsma. Pitch perception. In B. C. J. Moore, editor, *Hearing*, Handbook of Perception and Cognition, 2nd Edition, pages 267–295, San Diego, 1995. Academic Press.
- [7] J. C. R. Licklider. A duplex theory of pitch perception. *Experientia*, 7 :128–133, 1951.
- [8] R. Meddis. Simulation of mechanical to neural transduction in the auditory receptor. *J. Acoust. Soc. Am.*, 79 :702–711, 1986.
- [9] B. C. J. Moore. *An Introduction to the Psychology of Hearing*, 4th ed. Academic Press, London, 1997.
- [10] R. D. Patterson. Auditory filter shape derived with noise stimuli. *J. Acoust. Soc. Am.*, 59 :640–654, 1976.
- [11] R. D. Patterson, M. H. Allerhand, and C. Giguère. Time-domain modelling of peripheral auditory processing : A modular architecture and a software platform. *J. Acoust. Soc. Am.*, 98 :1890–1894, 1995.
- [12] J. O. Pickles. *An Introduction to the Physiology of Hearing*. Academic Press, London, 2nd. edition, 1988.
- [13] A. N. Popper and R. R. Fay. *The Mammalian Auditory Pathway : Neurophysiology*. Springer Handbook of Auditory Research. Springer-Verlag, New York, 1992.
- [14] D. Pressnitzer, R. D. Patterson, and K. Krumbholz. The lower limit of melodic pitch with filtered harmonic complexes. *J. Acoust. Soc. Am.*, 105 :1152, 1999.
- [15] J. F. Schouten. The residue and the mechanism of hearing. *Proc. K. Ned. Akad. Wet.*, 43 :991–999, 1940.
- [16] M. Slaney and R. F. Lyon. On the importance of time - a temporal representation of sound. In M. Cooke, Beet. S., and M. Crawford, editors, *Visual representations of speech signals*, pages 95–116. Wiley & Sons, 1993.
- [17] K. N. Stevens. The relation of pitch to intensity. *J. Acoust. Soc. Am.*, 6 :150–154, 1935.
- [18] G. von Békésy. *Experiments in Hearing*. McGraw-Hill, New York, 1960.
- [19] Hermann L. F. von Helmholtz. *On the Sensations of Tone as the Physiological Basis for the Theory of Music*. 1877. 2nd. Ed. trans. A. J. Ellis (1885), from German 4th Ed., Dover, New York (1954).
- [20] W. A. Yost, R. D. Patterson, and S. Sheft. A time domain description for the pitch strength of iterated rippled noise. *J. Acoust. Soc. Am.*, 99(2) :1066–1078, 1996.