Son et Mathématiques

Maïtine Bergounioux

MAPMO - UMR 6628 - Département de mathématiques -
Université d'Orléans - BP 6759 - 45067 Orléans cedex 02, France
maitine.bergounioux@univ-orleans.fr

17 février 2005

Résumé

Nous montrons comment l'analyse de Fourier est utilisée en traitement du son

1 Problématique

Le traitement du son prend une part de plus en plus importante dans notre environnement et notre vie quotidienne. On peut se poser de nombreuses questions relatives à la perception, la restitution ou la transmission du son, comme par exemple :

- Pourquoi mon voisin n'a-t-'il pas la même voix que moi ?
- Pourquoi les mêmes notes jouées par un violon, une trompette ou un piano sont-elles différentes ?
- Pourquoi la voix de mon correspondant est-elle différente au téléphone ?
- Comment se déplace la chauve-souris ?
- Quel est le principe d'un enregistrement numérique ?
- Comment puis-je envoyer ou recevoir de la musique par internet ?
- etc .

On peut répondre partiellement à ces questions en considérant que le son est ce que l'on appelle un « signal » et qu'on peut l'analyser, le traiter et le modifier grâce à des techniques mathématiques particulières.

1.1 L'aspect physique du son

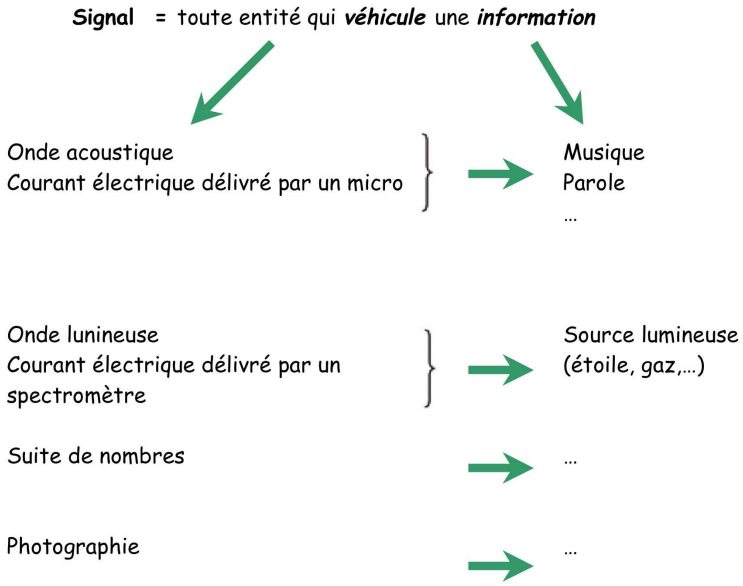
Le son est en fait une conséquence d'un mouvement matériel d'oscillation, une corde qui vibre ou la membrane d'un haut-parleur par exemple. Cette vibration provoque un mouvement des atomes l'avoisinant qui va se déplacer de proche en proche sous forme d'onde de pression. Dans ce mouvement, les atomes vibrent parallèlement à la direction de propagation de l'onde. C'est donc une onde progressive longitudinale. Parmi les ondes de nature mécaniques, seules les longitudinales peuvent se propager relativement loin dans un milieu gazeux. Ce qui nous permet, entre autres, d'entendre ce que notre interlocuteur nous dit. Dans le vide, le son ne peut se propager faute d'atomes autour de la source de vibration, aucune onde mécanique ne peut donc se créer.
En résumé :

– Une source sonore crée, dans l'espace, des zones de surpression et de dépression (pression sonore) qui varient dans le temps à la même fréquence et avec la même forme que cette source. Cette modification de pression se déplace à la vitesse de 340 m/s dans l'air.
– L'oreille est sensible à la pression sonore, et les caractéristiques perceptives de *hauteur, intensité, timbre* et *durée* sont étroitement liées aux paramètres physiques qui définissent la source sonore, c'est-à-dire respectivement la *fréquence, l'amplitude, la forme* et *le temps* de la vibration.

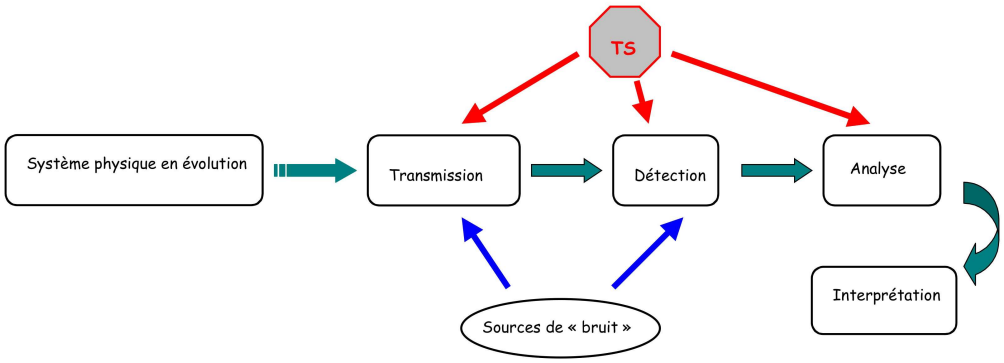
Le signal sonore est donc un signal unidimensionnel (une onde qui se propage dans une seule direction) capté et analysé par l'oreille et le cerveau. Les outils que nous allons présenter permettent de reproduire cette analyse « artificiellement » . Ainsi, on peut plus particulièrement étudier la voix parlée ou chantée (signal vocal).

1.2 Qu'est -ce que le traitement du signal ?

On appelle signal toute entité qui véhicule une information :



Le traitement du signal est une procédure pour extraire cette information (filtrage, détection, estimation, analyse spectrale...)



On peut aussi modifier, mettre en forme le signal (modulation, échantillonnage...) pour pouvoir le stocker (CDs audio ou vidéo) ou le transmettre (fichiers « sons » sur Internet).
Présentons maintenant un outil mathématique essentiel en traitement du signal : l'analyse de Fourier

2 L'analyse de Fourier

Comme le signal sonore que nous voulons traiter est unidimensionnel nous supposerons désormais que toutes les fonctions (signaux) que nous considérons sont à une seule variable (réelle ou complexe). On peut bien sûr adapter tout ce qui va suivre à des signaux bi- ou tri-dimensionnels : le principe est le même mais la technique est un peu plus compliquée.

2.1 Série de Fourier

Considérons un signal uni-dimensionnel (son par exemple) x_T signal périodique, de période T.

Remarque 2.1 *il faut savoir qu'un signal vocal qui semble ne pas être périodique (on ne dit pas toujours la même chose) est en fait* ***localement*** *périodique : si on zoome en temps (par exemple si on isole une voyelle sur une durée de 20 ms) le signal est alors clairement périodique. Nous reviendrons, un peu plus loin sur cette technique de « zoom » .*

On peut alors affirmer (moyennant certaines hypothèses liées à la régularité du signal et vérifiées en pratique) que x_T est égale à sa série de Fourier :

$$x_T(t) = a_o + \sum_{n=1}^{+\infty} a_n \cos(2\pi n f t) + \sum_{n=1}^{+\infty} b_n \sin(2\pi n f t)$$

$$\text{avec } \; a_o = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x_T(t)\,dt \quad \textit{valeur moyenne de } x \text{ et}$$

$$a_n = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x_T(t)\cos(2\pi n\, f\, t)\,dt \text{ et } b_n = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x_T(t)\sin(2\pi n f t)\,dt$$

C'est donc la superposition de signaux sinusoïdaux de fréquences $n\,f$, où n est entier et $f = \frac{1}{T}$ est la fréquence ou **fréquence fondamentale**. Le terme de fréquence $n \cdot f$ est la nième **harmonique**.

On peut écrire également la série de Fourier en notation complexe :

$$x_T(t) = \sum_{n=-\infty}^{+\infty} c_n\, e^{2i\pi n f t}$$

$$\forall n \in \mathbb{Z} \qquad c_n = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x_T(t)\, e^{-2i\pi n f t}\,dt.$$

On constate donc qu'on peut décrire un signal soit par son expression *temporelle* (valeur du signal x_T à l'instant t) soit par son expresion *fréquentielle* c'est-à-dire sous forme de série de Fourier. Chaque harmonique $n\,f$ est affectée d'un coefficient (complexe) c_n dont on peut tirer beaucoup d'informations.

On peut généraliser ce qui précède à un signal quelconque x non nécessairement périodique, mais de carré intégrable sur $\mathbb{R}$.

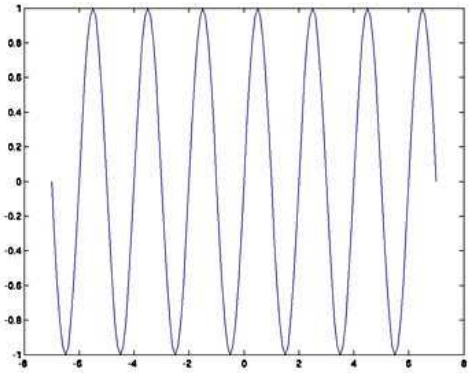
$$\mathcal{F}\{x(t)\} = \hat{x}(\lambda) \stackrel{def}{=} \int_{-\infty}^{+\infty} x(t)\, e^{-2i\pi\lambda t}\,dt$$

$$x(t) = \mathcal{F}^{-1}\{\hat{x}(\lambda)\} = \int_{-\infty}^{+\infty} \hat{x}(\lambda)\, e^{2i\pi\lambda t}\,d\lambda$$

En pratique, on a besoin à la fois de l'information en temps et en fréquence ; on fait alors de l'**analyse temps-fréquence**. Le principe consiste à déplacer une fenêtre temporelle sur le signal temporel (il faut imaginer un rectangle qui se déplace le long du signal et permet de zoomer), et on fait l'analyse fréquentielle sur la portion de signal ciblée. En effet, la transformée de Fourier tout comme la série de Fourier, utilise des valeurs moyennes sur l'intervalle de temps. On n'a donc qu'une information globale en temps. Le fait de passer par des fenêtres (de petite taille) qui se déplacent le long de l'axe temporel permet de réduire l'intervalle temps de manière significative et donc de compenser les effets de la moyennisation.

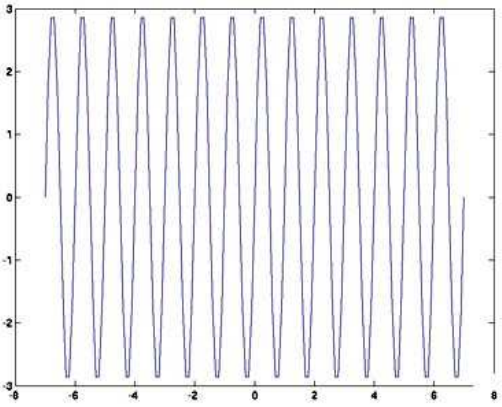
Principe de la superposition d'ondes élémentaires

$$sin(2\pi \cdot 0.5 \cdot x) + 3sin(2\pi \cdot 1 \cdot x) - 5sin(2\pi \cdot 1.5 \cdot x)$$

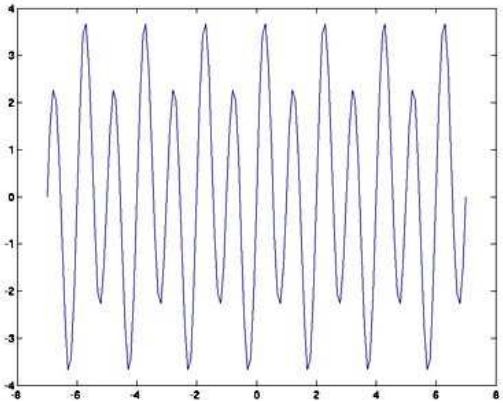


Fondamentale

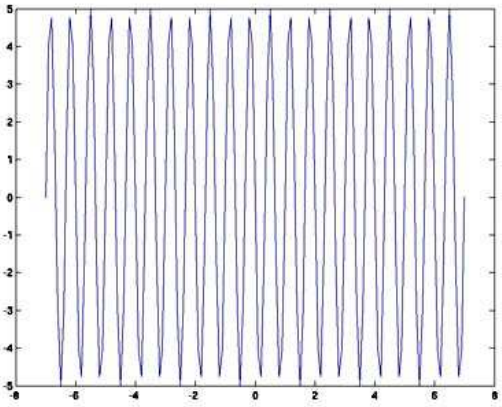
\+ Première harmonique



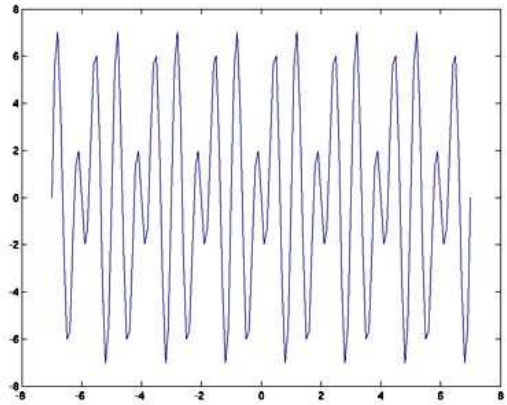
=



\+ Deuxième harmonique



=



2.2 Échantillonnage, FFT et calcul effectif de $\hat{x}$

Quand on veut calculer les coefficients de la série de Fourier du signal x, on doit calculer des intégrales. Comme on ne connaît pas la forme « analytique » de x il est exclus de calculer cette intégrale autrement que par des méthodes numériques. Celles-ci imposent le choix d'un certain nombre de valeurs de x, par exemple à des instants t_i où $1 \leq i \leq N$. Les valeurs $x_i = x(t_i)$ sont physiquement mesurables au cours du temps. On dit qu'on **échantillonne** le signal : au lieu de considérer le signal original, continu (on dit aussi **analogique**), on considère un (grand) nombre de valeurs (ou **échantillons**) x_i prises par le signal : on obtient alors un signal **numérique**.

Le choix des échantillons ne se fait pas au hasard : il faut en prendre suffisamment et pour un signal périodique, leur nombre N est deux fois l'inverse de la plus grande fréquence du signal (qui est supposé à spectre borné) : c'est le théorème de Shannon.

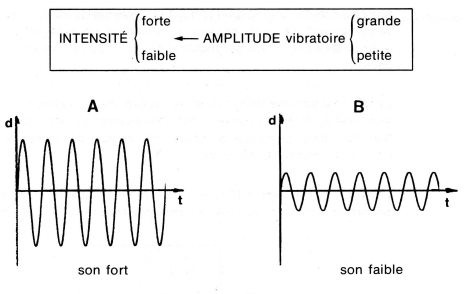
Une fois qu'on a les échantillons on peut calculer les coefficients de Fourier : toutefois, on ne le fait pas en calculant les intégrales mais en utilisant un algorithme célèbre et très performant, la FFT (*Fast Fourier Transform* ou Transformée de Fourier Rapide) qui permet de calculer très rapidement ces coefficients.

3 Le signal sonore (et vocal) : les « marqueurs » essentiels

3.1 L'amplitude du signal = intensité du son

L'**intensité** (ou force) des sons dépend de leur puissance sonore (celle-ci étant proportionnelle au carré de la pression sonore, elle-même proportionnelle à l'**amplitude**. Le son est **fort** ou **faible** suivant que la puissance sonore est élevée ou non. L'intensité d'un son s'exprime en dB (décibel), unité logarithmique de rapport.

La gamme des intensités sonores s'étend, en dB absolus, de 0 (seuil d'audibilité) à 140 (seuil de douleur). On peut aussi parler d'intensité sonore relative entre deux sons (tel son est 10 fois plus intense que tel autre). En ce cas, le dB n'est pas une unité fixe, mais relative (dB relatif) : il indique un rapport (ex. un son 100 fois plus intense qu'un autre est 20 dB plus intense).



L'intensité est déterminée en première approximation par l'intensité de **l'harmonique le plus intense.**

Notre oreille est un organe sensible aux ondes, ces ondes font entrer en vibration des cils se trouvant dans nos conduits auditifs. Ces vibrations sont captées par de petits muscles et par la voie des nerfs, transmises jusqu'au cerveau, où elles seront décodées et enfin « entendues ». Notre oreille n'est pas sensible de la même manière aux sons de toutes les fréquences. Nous n'entendons que les sons compris entre grosso modo 20 et 20 000 Hz. En dessous de 20 Hz, ce sont des infrasons et au dessus de 20 000 Hz, des ultrasons. La sensibilité de notre oreille est à son maximum pour les fréquences comprises entre 500 Hz et 5000 Hz. Toutefois la sensation de volume ne varie pas comme la puissance sonore réelle. C'est pourquoi, le décibel a été créé, pour que le rapport entre notre sensation et la mesure du volume sonore soit linéaire, c'est-à-dire, pour qu'un son qui sonne deux fois plus fort qu'un autre ait une mesure deux fois plus grande que celle de l'autre. Nous pouvons aussi comprendre que si 10 instruments jouent à un volume de 40dB, il en faudra 100 pour obtenir un volume de 50dB. C'est aussi pour cela que quand vous tournez le bouton de volume de votre chaîne stéréo, vous devez le tourner de plus en plus pour entendre une différence de volume.

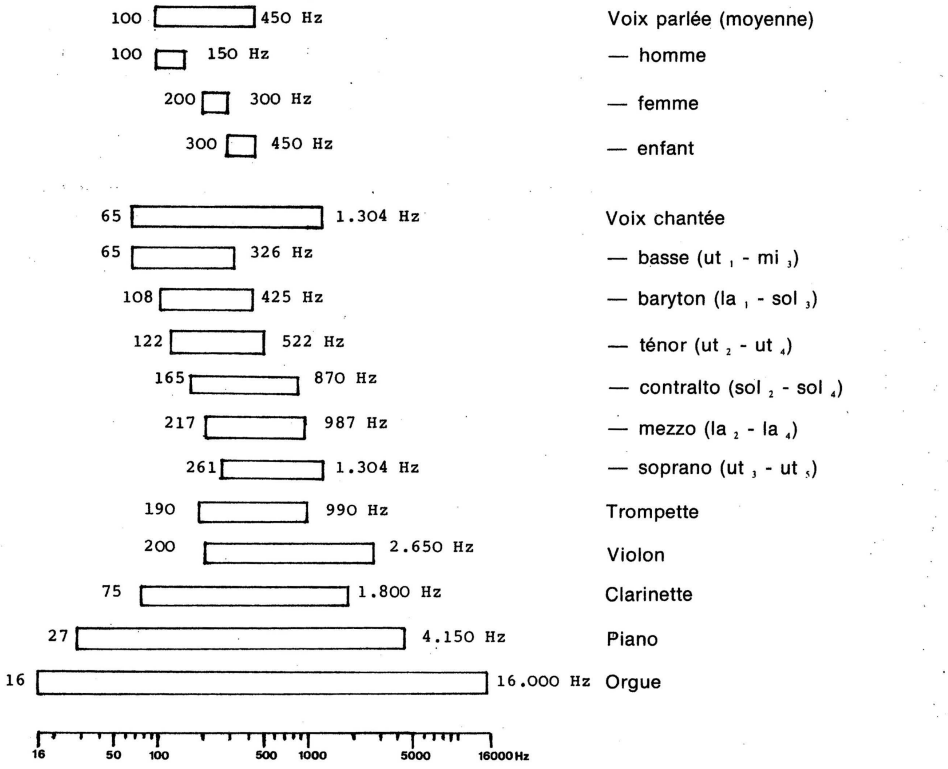
Notre oreille est fragile, une exposition à un son de plus de 120dB, même pendant un court instant, peut entraîner des lésions irréversibles sur notre système auditif. Voici un tableau explicatif des sensations et effets sur nos oreilles de bruits trop importants.

Echelle des niveaux sonores

Niveau	Impression ressentie	Effets	Exemples
140 dB	Très douloureuse	Lésions irréversibles du système auditif	Banc d'essai de réacteur
130 dB			Avion au décollage
120 dB	Douloureuse		Burin pneumatique
110 dB	Insupportable	Perte d'audition après une exposition brève	Atelier de presse
100 dB	Difficilement supportable		Atelier de tolerie
90 dB	Très bruyant	Perte d'audition après une expostion longue	Poids lourd à 3 mètres
80 dB	Bruyant		Réfectoire scolaire
60 dB	Bruit courant		Rue bruyante
50 dB			Bureau tranquille, aspirateur
40 dB	Faible		Radio à faible niveau voix chuchotée
30 dB	Calme		Zone résidentielle calme
20 dB	Très calme		Pièce très protégée
10 dB	Silence	L'observateur entend le bruit de son organisme	Désert
0 dB	Silence absolu		Ne peut être obtenu qu'en laboratoire

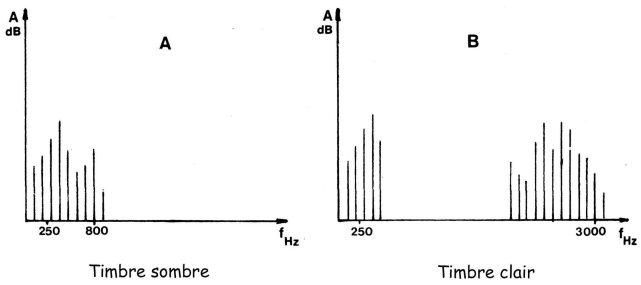
3.2 La fréquence fondamentale : « Hauteur » du son

La **hauteur** des sons dépend de leur **fréquence**. Plus la fréquence est élevée, plus le son paraît **aigu** ; plus elle est basse, plus le son paraît **grave**. La gamme des fréquences audibles se situe entre 16 et 16.000 Hz ; elle est divisée en 10 octaves.



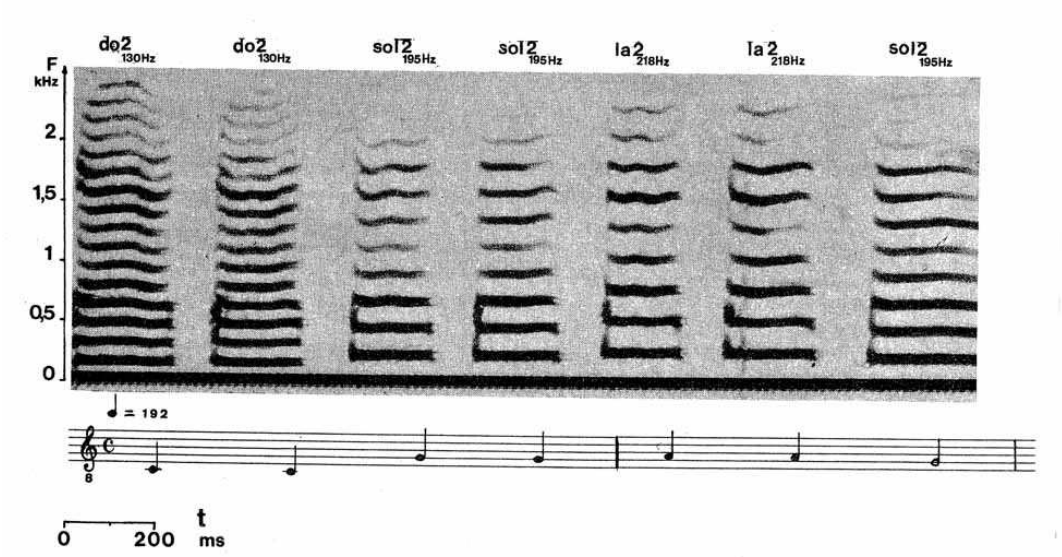
3.3 La répartition des harmoniques = timbre de la voix ou de l'instrument

La représentation du timbre de la voix est obtenue par un **sonagramme**, qui permet d'afficher dans le temps la décomposition d'un son sur les fréquences qui le composent (analyse temps/fréquence) . Le **timbre** est déterminé par la **densité** relative des harmoniques. Il est qualifié de **clair** si cette répartition est essentiellement située dans les *hautes fréquences* et de **sombre** si elle l'est dans les *basses fréquences.*

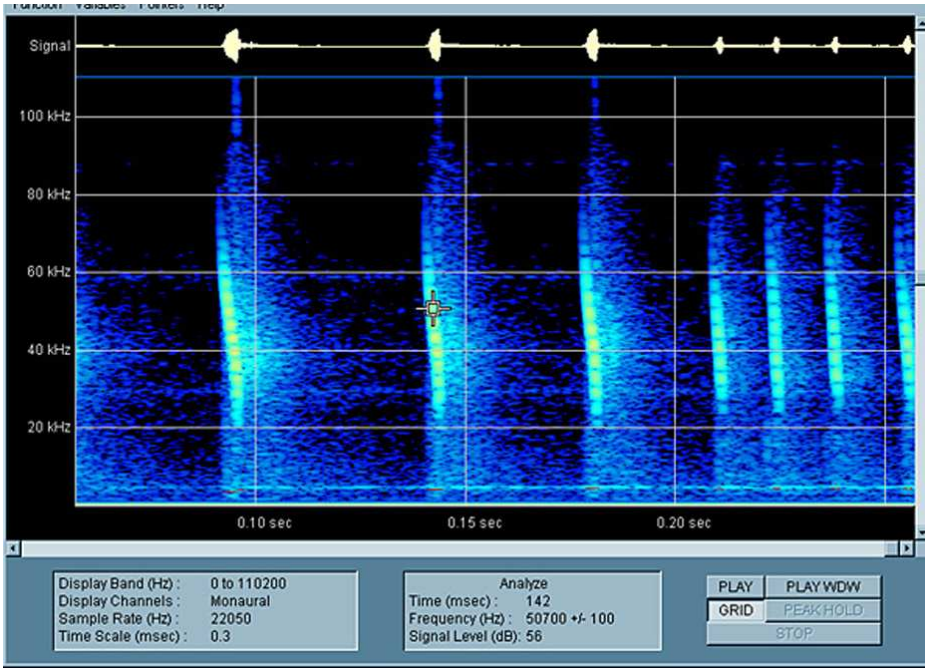


Exemples de sonagrammes

Une mélodie

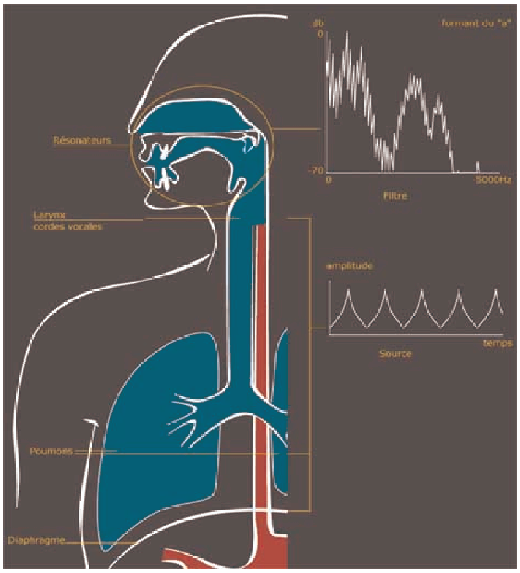


Une chauve-souris



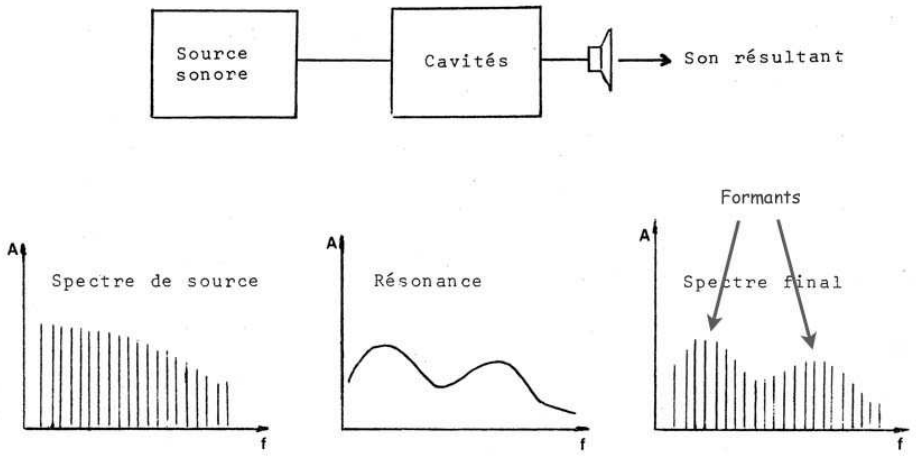
3.4 Les formants

Ce sont les harmoniques (c'est-à-dire des fréquences) dont les amplitudes sont les plus élevées. Elles caractérisent les sons prononcés.

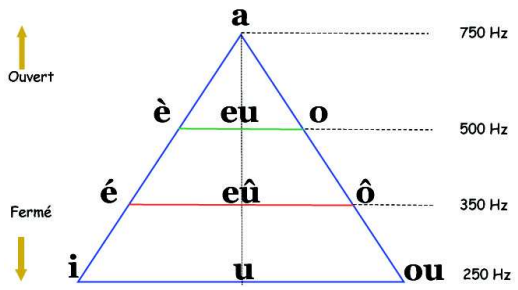


L'air est émis par les poumons puis est mis en vibration par les cordes vocales : cela donne un son (sous forme d'onde complexe). L'onde est modifiée par son passage dans des résonateurs (cavités nasales, palais, bouche, lèvres, etc...)

Appareil phonatoire



Principe de la formation des formants par résonnance



Premier formant des voyelles principales

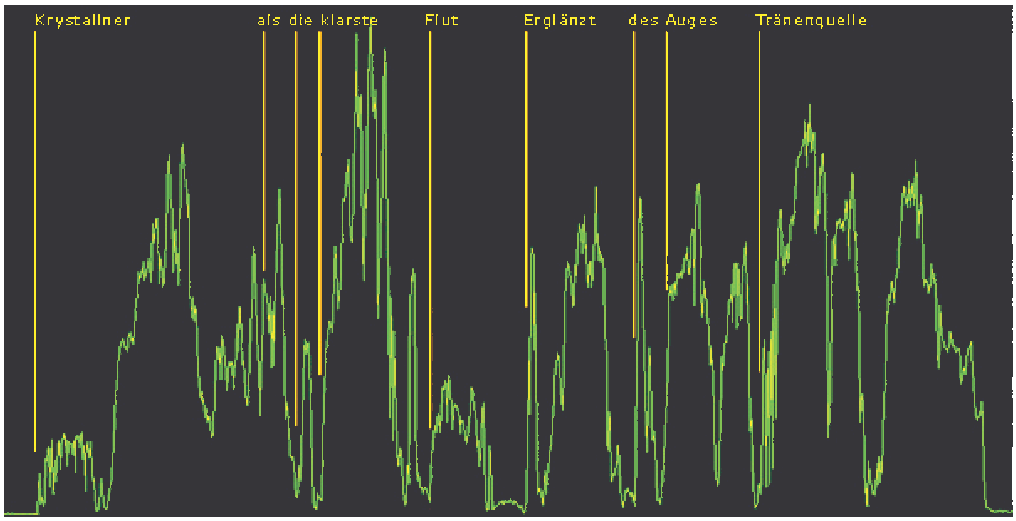
4 L'analyse du son

Notre oreille analyse les sons et leurs variations en permanence, notamment leur timbre ou couleur. Par exemple, nous trouvons que la voix au téléphone prend un aspect « métallique » ou alors nous ne reconnaissons plus un interlocuteur fortement enrhumé. Très souvent, nous tentons d'identifier l'origine du son, et si possible, de lui trouver une ressemblance avec un son que nous connaissons déjà.

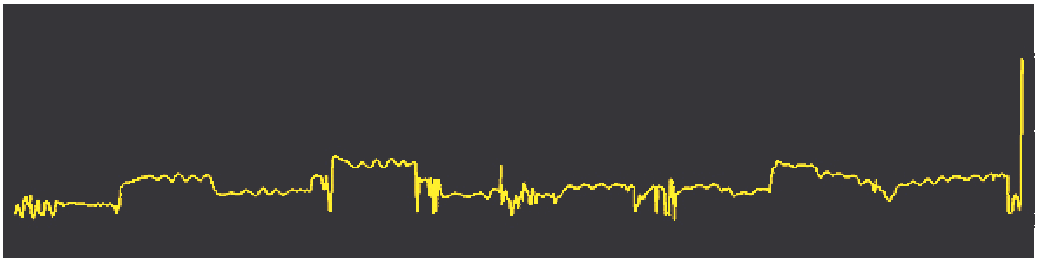
Avec l'ordinateur, analyser le son consiste à calculer les indicateurs précédents : il s'agit d'en dresser une sorte de carte d'identité quantitative, qui vient compléter l'analyse faite par l'oreille, plus qualitative. L'analyse par l'ordinateur permet d'extraire les courbes d'intensité et de hauteur ainsi qu'une représentation du timbre de la voix, et d'obtenir les paramètres de l'interprétation vocale.

Examinons ces trois courbes obtenues à partir du même extrait du lied de Richard Strauss opus 15 n°3 « Lob des Leidens » :

– la courbe de **l'intensité en fonction du temps** : elle permet de repérer les endroits où la voix est forte (maxima) et ceux où elle est faible (minima)

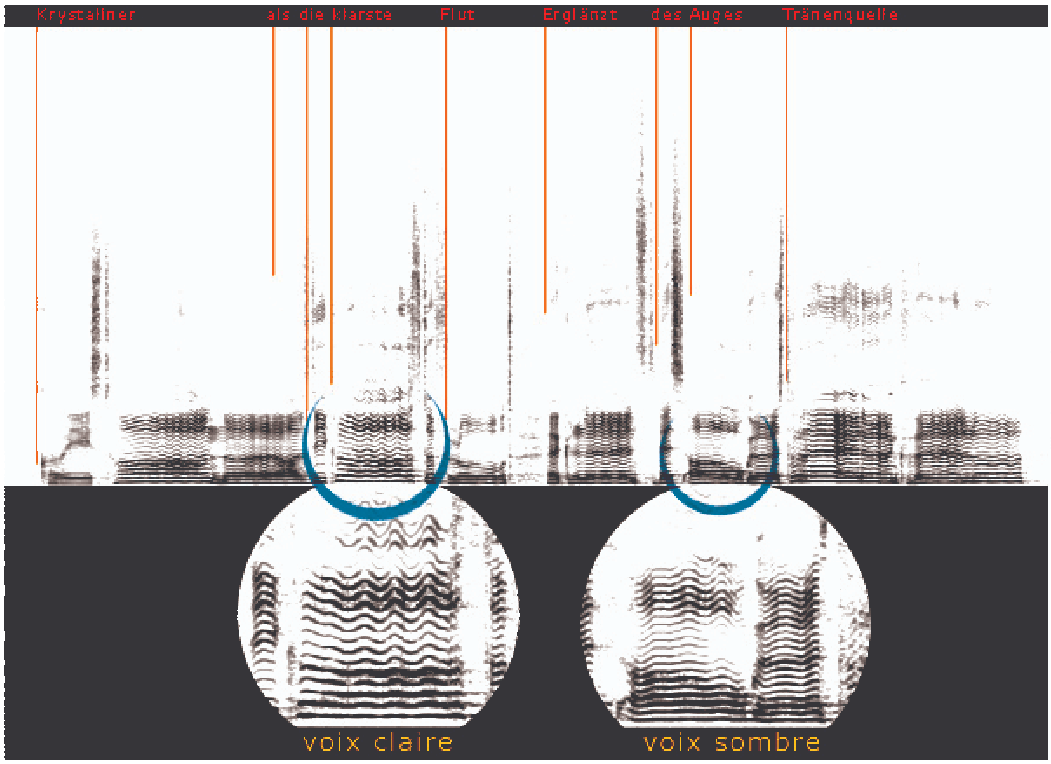


– la courbe de **hauteur en fonction du temps** : elle permet de suivre la hauteur de la note chantée ;



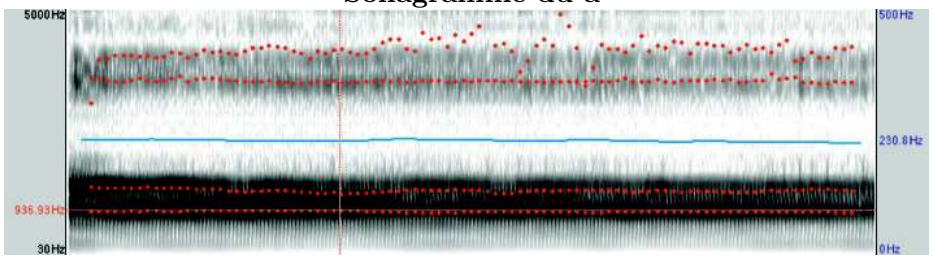
– La représentation du timbre de la voix est obtenue par un sonagramme. Pour une même hauteur chantée, la répartition d'énergie sur ces fréquences peut varier. Elle est visualisée sur le sonagramme en niveaux de gris : une fréquence associée à une grande énergie apparaît en gris foncé, alors qu'une fréquence moins soutenue apparaît en gris clair. Une voix sombre placera plus d'énergie sur les fréquences graves du son, alors qu'une voix claire placera plus d'énergie sur ses fréquences aiguës.

Courbe du timbre du début du Lied opus 15 n°3 de Richard Strauss « Lob des Leidens »

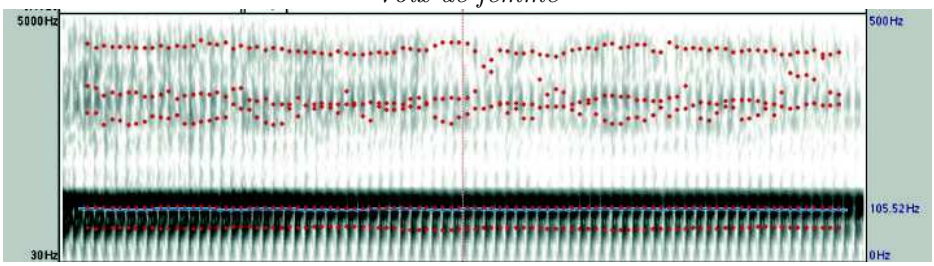


Autre exemple : « photographie » et caractéristiques d'une voyelle

Sonagramme du a



Voix de femme



Voix d'homme

5 Applications : manipulations du signal sonore

5.1 Filtrage

Les **filtres** sont des appareillages qui permettent d'isoler des bandes de fréquences. Ils sont principalement définis par :

- leur **fréquence de coupure** f_c, c'est-à-dire la limite de fréquence à partir de laquelle ils agissent.
- leur **pente**, c'est-à-dire la rapidité avec laquelle ils éliminent les fréquences à partir de f_c.

Mathématiquement, appliquer un filtre correspond à faire une convolution avec le signal. On distingue quatre types de filtres :

- **passe-bas**, qui laissent passer toutes les fréquences inférieures à f_1 ;
- **passe-haut**, qui laissent passer toutes les fréquences supérieures à f_2 ;
- **passe-bande**, qui laissent passer toutes les fréquences comprises entre f_1 et f_2 ;
- **de réjection**, qui laissent passer toutes les fréquences, sauf celles comprises entre f_1 et f_2.

5.2 Synthèse

Pour synthétiser la voix humaine avec un ordinateur, il faut :

- connaître son fonctionnement et les organes mis en jeu (poumons, cordes vocales, résonateurs) ;

– analyser des enregistrements de la voix, afin de trouver des équations mathématiques nécessaires pour effectuer la synthèse.

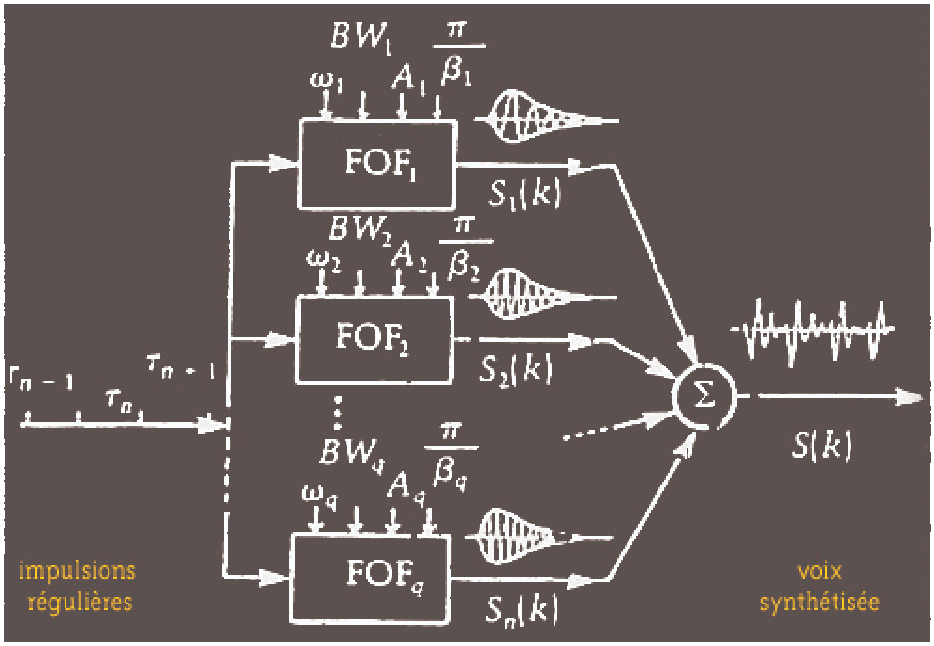
C'est ainsi qu'a été conçu à l'Ircam le programme *Chant* pour la synthèse de la voix. Il est articulé en deux étages successifs :

– une source qui émet un son pseudo périodique qui donne la hauteur de la voyelle ;
– un filtre qui donne une résonance particulière selon la nature de la voyelle. Cette source correspond, dans le corps humain, à l'ensemble constitué des poumons et des cordes vocales, le filtre est l'équivalent des résonateurs.

Le fonctionnement des résonateurs de l'appareil vocal humain est simulé en utilisant :

– soit des filtres agissant sur le spectre en fréquences
– soit des Fonctions d'Onde Formantiques (FOF) opérant dans le domaine temporel.

Schéma de principe d'un synthétiseur à base de FOF



La position des résonateurs chez l'homme crée dans le son généré des formants. Dans le programme *Chant*, on modélise cinq formants, en utilisant cinq filtres en parallèle, chacun émettant des fonctions d'ondes correspondant à un formant donné, la somme des cinq donnant le signal souhaité.

Le schéma ci-dessus montre ce principe : sur l'entrée à gauche sont envoyées des impulsions régulières simulant les coups de glotte, chaque « boîte » rectangulaire FOF fabrique des fonctions d'ondes correspondant à un formant donné, et le résultat final apparaît à la sortie à droite.

5.3 Transmission et compression

Présentons sommairement le principe de compression du son. Les techniques sont, dans ce domaine, très sophistiquées. Remarquons simplement que tout signal sonore est égal à sa série de Fourier sur un intervalle de temps suffisamment petit pour qu'il soit considéré comme périodique. Il suffit alors de tronquer la série de Fourier pour ne garder que les coefficients correspondants aux fréquences « utiles » (par exemple audibles) (en nombre fini). On calcule ces coefficients par FFT et échantillonnage et il suffit alors de les stocker dans un fichier (dit fichier « son ») pour les transmettre (via le WEB) ou les graver (techniques des CD dits « numériques »). En gros, comme on connaît la base des fonctions (sinusoïdales) toute l'information est contenue dans les coefficients de Fourier.

La fréquence d'échantillonnage correspond au nombre d'échantillons du signal par seconde. Un signal échantillonné à 32 000Hz comprend donc 32 000 échantillons par seconde. Le célèbre théorème de Shannon indique que si on choisit suffisamment d'échantillons (en fait si la fréquence d'échantillonnage est le double de la plus grande fréquence du signal étudié) on reconstruit **exactement** ce signal à partir des échantillons (alors qu'il fallait a priori une série infinie).

Si on prend une fréquence d'échantillonnage à 32 KHz, elle vaut deux fois ce que l'oreille humaine peut percevoir au maximum (16 000 Hz) . Un signal échantillonné à une telle fréquence restitue donc très fidèlement le son entendu. Si on échantillonne à une fréquence inférieure, on perdra de l'information et le signal sera de moins bonne qualité.

Toutefois, il faut garder en tête que si on code un échantillon sur 8 octets, 1 seconde de son échantillonné à 32KHz « pèsera » 256 KOctets, une minute environ 15 MO. Un CD de 800MO pourra contenir environ 55 mn de musique.

De nouvelles techniques (MP3) permettent d'améliorer la compression en gardant une restitution fidèle du signal. On doit changer de base pour cela et utiliser d'autres fonctions que les fonctions trigonométriques. L'analyse de Fourier est très générale et s'adapte à ces nouvelles fonctions.

Références

[1] A. Landercy, R. Renard, *Éléments de phonétique*, Editions Didier, Bruxelles, 1977

[2] Site WEB de l'IRCAM : http :// www.ircam.fr

[3] H. Reinhard, *Éléments de mathématiques du signal*, Dunod, 2002

[4] C. Gasquet, P. Witomski, *Analyse de Fourier et applications*, Masson, 1995