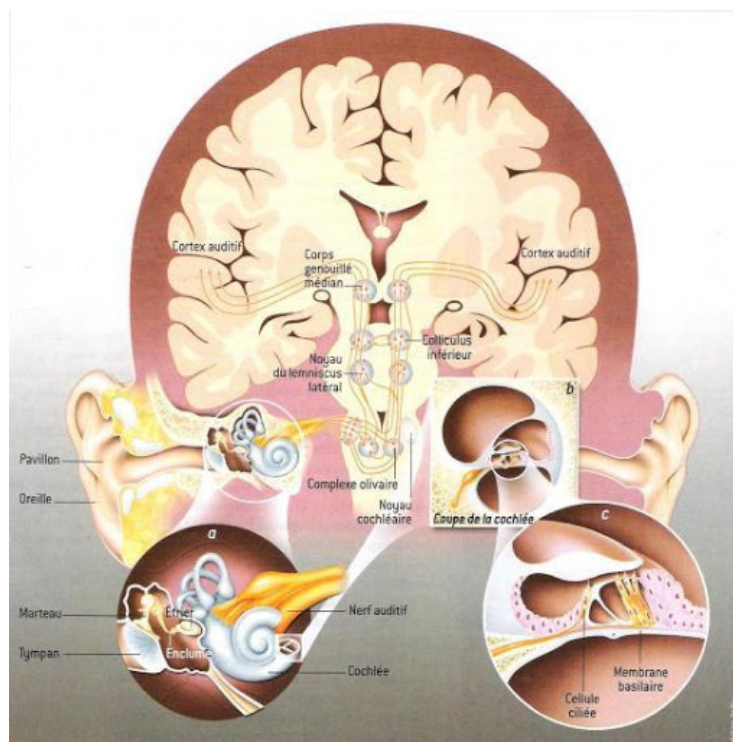


# Chapitre 3

## Perception du son

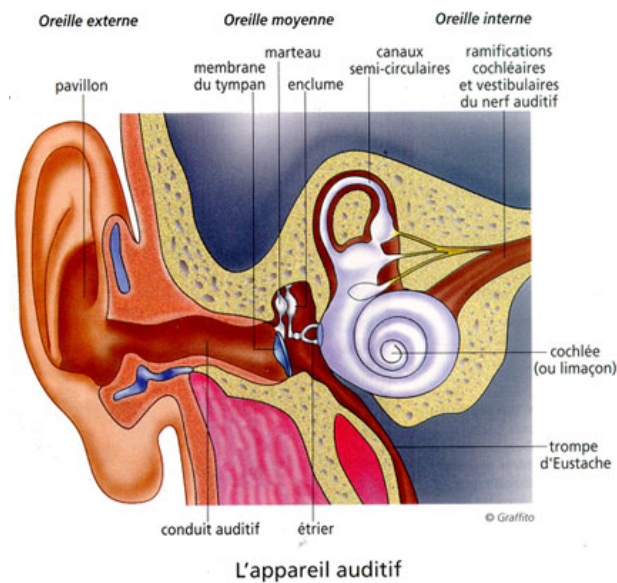
Dans ce chapitre on décrit le transfert d'information entre un signal sonore  $p(t)$ , i.e. **variation de pression**, à l'entrée d'une oreille (ou de chaque oreille) jusqu'à la **perception consciente d'un son** et ses caractéristiques par l'individu. Tous les mécanismes physiologiques qui participent à ce transfert d'information, s'appelle le **système auditif**. Cette description sera en fait très partielle pour la principale raison que la plupart des mécanismes sont encore mal compris voir totalement inconnus comme ce qu'est la "la conscience" qui est l'étape finale de ce transfert. De plus le signal n'est pas vraiment transféré mais plutôt "analysé", "filtré" et "transformé" de façon assez complexe par les étapes situés entre l'oreille et le conscient, appelé "inconscient". Seuls quelques unes des caractéristiques du signal sont extraites et transformées de multiples façon. Cela conduit à des illusions auditives que nous décrirons.

### 3.1 Description du système auditif

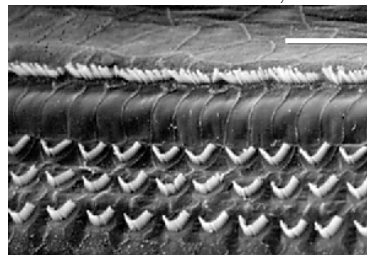


**Voici une description rapide des différentes étapes (plus détaillées ensuite) :**

- Une partie de l'onde de pression (le son) qui arrive au voisinage des oreilles est diffusée par **le pavillon** (voir aussi **pavillon**). Une partie de l'onde entre dans le conduit auditif et atteint la membrane du tympan.
- La **membrane du tympan** se met en mouvement et transmet ce mouvement à la cochlée via trois os qui sont articulés : marteau, enclume, étrier. Il semblerait que la présence de ces os intermédiaires permette une bonne transmission de l'énergie sonore qui seraient réfléchi sinon. Un autre rôle de ces petits os est qu'ils servent de "fusible" : si le signal est trop fort, les os se bloquent et évitent de transmettre une énergie trop forte aux organes auditifs suivants.



- Dans la cochlée il y a un fluide qui reçoit les mouvements de vibration de la part de l'enclume, et ces mouvements pénètrent dans la **cochlée** qui est formée de conduits enroulés en colimaçon. Le long de la cochlée, il y a une membrane appelée **membrane basilaire** dont la taille dépliée est 10cm par 1cm, et qui est en contact de tout son long avec des milliers de petits cils, appelées **cellules ciliées**. Les fréquences aiguës restent au début de la cochlée alors que les fréquences graves pénètrent tout le long de la membrane basilaire jusqu'au bout de la cochlée.
- Chaque cil est comme un résonateur. Il peut se mettre en mouvement sous l'influence de la membrane basilaire. Il libère alors des ions qui excitent des terminaisons nerveuses de **neurones**, des dendrites. Photo de cils, à l'échelle =  $15\mu m$  :



- Il y a donc des milliers de **dendrites** issues de ces cils (i.e. fils conducteurs) qui transmettent des influx nerveux venant spécifiquement de chaque cil. Toutes ces dendrites sont regroupées en parallèle dans le “nerf auditif” qui envoie cette information très riche au cerveau.
- L'information nerveuse est analysée par différentes parties du cerveau. Ne nombreux calculs sont effectués par des algorithmes mal connus ou non connus. Tout cela est inconscient, c'est à dire non perçu par la personne elle même. Certains résultats de tous ces calculs et analyses sont envoyées vers des zones du cerveau qui participent à la “conscience”, c'est à dire aux événements perçus par la personne.

Voici une description de quelques aspects de cette analyse du son depuis son entrée dans

le pavillon de l'oreille jusqu'à ce qu'il atteigne la conscience.

### 3.1.1 Le pavillon de l'oreille

Nous avons deux oreilles, chacune captant un signal sonore et le pavillon de chaque oreille est très “sculpté”. Cela semble essentiel à la perception tri-dimensionnelle du son.

- Regarder cette [vidéo](#) et les liens donnés qui suggère que la forme complexe du pavillon est utile à la localisation 3 dimensionnelle du son. En particulier regarder cette [expérience](#) que vous pouvez faire vous même : elle consiste à coller de la pâte à modeler sur son oreille pour modifier sa géométrie et remarquer, les yeux bandés, que l'on ne sait plus localiser l'origine spatiale des sons. Il paraît que si on garde la nouvelle géométrie quelques semaines, le cerveau se reprogramme, grâce à la synchronisation avec la vue, et la perception spatiale revient (mais re-disparaîtrait si on revenait à la forme initiale de l'oreille!).



- Dans cette [vidéo de la BBC](#), on voit un renard qui se concentre sur la localisation 3 dimensionnelle de sa proie cachée sous une épaisse couche de neige.



Ref: Labo de Brams

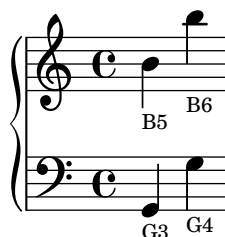
**Exercice 3.1.1.** (TP) “Pavillon de l'oreille et localisation du son”. Essayer de reproduire les expériences ci-dessus.

### 3.1.2 Cils

Les cils sont situés dans la cochlée et fonctionnent comme des résonateurs. On a étudié le modèle d'un résonateur en Section B.7. Les paramètres importants sont la fréquence de résonance  $f_0$  et la largeur de résonance  $\Delta f_0$ . On peut convertir  $f_0$  et  $\Delta f_0$  en unité de

pitch, c'est à dire demi-ton, d'après (2.3.5), donnant respectivement  $x_0, \Delta x_0$ . Voici l'ordre de grandeur pour différents cils, d'après Rossing [Fletcher & Rossing \(2012\)](#), p.74 :

$f$ (Hz)	100	200	500	1000	2000	5000	10000
$\Delta f$ (Hz)	90	90	110	150	280	700	1200
$x_0$ (en note)	$G_3$	$G_4$	$B_5$	$B_6$	$B_7$	$Eb_8$	$Gb_8$
$\Delta x_0$ (en $\frac{1}{2}$ ton)	15	7	4	2.5	2	2	3



Il est important de remarquer dans ces données que la précision  $\Delta x_0$  en unité de pitch (demi-ton) est supérieure à 1 alors que la perception consciente donne une précision bien meilleure, de l'ordre de 1/10 pitch, voir Section 3.3.4. Cela suggère que le raffinement de la précision se fait dans une étape ultérieure de l'analyse, dans le cerveau.

### 3.1.3 Physiologie du cerveau

A la date actuelle (2021), on connaît très peu, voire pas du tout, les mécanismes précis d'analyse du son dans le cerveau. Le fonctionnement collectif du cerveau est encore mystérieux<sup>1</sup>. Par exemple on ne sait pas du tout où et comment fonctionne la mémoire, ni même ce qu'est précisément la "conscience" et où et comment fonctionne t-elle. Référence : [Dehaene \(2014\)](#). Voici cependant quelques observations et considérations.

- Noter que le temps de réponse d'un neurone est supérieures à  $20ms$  soit une fréquence inférieure à  $50Hz$ . Ainsi l'analyse du timbre d'un signal audible  $u(t)$  sur des fréquences  $f \geq 100Hz$  ne peut pas être une analyse temporelle faite dans le cerveau mais serait plutôt une analyse de la transformée de Fourier  $|(\mathcal{F}u)(f)|$ ? Cela se confirme par le fait que notre perception est insensible aux phases? Ce temps de réponse détermine la **latence acceptable en musique** :  $\Delta t = 20ms$ , i.e. un décalage temporel perceptible. On a déjà vu que sur cette durée  $\Delta t$  le son se propage dans l'air sur une distance  $L = c\Delta t = 343 \times 20 \cdot 10^{-3}m = 7m$ . C'est donc une distance au dessus de laquelle la latence peut être perceptible.

1. Remarquer par exemple que par rapport à il y a 10000 ans, les progrès de la connaissance scientifique ont apporté des bouleversement dans nos conditions de vie : transport, logement, nourriture, médecine de la plupart des organes de notre corps. Une exception notable est le cerveau qui reste totalement mystérieux. Une conséquence est que tous les problèmes reliés aux mauvais fonctionnement du cerveau, états d'humeur des personnes, personnalités extravagantes, anxiété, maladie mentales, affectant les relations humaines, tous ces problèmes sont présents de nos jours comme il y a 10000 ans. Quasiment aucun progrès n'a eu lieu sur cet aspect déterminant de la qualité de vie.

- Par imagerie IRM, on observe que les calculs liés à la parole et la résolution temporelle du son sont effectués dans la partie gauche du cerveau, alors que l'analyse de la résolution fréquentielle, dont la musique sont effectués dans la partie droite, cf [Conférence de Christine Petit](#) à 32'.

## 3.2 La voix et les signaux periodiques

L'appareil auditif est très adapté à la perception de la voix humaine. Pour cette raisons, dans cette section nous décrivons la voix humaine, comment elle est générée et perçue.

Chez les humains, la voix est apparue entre  $-2$  millions d'années (sons) et  $-50000$  ans (parole). Des [recherches récentes](#) font remonter l'apparition de la parole à  $-20.10^6$  ans. La parole a un rôle social pour la communication, pour palabrer, pour l'échange d'informations

...



On peut imaginer que notre cerveau est programmé (de façon inné et/ou acquise) pour percevoir particulièrement (i.e. analyser) certains sons qui nous sont importants comme :

- Sons de la voix humaine
- Sons d'une source d'eau, de la pluie, de l'orage
- Bruits inquiétants dans le silence : craquements de branches, respirations
- Bruits du cœur de sa maman, de la respiration ...

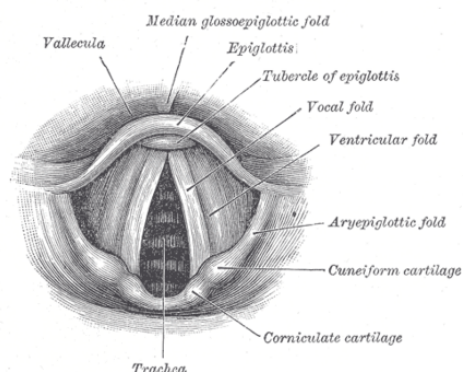
### Références :

- voir cet article sur [l'origine du langage](#).
- Ce livre de Harari [Dauzat & Harari \(2015\)](#) sur l'histoire et les aptitudes de l'espèce Homo Sapiens.
- [Schroeder et al. \(2007, chap.16\)](#).

### 3.2.1 Observations générales sur la voix

#### 3.2.1.1 Cordes vocales, forme de la cavité buccale

La voix est générée par la mise en vibration des “**cordes vocales**” qui ne sont pas véritablement des “cordes” mais des muscles sous tension, sous le flux d’air venant des poumons :

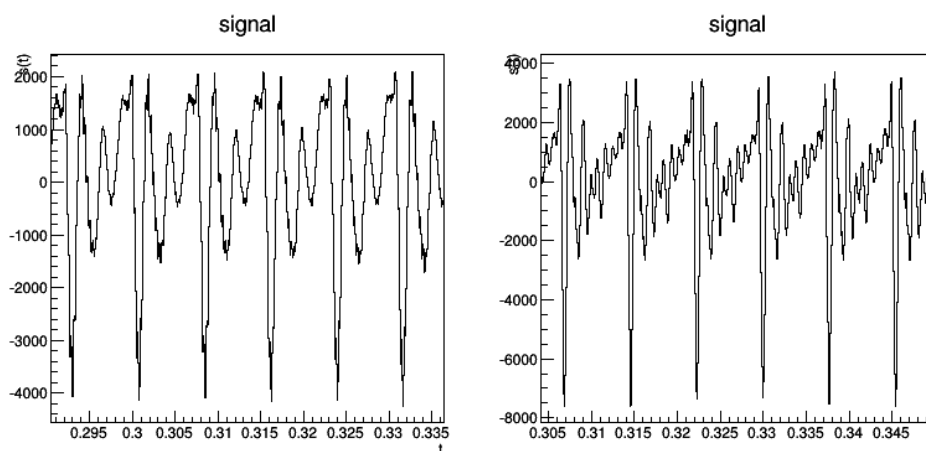


**Video** des “cordes vocales” en action.

Les cordes vocales produisent un signal sonore (presque) périodique que l’on qualifie de **voyelle**. Voir Section 2.3 sur les signaux périodiques. En fait toute la cavité buccale participe à sculpter ce signal. Cela correspond à des positions précises de la langue, du palais, etc.. Voir cette **Video**.

Le signal sonore produit et donc la voyelle prononcée dépend de la forme de la cavité buccale.

**Exemple 3.2.1.** Par exemple, voici des **voyelles chantées par Malik Mezzadri** sur une fréquence fixe de  $f = 130\text{Hz}$  (note  $C_4$ ). Voici le signal (mesuré avec un micro) pour les voyelles “ON” et “A” sur quelques périodes  $T = 1/f = 7\text{ms}$ . On observe en effet que le signal est (presque) périodique et dépend de la voyelle :



*Remarque 3.2.2.* la vitesse du son dans l'air étant  $v = 330\text{m/s}$  la longueur d'onde est donc  $\lambda = v/f = 2,50\text{m}$ . L'intensité est de l'ordre de 70dB.

On a expliqué dans la Section 2.3 toutes les propriétés des signaux périodiques, en particulier que la transformée de Fourier est discrète et répartie sur les multiples  $f_n = n f, n \geq 1$ , appelées harmoniques. Pour chaque harmonique  $n \geq 1$ , il y a un coefficient de Fourier de module  $|\hat{u}_n| \geq 0$ . Cet ensemble  $(|\hat{u}_n|)_{n \geq 1}$  s'appelle le **timbre** du signal.

On peut aussi communiquer en parlant sans les cordes vocales, i.e. en "chuchotant". L'avantage de la voix chantée (i.e. avec les cordes vocales) est peut être qu'elle est mieux perçue parmi d'autres sons ou bruits.

### 3.2.1.2 Les formants

Pour la voix humaine, on observe que l'enveloppe des coefficients de Fourier  $(|\hat{u}_n|)_n$  possède quelques maxima à des fréquences  $F_1, F_2, \dots$  appelés **formants**. Ces fréquences  $F_1, F_2, \dots$  dépendent de la voyelle prononcée.

D'après Chevalier p.53, la valeur des deux premiers formants  $F_1, F_2$  suffit à déterminer la voyelle prononcée d'après le schéma suivant. Les formants  $F_3, F_4$  apportent des nuances. Certains chanteurs ont un formant supplémentaire spécial entre  $F_3$  et  $F_4$ .

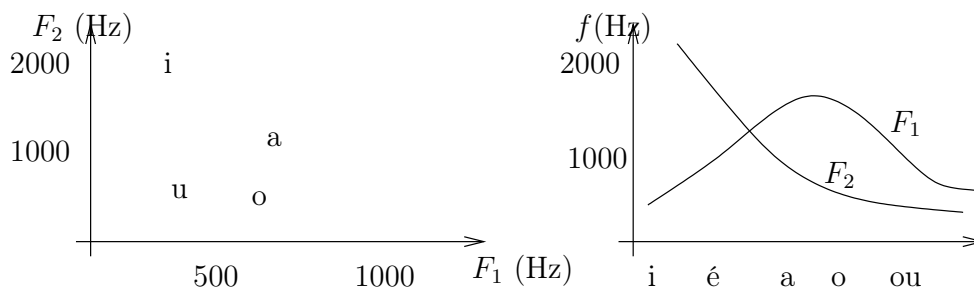


FIGURE 3.2.1 – Valeurs des formants  $F_1, F_2$  en Hz, pour différentes voyelles.

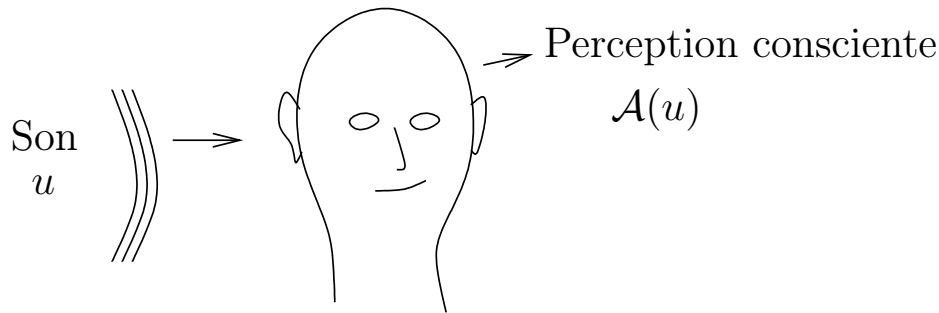
## 3.3 Du signal sonore à la perception consciente

Dans la section 3.1, on a décrit le système auditif. Dans cette Section, on va considérer le système auditif comme une "boîte noire", c'est à dire dont on ne connaît ni le contenu ni le fonctionnement. Seulement, on va considérer que le système auditif reçoit en entrée : un signal sonore, c'est à dire une fonction  $t \in \mathbb{R} \rightarrow u(t) \in \mathbb{R}$  et donne en sortie une "perception consciente" c'est à dire des sensations que l'on peut exprimer.

### 3.3.1 Définition de la perception sonore

**Question 3.3.1.** Pour un son donné  $u$ , identifier les caractéristiques  $\mathcal{A}(u)$  qui sont détectées au niveau de la perception consciente et les représenter, et décrire l'association  $u \rightarrow \mathcal{A}(u)$ .





On notera  $\mathcal{S}(\mathbb{R}_t)$  l'espace des fonctions  $t \rightarrow u(t)$  en entrée (précisément c'est l'espace de Schwartz (B.3.1), mais cela n'a pas d'importance ici).

On notera  $\mathcal{E}$  l'espace des perceptions conscientes des sons et

$$\mathcal{A} : \begin{cases} \mathcal{S}(\mathbb{R}_t) & \rightarrow \mathcal{E} \\ u & \rightarrow \mathcal{A}(u) \end{cases}$$

l'application qui à un signal sonore  $u$  associe une perception consciente  $\mathcal{A}(u)$ .

*Remarque 3.3.2.*

- L'espace d'entrée  $\mathcal{S}(\mathbb{R}_t)$  est bien défini. (il faudrait cependant rajouter une norme, ce que l'on fera). Mais la suite est très vague et pas bien définie : quel est cet espace  $\mathcal{E}$  ? quelle est cette application  $\mathcal{A}$  ? Comme mieux les définir ?
- On peut déjà remarquer que  $\mathcal{E}$  et  $\mathcal{A}$  doivent dépendre de l'individu. On a sûrement pas tous les même perceptions des sons, même si de nombreux aspects sont partagés (comme le fait que personne ne perçoit les ultrasons,  $f \gg 20000\text{Hz}$  et les infra sons  $f \ll 20\text{Hz}$ ).
- Pour deviner l'espace  $\mathcal{E}$  et l'application  $\mathcal{A}$  on peut effectuer des expériences de **psychoacoustique**.

Voici une définition de  $\mathcal{A}$  et  $\mathcal{E}$  très grossière mais qui nous permettra de progresser vers leur recherche.

*Définition 3.3.3.* Pour une personne donnée, on définit la relation d'équivalence  $\sim$  entre deux signaux sonores :

$$u_1 \sim u_2$$

si la personne ne les distingue pas consciemment. On définit l'**espace des perceptions sonores**  $\mathcal{E}$  comme l'**espace quotient**

$$\mathcal{E} := \mathcal{S}(\mathbb{R}_t) / \sim$$

et l'application **perception sonore**  $\mathcal{A} : u \rightarrow [u] \in \mathcal{E}$ .

Commençons donc à explorer cette relation d'équivalence  $\sim$  par quelques observations.

### 3.3.2 Perception du temps

On estime que l'on perçoit la position temporelle d'un signal avec une précision supérieure à

$$\Delta t \geq 0.1s \quad (3.3.1)$$

### 3.3.3 Perception de l'intensité

En Section 1.3.6 on a défini l'intensité d'un signal sonore en décibels (dB). L'intervalle maximal de perception à  $f \sim 1000\text{Hz}$  est

$$I \in [0, 120] \quad \text{dB.}$$

où  $I = 0\text{dB}$  est le seuil minimal de perception,  $I = 120\text{dB}$  est le maximum (ensuite c'est intolérable, voir destructeur) et la précision de perception vers  $1000\text{Hz}$  est

$$\Delta I \geq 1 \text{ dB}$$

C'est à dire que l'on ne distingue pas de différence entre deux signaux  $u_2 = \lambda u_1$  avec une constante de proportionnalité

$$\frac{1}{10^{1/10}} = 0.8 < \lambda < 10^{1/10} = 1.26.$$

.

#### 3.3.3.1 Courbes isophones

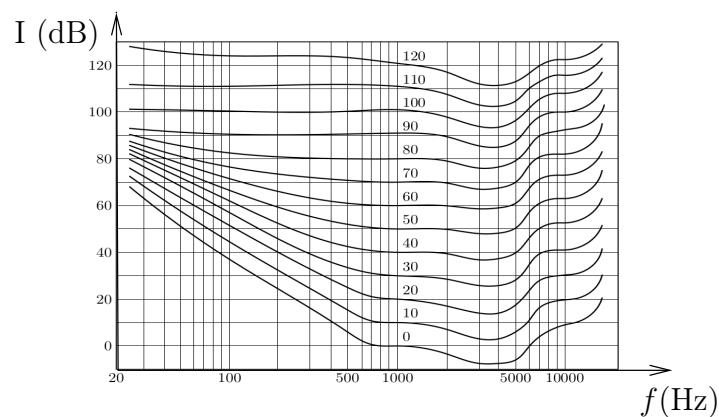


FIGURE 3.3.1 – Ce diagramme appelé courbes isophones de Fletcher-Muson montre des valeurs  $(f, I)$  dans le plan fréquence-Intensité qui donne la même perception d'intensité. Par exemple un signal  $I = 40\text{dB}$  à  $f = 100\text{Hz}$  est à peine audible, comme  $I = 0\text{dB}$  à  $f = 1000\text{Hz}$ .

Voir aussi la définition des **décibels dBA SPL**, voir [Benson \(n.d., p.10\)](#). Les courbes de la figure [3.3.1](#) sont mesurées par des expériences de psychoacoustiques avec des signaux sinusoïdaux :  $u(t) = A \sin(\omega t)$ , avec  $\omega = 2\pi f$

Observer que sur la figure [3.3.1](#) l'écart entre les courbes isophones dépend de la fréquence. Une conséquence importante est que en musique, pour un "mixage" de deux (ou plus) instruments jouant à des hauteurs (fréquences) différentes, si l'on souhaite réduire l'ensemble de "10dB" par exemple, il faudra réduire différemment les intensités de chacun selon sa fréquence.

### 3.3.3.2 Réverbération

Par définition, en acoustique musicale, si un signal s'éteint progressivement, on dit qu'il y a de la **réverbération**. Plus précisément

**Définition 3.3.4.** Si un signal décroît exponentiellement comme

$$u(t) = u(0) 10^{-t/\tau}$$

alors on appelle

$$\tau_{\text{reverb}} := 6\tau$$

le **temps de réverbération**. De façon équivalente

$$u(t + \tau_{\text{reverb}}) = 10^{-6} u(t)$$

$$I(t + \tau_{\text{reverb}}) = I(t) - 60 \text{ dB}$$

### 3.3.3.3 Mesure de l'intensité dans les synthétiseurs

Les synthétiseurs communiquent avec le langage MIDI. L'intensité d'un son est codé par le paramètre appelé "vitesse"  $v \in [0, 127]$ , qui est proportionnel à l'amplitude du signal d'après les spécifications MIDI. Cela est curieux, on penserait plutôt à  $v$  fonction affine du log de l'amplitude?

— Voir ce [document](#).

### 3.3.3.4 Effet de masquage

Un son grave suffisamment intense masque un son aigu plus faible, c'est à dire que l'on ne perçoit plus le son aigu. Mais l'inverse n'est pas vrai : un son grave faible est toujours perçu même en présence d'un fort son aigu.

**Exemple 3.3.5.** si il passe un train générant des sons graves forts très de nous, alors on ne s'entend plus parler. Mais si une alarme aiguë sonne, on s'entend encore parler.

L'explication physiologique de ce phénomène serait que le long de la membrane basilaire dans la cochlée, les capteurs des sons aigus sont à l'entrée et ceux des sons graves sont au

bout terminal. Le signal d'un son grave passe devant les capteurs des sons aigu pour aller au bout, mais pas l'inverse.

### 3.3.4 Perception du pitch des notes (fréquences)

On entend seulement les fréquences  $f$  comprises dans l'intervalle

$$f \in [20 \text{ Hz}, 20000 \text{ Hz}] .$$

Cela dépend des animaux [Benson \(n.d., p.9\)](#).

Pour des notes pures (i.e. signal périodique sinusoïdal qui est un mode de Fourier) la précision de perception de la fréquence  $f$  est en unité de pitch

$f$ (Hz)	250	500	1000	2000	8000
$\Delta x$ (pitch)	0.17	0.03	0.05	0.04	0.16

Cela est extrêmement précis, largement inférieur à la précision en fréquence des cils, voir Section [3.1.2](#), par exemple à  $f = 1000 \text{ Hz}$  :

$$\Delta x = \frac{1}{40} (\Delta x)_{cils}$$

On pense que cette grande précision de perception du pitch est obtenue grâce à des algorithmes dans le cerveau, non encore observés ni compris.

*Remarque 3.3.6.* La précision de perception du pitch indiquée dans le tableau précédent montre que l'échelle des notes tempérées, Définition [2.3.37](#), écartées de  $\Delta x = 1$  pitch, est très grossière par rapport à notre perception.

*Remarque 3.3.7.* A cause de cette grande précision en pitch de notre perception, la fabrication d'un instrument de musique demande une grande précision, inférieure à  $10^{-2} \text{ mm}$  [Hirschberg et al. \(1996\)](#). (ref : [article 1996, Aeroacoustics of Musical Instruments](#))

### 3.3.5 Perception et principe d'incertitude en temps-fréquence

Nous avons mentionné en Section [??](#), le principe d'incertitude en temps fréquence, concernant la quantité d'information limitée par la mesure  $dt df$ .

Par exemple, considérons une note  $C_5$  de pitch  $x = 60$  avec une précision  $\Delta x = 0.5$  (quart de ton) soit  $\frac{\Delta f}{f} = 0.057 \Delta x$ , et de durée  $\Delta t = 0.2 \text{ s}$ , supérieure au plus petit intervalle de temps perceptible ([3.3.1](#)). Cela donne  $f = 260 \text{ Hz}$  et

$$\Delta t \Delta f = 0.2 \times 260 \times 0.057 \times 0.5 \sim 1.5 \geq 1$$

juste supérieure à 1 donc à la limite du principe d'incertitude. Cela montre que notre perception de l'information dans le plan temps-fréquence est quasi optimale.

La formule générale en fonction du pitch  $x$  est

$$\Delta t \Delta f \stackrel{(2.3.7)}{=} 0.057 \times \Delta t \Delta x 2^{\frac{1}{12}(x-x_{A_5})} f_{A_5}$$

et cela montre que pour les notes graves, i.e.  $x$  petit, la limite  $\Delta t \Delta f = 1$  est plus facilement atteinte. Par exemple il est impossible en principe de détecter le pitch précis (au demi-ton) des notes d'une contrebasse qui joue un **walking bass** rapide.

### 3.3.6 Non perception de la phase

La transformée de Fourier d'un signal périodique  $u(t)$  donne des coefficients de Fourier complexes  $\hat{u}_n \in \mathbb{C}$ , en (??) que l'on peut écrire

$$\hat{u}_n = |\hat{u}_n| e^{i\phi_n} \tag{3.3.2}$$

avec une phase  $e^{i\phi_n}$ . D'après (??), le signal temporelle s'écrit alors

$$u(t) = \sum_{n \in \mathbb{Z}} \left( \frac{\hat{u}_n}{\sqrt{2\pi}} \right) e^{i2\pi n \frac{t}{T}} = \sum_{n \in \mathbb{Z}} \left( \frac{|\hat{u}_n|}{\sqrt{2\pi}} \right) e^{i(2\pi n \frac{t}{T} + \phi_n)}$$

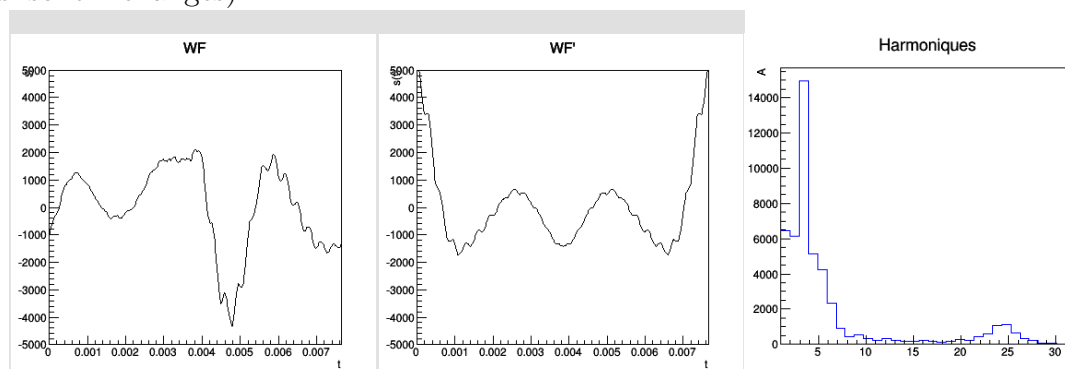
montrant que la phase induit un décalage temporel spécifique à chaque harmonique  $n$ .

**Proposition 3.3.8. “Non perception de la phase”.** Dans les exemples suivants on observe que nous ne percevons pas du tout les phases  $(\phi_n)_n$  de (3.3.2), autrement dit on ne distingue pas le signal  $u$  du signal suivant

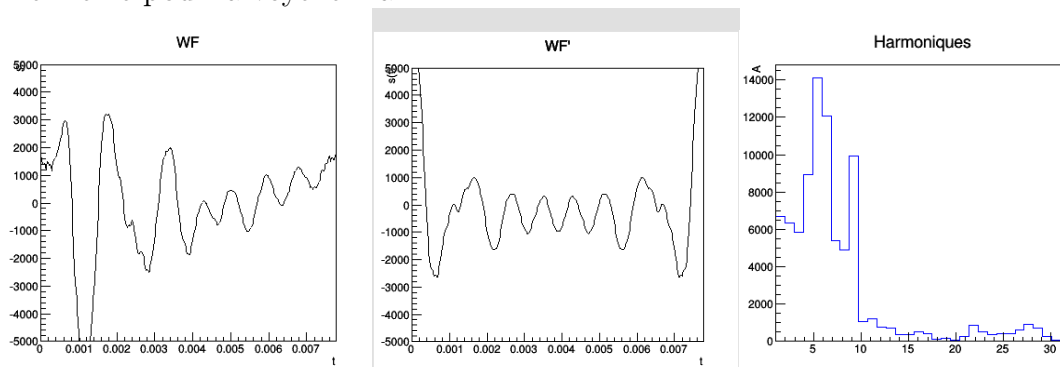
$$v(t) = \sum_{n \in \mathbb{Z}} \left( \frac{|\hat{u}_n|}{\sqrt{2\pi}} \right) e^{i(2\pi n \frac{t}{T} + \varphi_n)}$$

qui a les même modules  $|\hat{u}_n|$  mais des phases  $(\varphi_n)_n$  différentes et arbitraires.

**Exemple 3.3.9.** Voici un **Son**  $u(t)$ . Voici maintenant un **Son**  $v(t)$  où toutes les phases sont mises à  $\varphi_n = 0, \forall n$ . Voici pour la voyelle « on », les formes d'ondes d'origine  $u(t)$  et  $v(t)$  sur une période, ainsi que les modules des coefficients de Fourier  $|\hat{u}_n|$  pour  $0 \leq n \leq 32$  (qui sont inchangés) :



De même pour la voyelle « a » :



On remarque que  $u(t)$  et  $v(t)$  sont très différentes. En particulier  $v(t)$  a un pic en  $t = 0$  à cause du choix particulier  $\varphi_n = 0$ .

**Exemple 3.3.10.** Voici un **autre exemple** de son artificiel  $u(t)$ , où chaque harmonique  $n = 1, 2, \dots, 30$  a un module  $|\hat{u}_n| = 1/n$  et une **phase**  $\phi_n$  **aléatoire**. On représente aussi le signal  $u(t)$  sur une période. Remarquer que lorsque les phases sont nulles, le signal  $u(t)$  est en « dent de scie ». On le perçoit légèrement, peut être à cause de la « singularité » en zéro qui fausse le rendu.

*Remarque 3.3.11.* Cette non perception des phases  $\phi_n$  suggère que parmi toute l'information du signal, le cerveau perçoit des notes pures quasi-indépendantes  $|\hat{u}_n| e^{i(2\pi n \frac{t}{T} + \phi_n)}$  et mesure seulement leur fréquence  $f_n = \frac{n}{T}$  et intensité  $|\hat{u}_n|$ .

**Exercice 3.3.12.** « La phase est imperceptible » (TP) Comme dans l'exercice 1, avec python, créer le signal suivant qui est la superposition de deux sinus de fréquences respectives  $f_1$  et  $f_2$ , avec une phase  $\varphi$  :

$$u(t) = A_1 \sin(2\pi f_1 t) + A_2 \sin(2\pi f_2 t + \varphi)$$

1. Fixer  $f_1 = 100\text{Hz}$ ,  $f_2 = 150\text{Hz}$ ,  $A_1 = A_2 = 0.5$  par exemple et  $\varphi = 0$  puis observer et écouter le signal. Recommencer avec une autre valeur de  $\varphi$ , par exemple  $\varphi = 1$ . Qu'est ce qui a changé? est-ce perceptible à l'écoute? Si vous changez  $A_1, A_2$  ou  $f_1, f_2$  est-ce perceptible?
2. (\*) Faire que la phase  $\varphi$  change lentement avec le temps, par exemple :  $\varphi(t) = 10t$  (ainsi  $\Delta\varphi = 1$  radian pour  $\Delta t = 0.1\text{sec.}$ ). Observer et écouter.

### 3.3.7 Perception du timbre

On a observé dans la Section précédente, que en présence d'un signal périodique  $u(t)$ , appelé **note** (d'après la Définition 2.3.1), nous ne percevons pas la phase des coefficients de Fourier. Cependant que perçoit on? On ne perçoit pas individuellement chaque harmonique mais un effet collectif de  $|\hat{u}_n|_{n \geq 0}$ , que l'on appelle le **timbre** et la fréquence fondamentale  $f_1 = \frac{1}{T}$ , équivalente à un pitch  $x = x_1$  d'après (2.3.5) appelé le **pitch de la note**. Dans cette section, on discute de la perception du timbre et du pitch  $x$ .

### 3.3.7.1 Le timbre et les voyelles

**Définition 3.3.13.** Pour un signal périodique  $u(t)$ , (appelé **note** d'après la Définition 2.3.1), de pitch  $x \in \mathbb{R}$ , (2.3.5), le **timbre** est la suite des modules

$$(|\hat{u}_n|)_{n \geq 1} = (|\hat{u}_1|, |\hat{u}_2|, |\hat{u}_3|, \dots)$$

i.e. le module des coefficients de Fourier de  $u(t)$ , (??), sans la phase que l'on ne perçoit pas d'après la proposition 3.3.8.

Dans la voix parlée, le timbre est associé à des voyelles, voir l'exemple 3.3.9.

**Question 3.3.14.** Comment associer une suite  $(|\hat{u}_n|)_{n \geq 1}$  à une certaine voyelle ?

**Observations :** Considérons cet **enregistrement** d'un signal  $u(t)$  de voyelles différentes, i.e. timbres différents mais de même pitch  $x = 48$ , i.e.  $f = 130\text{Hz}$ , la note est  $C_4$ .

— Dans cet **exemple**, le signal  $u(t)$  est filtré, ne gardant que avec les harmoniques 1,2,3 seulement, i.e. on entend le signal

$$v(t) = \sum_{n=1,2,3} \left( \frac{\hat{u}_n}{\sqrt{2\pi}} \right) e^{i2\pi n \frac{t}{T}}$$

On constate que **on ne perçoit plus les voyelles**. On perçoit toujours le même pitch  $x$ .

— Dans cet **exemple**, le signal  $u(t)$  est filtré, ne gardant que avec les harmoniques  $n \geq 5$ , i.e. on entend le signal

$$v(t) = \sum_{n \geq 5} \left( \frac{\hat{u}_n}{\sqrt{2\pi}} \right) e^{i2\pi n \frac{t}{T}}$$

On constate que **l'on perçoit les voyelles**. De façon plus surprenante, on perçoit toujours le même pitch  $x$ , i.e. la fréquence  $f_1$  alors qu'elle n'est pas présente dans la transformée de Fourier. La perception des voyelles est surtout liée aux formants expliqués en Section 3.2.1.2.

### 3.3.7.2 Paradoxe de la basse virtuelle

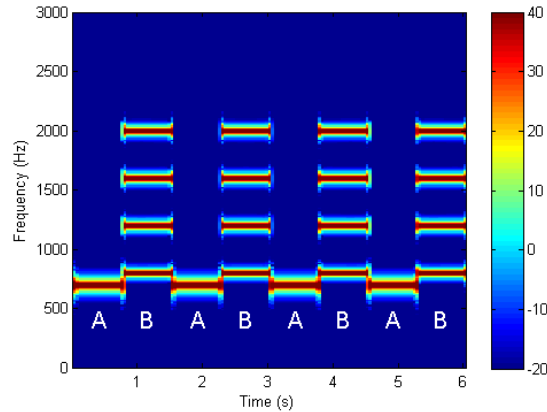
La conclusion importante de observations précédentes est que :

*Conclusion 3.3.15.* Lorsqu'on entend un signal périodique, on perçoit le timbre

$$(|\hat{u}_n|)_{n \geq 1} = (|\hat{u}_1|, |\hat{u}_2|, |\hat{u}_3|, \dots)$$

et le pitch  $x$  même si les premiers coefficients sont nuls, par exemple  $|\hat{u}_1| = 0, |\hat{u}_2| = 0, |\hat{u}_3| = 0, |\hat{u}_4| = 0$ .

**Exemple 3.3.16.** Dans cet exemple on entend successivement deux notes  $A, B$ . La note  $A$  est une note pure de une fréquence fondamentale  $f_1(A) = 700\text{Hz}$ . La note  $B$  est une note timbrée avec  $f_1(B) = 400\text{Hz}$ , mais pour la note  $B$  on a enlevé l'harmonique  $f_1(B) = 400\text{Hz}$  (en posant  $\hat{u}_1 = 0$ ) de sorte que la fréquence la plus basse émise est  $f_2(B) = 800\text{Hz}$ . A l'écoute on entend  $f_1(B) = 400\text{Hz}$  qui n'est pas présente. Remarquer que  $f_1(A)/f_1(B) = 7/4$  de pitch  $x_1(A) - x_2(B) = \frac{12}{\ln 2} \ln\left(\frac{7}{4}\right) = 9.68 \sim 10$  qui est une septième mineure.



**Exercice 3.3.17.** (TP) « Illusion de la basse fondamentale manquante ». Avec audacity Menu/Générer/Tonalité, créer différents signaux sinusoïdaux d'amplitude  $A = 0.1$  et de fréquences  $f_1 = 200\text{Hz}$ ,  $f_5 = 1000\text{Hz}$ ,  $f_6 = 1200\text{Hz}$ ,  $f_7 = 1400\text{Hz}$ . Remarquer que ce sont des harmoniques de la fréquence fondamentale  $f_1$ . Ecouter simultanément les signaux de fréquence  $f_5, f_6, f_7$ . Entendez vous la fréquence  $f_1$ ? Cela s'appelle le **mystère de la fondamentale manquante** actuellement pas très bien compris. On pense que le mécanisme est au niveau du cerveau, et serait utile pour analyser les stimuli venant d'une voix humaine incomplète (dans la nature, la fondamentale peut être brouillée par d'autres bruits).

Remarque : le phénomène du « troisième son » ou **Combination tone** est différent : il fait ressort  $f_2 - f_1$  (effets non linéaire dans l'oreille?). Ecouter par exemple **ex. de 3eme son** avec  $f_1 = 440 (A_5)$ ,  $f_2 = 704 (F_6)$ , on perçoit  $f_2 - f_1 = 264 (C_5)$ , mais ici la basse manquante serait  $f = f_2/8 = f_1/5 = 88 (F_4)$ .

### 3.3.7.3 Hypothèse d'une reconnaissance inconsciente des rapports rationnels de fréquences

Ce phénomène est assez paradoxale car d'après la structure de l'oreille interne, on pense que l'information qu'il y a un signal sur les fréquences  $f_n = n f_1$  des différentes harmoniques est envoyé de façon séparée au cerveau par des neurones. Mais cependant le cerveau arrive à détecter que ces différentes fréquences sont multiples d'une fréquence  $f_1$  pourtant absente. Comment le fait-il? En arithmétique, on utiliserait le fait que étant donné  $f_5 = 5 f_1$ ,  $f_6 = 6 f_1$ ,  $f_7 = 7 f_1$  on peut retrouver  $f_1$  en observant que  $f_6/f_5 = 6/5$  et  $f_7/f_5 = 7/5$  sont des (petits) nombres rationnels et supposer donc que  $f_1 = f_5/5 = f_6/6 = f_7/7$ .



**Hypothèse 3.3.18.** Une hypothèse avancée par des neurologues est que le cerveau procède ainsi, grâce à des algorithmes inconscients de reconnaissance de rapports rationnels

$$\frac{f'}{f} = \frac{a}{b}, \quad a, b \in \mathbb{N} \text{ "petits" i.e., } a, b \leq 11, 13 \quad (3.3.3)$$

entre les fréquences des signaux perçus.

On peut imaginer que cette capacité soit arrivée par adaptation évolutive, car serait bien utile pour détecter une voix humaine parmi d'autres voix ou parmi du bruit.

*Résumé 3.3.19.*

- A la sortie des oreilles (oreille interne), et à l'entrée du cerveau, on a effectué (inconsciemment) une décomposition du son en "notes pures" (mode de Fourier) et on ne retient que les modules  $|\hat{u}_n|$ .
- Ensuite le cerveau effectue un "traitement complexe" de cette information. [Schnupp et al. \(2011, p.138\)](#) (dans le cerveau il y a un processus de reconnaissance de pattern arithmétique?)
- Au final on est conscient de l'ensemble du spectre : "le timbre" et son "pitch" (fondamentale), et non pas des fréquences individuelles.

#### 3.3.7.4 Fréquences aléatoires :

Voici un exemple artificiel, où des notes pures sont d'amplitude  $A_k = 1/k$ , de phase nulle, et de fréquence  $f_k = k \cdot f_1 \cdot (1 + \epsilon_k)$  avec  $f_1 = 130\text{Hz}$  (do) et  $\epsilon_k \in [-\epsilon, \epsilon]$  aléatoire. On [écoute et représente](#)  $(f_k, A_k)$  pour différentes valeurs de  $\epsilon$  qui mesure la fluctuation des fréquences. Remarquer que l'on perçoit parfois plusieurs notes (par exemple pour  $\epsilon = 0.084$ ). Remarquer que les pics de fréquences  $f_k, f_{k+1}$  peuvent se croiser si  $\epsilon > 1/k$ .

Est-ce en rapport avec l'expérience suivante ?

#### Y. Menuhin, racontant un étrange concert dans un champ de fleurs à Mykonos :

Référence : "La légende du violon", Y Menuhin, Flammarion

« Je m'arrêtai pour écouter ce bourdonnement entêtant quand, soudain, je m'aperçus qu'il ne s'agissait pas d'un bruit confus et désordonné, mais bien au contraire d'un accord secret de la création : les abeilles émettaient deux notes, et qui plus est ces deux notes formaient une quinte. Je ne sais comment analyser ce prodige de la nature. Toujours est-il que la recherche de la quinte semble constituer une nécessité primitive de l'univers : nous émettons et nous recevons des vibrations. Ces vibrations obéissent à un ordonnancement secret. Une harmonie occulte se met en place. Certains prétendent même que les intervalles des planètes, c'est-à-dire les distances qui les séparent les unes des autres, sont en correspondance intime avec les intervalles de la musique : quintes, tierces, octaves ne seraient que des images musicales des intervalles célestes. »

## 3.4 Perceptions des intervalles justes et accords justes

Voir la section B.1.2 en appendice, concernant la représentation de fractions rationnelles  $\frac{a}{b} \in \mathbb{Q}$ .

### 3.4.1 Perception des intervalles justes

Comme les rapports de fréquence rationnels  $\frac{f'}{f} = \frac{a}{b}$  en (3.3.3) correspondent aux intervalles juste, voir Section ??, l'hypothèse (3.3.3) signifierait que nous percevions particulièrement les intervalles justes.

- **Rappel** : pour une note individuelle, i.e. un signal périodique provenant d'une voix ou d'un instrument, on ne perçoit pas ses harmoniques individuelle, mais on perçoit le timbre  $|\hat{u}_n|_{n \geq 1}$  qui donne une impression collective, interprétée par le cerveau.
- **Question** : que perçoit-on si deux notes (timbrées) de fréquences  $f_1 \leq f_2$  sonnent ensemble ?

Voici quelques exemples.

#### 3.4.1.1 Exemples

- **Video** qui montre une résonance à l'unisson  $\frac{f_2}{f_1} = \frac{1}{1}$ .
- **Video** qui montre une résonance à l'octave  $\frac{f_2}{f_1} = \frac{2}{1}$ .
- **Video** qui montre une résonance à la quinte  $\frac{f_2}{f_1} = \frac{3}{2}$ .
- **Video** qui montre les résonances  $\frac{f_2}{f_1} = \frac{a}{b}$  importantes.
- **Video** qui montre les résonances  $\frac{f_2}{f_1} = \frac{a}{b}$  dans l'ordre descendant.
- **Video** qui montre les résonances  $\frac{f_2}{f_1} = \frac{a}{b}$  avec le rajout de la “basse”  $f_b = \frac{f_2}{a} = \frac{f_1}{b}$ . (la fondamentale manquante en rouge)
- **Video** qui montre les résonances  $\frac{f_2}{f_1} = \frac{a}{b}$  avec la “basse”  $f_b = \frac{f_2}{a} = \frac{f_1}{b}$  en rouge et le “sifflet”  $f_s = af_1 = bf_2$  en vert.

*Résumé* 3.4.1. un **intervalle juste** est un ensemble de deux notes dont le rapport des fréquences est  $f_2/f_1 = a/b$  avec  $a, b$  petits entiers. On ressent cet intervalle comme “résonant” ou “consonant”.

#### 3.4.1.2 Perception des intervalles justes

Sur cette [page web](#).

### 3.4.2 Perception des accords justes

Voir la Section B.1.3 qui définit le réseau tonnetz, permettant de représenter les accords justes.