NOTES de COURS
Audionumérique
M2P informatique
Université Paris-Sud 11
2009 - 2010

Christophe d'ALESSANDRO
LIMSI-CNRS BP 133, F-91403 Orsay

Table des matières

1. Introduction 4
2. Signal et spectre 7
   2.1. Introduction 7
   2.2. Energie et puissance 8
   2.3. Signaux périodiques et séries de Fourier 9
   2.4. intégrale de Fourier 11
3. Signaux et systèmes numériques 12
   3.1. échantillonnage 12
   3.2. Quantification 13
   3.3. Transformée de Fourier du signal échantillonné 13
   3.4. Transformée de Fourier discrète 14
   3.5. Systèmes discrets 15
   3.6. Transformée en z 17
4. Filtres linéaires 19
   4.1. Filtres dynamiques 19
   4.2. Types de filtres 20
   4.3. Exemples de filtres 22
5. Compléments sur la numérisation 24
   5.1. La chaîne du traitement numérique 24
   5.2. échantillonnage 25
   5.3. Reconstruction du signal analogique 28
   5.4. Quantification 30
   5.5. Tremblotement (dither) 32
   5.6. Suréchantillonnage 33
   5.7. Convertisseurs $\Delta\Sigma$ 34
6. Transformée de Fourier à court terme 37
   6.1. Spectrographe 37
   6.2. Vocodeur de phase 39
   6.3. Interprétation en banc de filtre de la TFCT 40
   6.4. Interprétation par blocs de la TFCT 42
   6.5. Modification et reconstruction 44
   6.6. Applications de la TFCT 45
   6.7. Représentation sinusoïdale 46

7. Les sons du Français 49
7.1. Production de la parole 49
7.2. Phonèmes 52
7.3. Traits distinctifs 56
7.4. Prosodie 57
8. Lecture de spectrogrammes 59
8.1. Indices acoustiques 59
8.2. Voyelles orales 60
8.3. Voyelles nasales 60
8.4. Fricatives sourdes 61
8.5. Fricatives voisées 61
8.6. Plosives sourdes 61
8.7. Plosives voisées 62
8.8. Liquides 62
8.9. Nasales 62
8.10. Semi-voyelles 62
9. Les sons instrumentaux 63
9.1. Les familles d'instruments 63
9.2. Propriétés temporelles des sons instrumentaux 65
9.3. Sons périodiques et apériodiques 66
9.4. Enveloppe spectrale : le modèle source/filtre 67

1. Introduction

L'introduction de l'électricité dans le domaine de la musique date de plus d'un siècle. La première application réelle à un instrument de musique est l'utilisation de transmission électrique entre les claviers et les soupapes, au grand orgue de Saint-Augustin, à Paris. Le brevet de Peschard et Barker date des années 1860. Le titulaire de cet instrument était Eugène Gigout, qui a été professeur d'orgue au Conservatoire de Paris et qui à laissé une oeuvre d'orgue importante. Il s'agit ici de l'application d'une technique nouvelle au mécanisme d'un instrument traditionnel.

Depuis cette époque, les champs d'application du génie électrique à la musique s'est considérablemet diversifié :

- création instruments électriques et électroniques, synthèse. De nombreux instruments électriques (guitare électrique, piano électrique etc.) ont transformé le son musical. Ensuite sont apparus les synthétiseurs électroniques (analogiques), en général pilotés par clavier, ou bien modulaires et voués aux sons répétitifs. La génération suivante est celle des synthétiseurs numériques, et de la normalisation MIDI.
- acoustique musicale. L'acoustique musicale est aussi vieille que l'acoustique et la théorie musicale. Elle s'occupe du fonctionnement acoustique des instruments de musique, des théories musicales (gammes, tempéraments, intervalles musicaux, rythmes), de la perception et de la cognition musicale, de l'acoustique des salles etc.
- enregistrement et restitution, codage du signal musical. Depuis le téléphone et le phonographe, jusqu'à la numérisation du son, l'archivage la transmission et la diffusion du son est un enjeux majeur pour la musique.
- analyse/synthèse et transformation des sons. De plus en plus d'applications de transformation de son naturels, et des techniques de plus en plus raffinées d'analyse/synthèse se développent grâce à l'ordinateur.
- informatique musicale. Ce terme regroupe toutes les applications de l'informatique, en particulier de l'informatique personnelle à la musique. Citons l'édition et le jeux de partition, l'enregistrement et l'échantillonnage de séquences, le studio numérique pour le montage, les effets, la production de CD etc.
- diffusion du son, spatialisation. Depuis la diffusion stéréophonique, puis la diffusion avec des orchestres de haut-parleurs, l'outil informatique permet maintenant de reconstituer l'espace sonore d'une salle virtuelle, ou bien de positionner et de mettre en mouvement dans l'espace des sources sonores virtuelles ("spacialisation").

- musique électroacoustique, éléctronique, de studio. Il s'agit ici, depuis la "musique concrête" des années 50, de la musique créée en studio, sans instrumentistes, par des sons naturels ou synthétique.

- musique électronique de scène. Depuis les grands groupes de musique pop planante des année 60-70, jusqu'aux groupes actuels de musique techno. L'instrument élecronique ou numérique est intégré dans un jeux de scène en direct ("live"). Il est remarquable que maintenant même les formation purement instrumentales soient sonorisée, et donc médiatisée par l'électronique.

- composition algorithmique, ou assistée par ordinateur. Le processus même de composition musicale peut être assisté par l'ordinateur. L'exemple le plus simple est le calcul d'accompagnements automatiques (accords, rythmes) pour une ligne mélodique jouée. Les pianos numériques avec enregistrement et restitution ("replay") permettent de capter très finement le jeux instrumental et d'appliquer à volonté des algorithmes intéractifs. Enfin, une partition classique peut être calculée par un algorithme, qui implémente une formule mathématique, ou une autre forme de processus.

De toutes ces applications musicales directes, les changements les plus importants et les plus profonds apportés par l'électricité sont la diffusion et l'enregistrement. Le son autrefois ne pouvait être entendu que sous la condition d'une présence physique réelle du musicien et de l'instrument. Le son, le jeux d'un instrumentiste, le concert n'existaient que dans l'instant, et il ne pouvaient jamais être produit deux fois de la même façon, ou pour un public différent. Aujourd'hui, il peut être produit n'importe où, n'importe quand et n'importe comment, et néanmoins être diffusé, écouté, reécouté, reproduit, analysé, transformé et synthétisé, par tous ceux qui ont ou auront un enregistrement. Par l'électricité, le son a perdu son statut immatériel pour prendre celui d'objet. Ainsi, a part peut être pour les musiciens instrumentistes "acoustiques" , qui passent de longues heures en compagnie de leurs instruments, la grande majorité de la musique diffusée actuellement passe par des moyens électroniques, et de plus en plus, par des moyens numériques.

Le propos de cet exposé est d'aborder quelques techniques de traitement numérique du signal musical, ou traitement du signal audionumérique. Les ouvrages suivants m'ont aidé à préparer ces notes de cours :

1. Arthur Benade "Fundamental of Musical Acoustics", Dover publications, New York, 1976, 1990.

2. Pierre Schaeffer, "Traité des objets musicaux", Edition du Seuil, Paris, 1966, 1977.

3. Emile Leipp, "Acoustique et musique", Masson, Paris, 1977.

4. Jürgen Meyer, “Acoustics and the performance of music”, Verlag das Musikinstrument, Frankfurt/Main, 1978.

5. John Pierce, “Le son musical”, Pour la science, Belin, 1983.

6. Ronald Crochiere, Lawrence Rabiner, “Multirate digital signal processing”, Prentice Hall, Englewood Cliffs, 1983.

7. Alan Oppenheim, Ronald Schaffer, “Digital signal processing”, Prentice Hall, Englewood Cliffs, 1991.

8. Lawrence Rabiner, Bernard Gold, “Theory and application of Digital signal processing”, Prentice Hall, Englewood Cliffs, 1988.

9. William Strong, G.R. Plilnik, “Music, speech, audio”, Soundprint, Provo, UT, 1992.

10. Denis Mercier (éditeur), “Le livre des techniques du son”, Eyrolle, 1987-1993 (3 tomes).

11. Sophocles J. Orfanidis, "Introduction to signal processing", Prentice Hall International Editions, 1996.

12. Curtis Road, "L’Audionumérique", Dunod, Paris, 1998 (traduction de "Computer Music Tutorial", MIT Press.)

Par ailleurs, les informations pertinentes sur les recherches dans ce domaine se trouvent dans des périodiques comme :

1. Computer Music Journal (MIT Press)

2. Journal of the Audio Engineering Society (Audio Engineering Society)

3. IEEE Transaction on Speech and Audio Processing (Institute of Electrical and Electronics Engineers)

4. Journal of the Acoustical Society of America (Acoustical Society of America)

5. Acta Acustica (European Acoustical Association)

ou bien à l’occasion de conférences périodiques comme :

1. International Computer Music Conférence (ICMC)

2. International Conference on Acoustics, Speech and Signal Processing (ICASSP)
3. Convention of the AES
4. Meeting of the ASA
5. Forum Acusticum (EAA)
6. Journées d'Informatique Musicale (JIM)
7. International Conference on Acoustics (ICA)

2. Signal et spectre

2.1. *Introduction*

Il est indispensable de rappeler quelques éléments de traitement du signal utiles pour aborder l'audionumérique.

Le son, ou signal acoustique est une variation de pression atmosphérique qui rayonne depuis la source sonore.

Le signal parvenant au récepteur (oreilles, microphones, ...) se trouve considérablement déformé en fonction de la localisation relative de la source et du capteur, mais aussi des conditions d'environnement et d'ambiance sonore. En particulier, les réflexions multiples que provoque l'onde sonore sur les parois ou les objets environnants, et le déphasage simplement dû à la distance, déforment le signal initial de façon considérable dans sa dimension amplitude/temps. L'analyse temporelle des signaux, qui offre néanmoins de précieuses méthodes, ne saurait ainsi suffire pour tenter d'approcher les performances remarquables réalisées par le système auditif humain.

Parmi les signaux possédant des propriétés remarquables citons :

- les signaux causaux :

$$x(t) = O \; si \; t < 0 \qquad [1]$$

  Le son, dans le monde réel, est toujours causale, mais dans le monde virtuel, sa représentation sur ordinateur peut ne pas l'être, si l'on ne travaille pas en temps réel.

- les signaux périodiques, qui possède la propriété :

$$x(t + T_0) = x(t) \qquad [2]$$

dont l'archétype est l'exponentielle complexe $e^{2i\pi t}$ et ses parties réelles et imaginaires, les fonctions sinus et cosinus. Les voyelles sont des signaux (quasi-) périodiques, dont la période est l'inverse de la fréquence de voisement.

- l'impulsion idéale, ou distribution de Dirac $\delta(t)$ qui n'est pas une fonction mais une mesure. Elle est définie par :

$$< \delta, x >= \int_{-\infty}^{+\infty} x(t)\delta(t)dt = x(0) \qquad [3]$$

et que l'on peut voir comme une pseudo-fonction valant 0 sur R^* et $+\infty$ en 0 (ce n'est pas une fonction réelle, puisque $+\infty$ n'est pas une valeur réelle). Les explosions de plosives sont des impulsions acoustiques, que l'on peut considérer idéalement comme des impulsions de Dirac.

- les bruits. Ce sont des signaux dont les amplitudes sont imprévisibles, et donc ne peuvent être décrites par une fonction mathématique $x(t)$ connue. On peut les considérer comme la réalisation (une épreuve particulière) d'un signal aléatoire. En parole, les bruits se rencontrent dans les fricatives, par exemple.

2.2. *Energie et puissance*

Le signal tel qu'on l'obtient grâce à des capteurs physiques se représente dans la dimension amplitude/temps. Un signal est donc une fonction $x(t)$ réelle et continue (il est parfois utile de considérer des signaux complexes) d'une seule variable réelle, le temps.

Pour un traitement numérique, il faudra représenter ce signal par une suite de nombres, par échantillonnage et quantification, ce qui sera abordé plus loin.

Une première notion relative au signal est son énergie totale $E(x)$. L'énergie est une puissance par unité de temps, l'énergie totale du signal est donc l'intégrale de la puissance instantanée $e(t)$:

$$e(t) = |x(t)|^2 \qquad [4]$$

Le son est un signal d'énergie finie (ou de carré sommable, $x(t) \in L^2(R)$), donc :

$$E(x) = \int_{-\infty}^{+\infty} |x(t)|^2 dt < +\infty \qquad [5]$$

son	dB	rapport
seuil d'audition 1000 Hz	0	1
chuchotement	20	10
conversation	80	10 000
hurlements	100	100 000
groupe de Hard Rock	120	1000 000
seuil de la douleur	140	10 000 000

Tableau 1. *Exemple de sons et leurs intensités.*

En acoustique, la grandeur mesurée, qui est représentée par $x(t)$ est généralement la pression. Il est commode de mesurer les amplitudes de $x(t)$ sur une échelle logarithmique plutŞt que linéaire, à cause du très grand domaine de variation du signal audible. Cette unité logarithmique rend bien compte d'autre part des propriétés de la perception (sensation se rapportant au logarithme de l'excitation). L'unité relative de mesure est le décibel (dB), qui est défini comme dix fois le logarithme (de base 10) du rapport entre la puissance e (ou l'énergie) et une puissance P_0 (ou une énergie) de référence :

$$e_{dB} = 10 \log_{10}(\frac{e}{e_0}) \quad [6]$$

le niveau de référence e_0 doit être précisé dans chaque cas. Un microphone mesure des amplitude de variation de pression, et non des energies. Pour une amplitude (ou une intensite) la formule est :

$$i_{dB} = 20 \log_{10}(\frac{i}{i_0}) \quad [7]$$

Pour une pression on prendra comme niveau de référence la pression atmosphérique correspondant au seuil d'audition a 1000 Hz (ce sont les dB SPL : Sound Pressure Level).

La table 1 donne quelques exemples de DB SPL pour des signaux acoutiques.

2.3. *Signaux périodiques et séries de Fourier*

Les travaux de Fourier ont posé les bases de *l'analyse spectrale*, ou représentation du signal dans le domaine fréquentiel, conjugué du domaine temporel (le produit d'un temps et d'une fréquence est sans dimension). Notons qu'un traitement par des méthodes de type fréquentiel ne saurait, pas plus qu'un traitement purement temporel, apporter une réponse satisfaisante au problème de

son	frequence (Hz)
limite des sons audibles	15-20
note la plus basse du piano	27.5
voix d'homme (parole)	90-150
voix de femme (chant)	175-1300
note la plus haute du piano	4180
limite des sons audibles	16000-20000

Tableau 2. *Exemple de sons et leurs fréquences.*

la représentation du signal sonore. En effet le contenu fréquentiel de la parole évolue au cours du temps : le son est un signal non-stationnaire.

Pour un signal périodique, la notion de fréquence fondamentale F_0 apparait naturellement : c'est le taux de répétition par unité de temps de la fonction (ou inverse de sa période, dite période fondamentale T_0). Une fréquence se mesure en nombre de cycles par seconde ou Hertz (Hz).

Fourier a montré que toute fonction périodique peut se décomposer en série trigonométrique, dite *série de Fourier*. Ainsi des fonctions particulières, les exponentielles complexes, forment une base de l'espace des fonctions périodiques.

La notion de spectre fréquentiel s'introduit, comme représentation du signal en somme d'exponentielles complexe (ou ce qui revient au même, de sinus et de cosinus). Chaque composante est située à une fréquence multiple de la fréquence fondamentale, et se nomme *harmonique*. La fréquence de l'harmonique 1 est la fréquence fondamentale, celle de l'harmonique 2 est le double de la fréquence fondamentale (ou octave), celle de l'harmonique 3 est le triple de la fondamentale etc. La Table 2.3 donne la fréquence de quelques sons. Lorsque l'on considère des rapports de fréquence, on utilise couramment l'octave (rapport 2 entre deux fréquences), le demi-ton tempéré (division de l'octave en 12 parties égales) ou le Cent (centième de demi-ton tempéré) :

$$Cent = \sqrt[1200]{2} \simeq 1.00057779 \qquad [8]$$

La décomposition d'un signal périodique réel en somme de sinusoïdes s'écrit dans la dimension temporelle :

$$x(t) = \sum_{k=-\infty}^{\infty} c_k e^{2i\pi k f_0 t} = c_0 + \sum_{k=1}^{\infty} 2|c_k| \cos(2\pi k f_0 t + \varphi(c_k)) \qquad [9]$$

$$c_k = |c_k| e^{i\varphi(c_k)} \in C \qquad [10]$$

les coefficients de Fourier sont obtenus par :

$$\tilde{x}(k) = c_k = \frac{1}{T_0} \int_0^{T_0} x(t) e^{-2i\pi kt/T_0} dt \qquad [11]$$

On appelle spectre du signal périodique x l'ensemble de ses coefficients de Fourier c_k. On constate donc que le spectre d'un signal périodique est discret (spectre de raies), on l'appelle un spectre harmonique. La décomposition du signal par série de Fourier est donc appelée analyse harmonique. Les c_k étant complexes, ils possèdent une amplitude et une phase (qui sont l'amplitude et la phase des coefficients c_k associés aux sinusoïdes). On parlera donc de spectre complexe (c_k), de spectre d'amplitude $(|c_k|)$, et de spectre de phase $(\varphi(c_k))$.

2.4. *intégrale de Fourier*

La décomposition d'une fonction complexe de la variable réelle en somme de fonctions exponentielles complexes, naturelle pour les fonctions périodiques, peut être étendue pour certains types de signaux. En particulier, pour les signaux d'énergie finie, on définit la transformée de Fourier (ou intégrale de Fourier) par :

$$\tilde{x}(\nu) = \int_{-\infty}^{+\infty} x(t) e^{-2i\pi\nu t} dt \qquad [12]$$

et de la transformée de Fourier inverse :

$$x(t) = \int_{-\infty}^{+\infty} \tilde{x}(\nu) e^{2i\pi\nu t} d\nu \qquad [13]$$

La convergence de l'intégrale est assurée si le signal est d'énergie finie. La variable ν est une fréquence. On peut considérer intuitivement l'intégrale de Fourier comme un passage à la limite du cas des fonctions périodiques : en effet, si une fonction n'est pas périodique, mais d'énergie finie par exemple, on peut la considérer comme la limite d'une suite de fonctions périodiques dont la période devient infinie. Ainsi les raies du spectre des fonctions périodiques se rapprochent jusqu'à former un spectre continu, à la limite, et non plus un spectre harmonique, comme dans le cas des signaux périodiques.

La largeur du support du spectre d'un signal se nomme sa bande passante. On remarque que dans le cas de l'intégrale de Fourier le signal se développe en une somme infinie de fonctions exponentielles complexes, de durée infinie. Il

faut donc une durée infinie pour calculer la valeur du spectre à une fréquence donnée, et un spectre de bande passante infinie pour calculer la fonction à un instant donné.

La transformée de Fourier possède un certain nombre de propriétés remarquables. Si l'on définit le produit de convolution :

$$\tilde{z}(\nu) = \int_{-\infty}^{+\infty} \tilde{x}(\tau)\tilde{y}(\nu - \tau)d\tau \qquad [14]$$

alors la transformée de Fourier transforme le produit simple en produit de convolution :

$$z(t) = x(t) \times y(t) \leftrightarrow \tilde{z}(\nu) = \tilde{x}(\nu) * \tilde{y}(\nu) \qquad [15]$$

La transformée de Fourier est linéaire :

$$\sum_{n=1}^{N} a_n x_n(t) \leftrightarrow \sum_{n=1}^{N} a_n \tilde{x_n}(\nu) \qquad [16]$$

3. Signaux et systèmes numériques

3.1. *échantillonnage*

L'ordinateur ne peut travailler que sur des séquences de nombres. Le traitement digital résulte d'une double discrétisation du signal analogique.

Un signal discret s'obtient par échantillonnage du signal analogique à des instants t_n. Les instants d'échantillonnage sont en général choisis régulièrement espacés :

$t_n = nT_e$ ou T_e représente le pas d'échantillonnage et $1/T_e = F_e$ la fréquence d'échantillonnage. Le signal digital, à temps discret, est une suite d'échantillons, dépendants de l'indice n, et obtenus par échantillonnage du signal à temps continu :

$$x(n) = x(nT_e) \qquad [17]$$

Par la suite l'indice n sera dénommé temps discret, par abus de langage (le temps discret est nT_e), en considérant une fréquence d'échantillinnage de 1, pour simplifier les expressions mathématiques.

Le théorème de Shannon, ou théorème d'échantillonnage se formule de la façon suivante : Si un signal $x(t)$ possède une transformée de Fourier dont le support est borné, alors il peut être reconstruit exactement et de manière unique à partir de ses échantillons si la fréquence d'échantillonnage est égale au double de la plus haute fréquence présente dans le signal.

On doit donc choisir la fréquence d'échantillonnage F_e telle que : $F_e \geq \Omega$ si $\tilde{x}(\nu) = 0$ pour $|\nu| > \Omega$.

Alors on peut reconstruire le signal à temps continu à partir de ses échantillons, par la formule :

$$x(t) = \sum_{-\infty}^{\infty} x(nT_e) \frac{\sin[\pi F_e(t - nT_e)]}{\pi F_e(t - nT_e)} \qquad [18]$$

3.2. *Quantification*

Les amplitudes d'un signal digital sont également discrétisées. Alors que le signa analogique est à valeurs réelles, le signal digital est à valeurs discrètes.En général, un codage binaire des amplitudes est utilisé, et le nombres de pas de quantification est noté en bits. Un mot de N bits donne un nombre de niveaux égal à 2^N.

Alors la dynamique que l'on peut représenter est d'environ $6.02N[dB]$. Pour une quantification sur 8, 12, 16 et 20 bits, le nombre de niveaux est 256, 4096, 65636 et 1048576, qui correspondent à 48, 72, 96 et 120 dB de dynamique.

3.3. *Transformée de Fourier du signal échantillonné*

Le signal échantillonné peut s'écrire :

$$x_e(t) = x(t) \times \sum_{n=-\infty}^{+\infty} \delta(t - nT_e) \qquad [19]$$

En notant $x(n) = x_e(nT_e)$, la transformée de Fourier du signal échantillonné est obtenue par intégrale de Fourier (pour simplifier on prend $T_e = F_e = 1$) :

$$\tilde{x}(\nu) = \sum_{n=-\infty}^{+\infty} x(n) e^{-2i\pi\nu n} \qquad [20]$$

et la transformée de Fourier inverse :

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \tilde{x}(\nu) e^{2i\pi\nu n} d\nu \qquad [21]$$

On remarque que la transformée de Fourier du signal échantillonné possède la forme d'une série de Fourier, et que le signal échantillonné lui même possède la forme des coefficients de Fourier. L'effet de l'échantillonnage est donc de périodiser (rendre périodique) le spectre du signal continu correspondant. La période du spectre périodisé est égale à la fréquence de l'échantillonnage F_e. Dans l'intervalle spectral fondamental $[-F_e/2, F_e/2[$ le spectre du signal échantillonné s'obtient par repliement du spectre du signal continu. Si le signal a été échantillonné sous les conditions du théorème de Shannon, il n'y a pas de repliement spectral (car le spectre est nul en dehors de la période spectrale principale).

3.4. *Transformée de Fourier discrète*

Un nouveau type de transformation de Fourier a été défini pour les signaux numériques. Les formules précédentes montrent qu'un signal discret possède un spectre périodique, et qu'un signal périodique possède un spectre discret. Si l'on considère un signal discret et périodique, de période N, et de fréquence d'échantillonnage 1, son spectre va également être discret et périodique, de période 1 et de fréquence d'échantillonnage $1/N$.

Soit une suite périodique, dont la période principale est constituée de N nombres $x(n)$, $0 \leq n \leq N-1$. On définit la transformée de Fourier discrète (TFD) par :

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-2i\pi kn/N} \qquad [22]$$

pour $0 \leq k \leq N-1$ et la TFD inverse (TFDI) :

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{2i\pi kn/N} \qquad [23]$$

On obtient ainsi une autre suite de N nombres, qui représentent exactement la période principale de la suite périodique temporelle initiale.

On peut considérer N échantillons d'une suite quelconque (pas forcément périodique) et calculer la TFD de la même façon : les N échantillons spectraux obtenus représenteront de façon exacte (inversible) les N échantillons temporels initiaux. Cette utilisation de la TFD est très utile en traitement du signal, mais il faut garder en mémoire que, la TFD d'un signal étant un signal échantillonné périodique, sa transformée inverse redonne un signal échantillonné périodique, égal au signal échantillonné initial dans l'intervalle temporel fondamental [0,N[.

Des algorithmes de calcul rapides, la transformation de Fourier rapide (FFT), existent si l'on restreint les valeurs de N à des puissances de 2.

3.5. *Systèmes discrets*

Voici quelques exemples de signaux discrets (ou suite numérique). Le signal discret correspondant à la distribution de Dirac est la suite numérique :

$$\delta(n) = \begin{cases} 1 & \text{si } n = 0 \\ 0 & \text{sinon} \end{cases} \qquad [24]$$

pour toute valeur entière k, on a :

$$\delta(n-k) = \begin{cases} 1 & \text{si } n = k \\ 0 & \text{sinon} \end{cases} \qquad [25]$$

un autre exemple important de suite est l'échelon unité :

$$u(n) = \begin{cases} 1 & \text{si } n \geq 0 \\ 0 & \text{si } n < 0 \end{cases} \qquad [26]$$

Une suite exponentielle (complexe) est de la forme :

$$x(n) = a^n \qquad [27]$$

avec $a = re^{i\varphi}$.

Le traitement du signal numérique implique la transformation de suites numériques en d'autres suites numériques. On appelle *système discret* une règle de transformation d'une suite numérique en une autre suite numérique :

$$y(n) = T[x(n)] \qquad [28]$$

la suite $x(n)$ est appelée "entrée du système", et la suite $y(n)$ est la "sortie", ou "réponse" du système.

Des exemples particulièrement importants de systèmes sont l'élément de retard :

$$y(n) = x(n-1) \quad [29]$$

et l'élément multiplicateur (avec le gain a complexe) :

$$y(n) = ax(n) \quad [30]$$

Un système est *linéaire* si :

$$T[a_1x_1(n) + a_2x_2(n)] = a_1T[x_1(n)] + a_2T[x_2(n)] \quad [31]$$

un système est invariant dans le temps si un décalage de l'entrée produit le même décalage pour la réponse :

$$T[x(n-k)] = y(n-k) \quad [32]$$

pour tout k.

La réponse impulsionnelle du système vaut :

$$h(n) = T[\delta(n)] \quad [33]$$

si cette réponse est nulle pour les temps négatifs, $h(n) = 0$ pour $n < 0$ le système est *causal*.

Si la réponse impulsionnelle est absolument sommable, le système est stable :

$$\sum_{n=-\infty}^{+\infty} h(n) < +\infty \quad [34]$$

et toute entrée bornée engendrera une sortie bornée.

La convolution discrète g de deux suites x et y est définie par :

$$g(n) = \sum_{n=-\infty}^{+\infty} x(k)y(n-k) = x(n) * y(n) = \sum_{n=-\infty}^{+\infty} y(k)x(n-k) = y(n) * x(n) \quad [35]$$

La suite de Dirac est l'élément neutre de la convolution discrète :

$$x(n) * \delta(n) = \sum_{n=-\infty}^{+\infty} x(k)\delta(n-k) = \sum_{n=-\infty}^{+\infty} \delta(k)x(n-k) = x(n) \qquad [36]$$

3.6. *Transformée en z*

Pour les signaux et systèmes discrets, on utilise une forme particulière de représentation dans le domaine fréquentiel : la transformée en z. La transformée en z (notée $X(z)$), de la variable complexe z, se définit pour une suite $x(n)$ par :

1. l'expression :

$$X(z) = \sum_{n=-\infty}^{+\infty} x(n)z^{-n} \qquad [37]$$

2. le domaine de convergence de la série précédente, c'est à dire dans le plan complexe l'anneau :

$$R_1 < |z| < R_2 \qquad [38]$$

la transformée en z possède une transformée inverse :

$$x(n) = \frac{1}{2i\pi} \oint_C X(z)z^{n-1}dz \qquad [39]$$

Différentes méthodes existent pour calculer cette transformée inverse.

Il est connu qu'une série absolument sommable est sommable, donc une condition suffisante pour la convergence est :

$$\sum_{n=-\infty}^{+\infty} |x(n)||z^{-n}| < +\infty \qquad [40]$$

Pour une suite causale, la transformée en z se réduit à :

$$X(z) = \sum_{n=0}^{+\infty} x(n)z^{-n} \qquad [41]$$

donc avec uniquement des puissances négatives de z, et le domaine de convergence est de la forme :

$$|z| > R_1 \quad [42]$$

On remarque que la transformée de Fourier des signaux discrets est de la forme d'une transformée en z, évaluée pour :

$$z = e^{2i\pi\nu} \quad [43]$$

c'est à dire évaluée sur le cercle unité du plan complexe, l'ensemble des z tels que $|z| = 1$. Si le cercle unité est dans le domaine de convergence de la transformée en z de la suite, alors elle possède une transformée de Fourier. La transformée en z est donc plus générale que la transformée de Fourier. Rappelons qu'une condition suffisante d'existence de la transformée de Fourier est d'être d'énergie finie.

Voici quelques propriétés de la transformée en z :

1. linéarité :

$$\sum_{n=1}^{N} a_n x_n(t) \leftrightarrow \sum_{n=1}^{N} a_n X_n(z) \quad [44]$$

2. décalage :

$$x(n + n_0) \leftrightarrow z^{n_0} X(z) \quad [45]$$

3. transformation de la convolution en produit :

$$x(n) * y(n) \leftrightarrow X(z)Y(z) \quad [46]$$

4. renversement du temps :

$$x(-n) \leftrightarrow X(z^{-1}) \quad [47]$$

4. Filtres linéaires

4.1. *Filtres dynamiques*

Une classe particulièrement utile de systèmes discrets est celle des filtres digitaux linéaires. Un filtre linéaire est un système linéaire invariant dans le temps.

La relation entre entrée x et sortie y d'un filtre linéaire s'exprime par un produit de convolution dans le domaine temporel entre la suite d'entrée et la réponse impulsionnelle du filtre h :

$$y(n) = \sum_{k=-\infty}^{+\infty} x(k)h(n-k) = x(n) * h(n) = \sum_{k=-\infty}^{+\infty} h(k)x(n-k) = h(n) * x(n) \quad [48]$$

dans le domaine fréquentiel, cette relation devient un simple produit :

$$Y(z) = H(z)X(z) \quad [49]$$

ou la transformée en z de la réponse impulsionnelle $H(z)$ est appelée fonction de transfert du système. Cette fonction de transfert évaluée sur le cercle unité (c'est à dire la transformée de Fourier de la réponse impulsionnelle) $H(e^{2i\pi\nu})$ est appelée gain complexe du filtre, ou réponse en fréquence.

La condition nécessaire et suffisante pour qu'un filtre linéaire soit stable est que sa réponse impulsionnelle soit sommable :

$$\sum_{n=-\infty}^{+\infty} |h(n)| < +\infty \quad [50]$$

Pour le système constitué d'un élément de retard, la fonction de transfert vaut :

$$H(z) = z^{-1} \quad [51]$$

et pour le système constitué d'un élément multiplicateur :

$$H(z) = a \quad [52]$$

Les filtres linéaires les plus utilisés en pratique sont les filtres dynamiques. Un filtre dynamique est un filtre linéaire, causal, stable, dont la fonction de transfert est une fraction rationnelle en z (à coefficients (a_k) et (b_l) réels :

$$H(z) = \frac{\sum_{l=0}^{M} b_l z^{-l}}{\sum_{k=0}^{N} a_k z^{-k}} \qquad [53]$$

une telle fonction de transfert peut s'exprimer en fonction de ses pôles d_k et de ses zéros c_l :

$$H(z) = \frac{A \prod_{l=1}^{M} (1 - c_l z^{-l})}{\prod_{k=1}^{N} (1 - d_k z^{-k})} \qquad [54]$$

La région de convergence de la fonction de transfert est de la forme $|z| > R$, puisque le système est causal. Comme il est stable, le cercle unité est contenu dans la région de convergence, donc $R < 1$. Tous les pôles d'un filtre dynamique sont donc situés à l'intérieur du cercle unité.

Dans le domaine temporel, un filtre dynamique est défini par une équation récurrente à coefficients constants :

$$\sum_{k=0}^{N} a_k y(n-k) = \sum_{l=0}^{M} b_l x(n-l) \qquad [55]$$

4.2. *Types de filtres*

On distingue parmi les formes de filtres possibles les filtres passe-bas, passe-haut, passe-bande, ou passe-tout, suivant leur gain. Un filtre passe bas coupe toutes les fréquences supérieures à sa fréquence de coupure. Un filtre passe haut coupe les fréquences inférieures à sa fréquence de coupure. Un filtre passe bande laisse passer un bande de fréquence, entre ses deux fréquences de coupure. Un filtre passe tout laisse passer toutes les fréquences, et n'agit donc que sur la phase.

Les filtres dynamiques peuvent se diviser en deux classes générales, en fonction de la forme de leur équation de récurrence. Si les termes a_k sont tous nuls pour $k > 0$, le filtre n'est pas récursif : c'est un filtre à réponse impulsionnelle finie, ou à horizon fini, d'ordre M :

$$y(n) = \sum_{l=0}^{M} b_l x(n-l) \quad [56]$$

un tel filtre ne possède pas de pôles, uniquement des zéros dans sa fonction de transfert :

$$H(z) = \sum_{l=0}^{M} b_l z^{-l} = \prod_{l=1}^{M} (1 - c_l z^{-l}) \quad [57]$$

la réponse impulsionnelle de ce filtre dure $M+1$ échantillons, et vaut (b_l), pour $0 \leq l \leq M$.

Si le filtre est uniquement récursif, les termes b_l étant nuls pour $l > 0$, le filtre est dit à réponse impulsionnelle infinie, d'ordre N :

$$y(n) = x(n) - \sum_{k=1}^{N} a_k y(n-k) \quad [58]$$

Sa fonction de transfert possède uniquement des pôles, et aucun zéro :

$$H(z) = \frac{1}{\sum_{k=0}^{N} a_k z^{-k}} = \frac{A}{\prod_{k=1}^{N} (1 - d_k z^{-k})} \quad [59]$$

Les filtres qui possèdent des pôles et des zéros sont également à réponse impulsionnelle infinie (récursifs).

Les filtres récursifs peuvent être réalisés sous plusieurs formes, en réécrivant leur fonction de transfert. Deux formes importantes sont la réalisation en série, et la réalisation en parallèle.

Pour la forme série, on remarque que les pôles et les zéros de la fonction de transfert sont tous complexes conjugués, puisque les coéfficients du filtre sont réels. On peut donc écrire la fonction de transfert sous la forme d'un produit de systèmes récursifs du second ordre :

$$H(z) = A \prod_{r=1}^{R} [\frac{(1 - b_{1r} z^{-1} + b_{2r} z^{-2})}{(1 + a_{1r} z^{-1} + a_{2r} z^{-2})}] \quad [60]$$

un produit de fonctions de transfert correspond à une mise en série des systèmes.

Pour la forme parallèle, on utilise la décomposition en éléments simples de la fraction rationnelle, pour écrire la fonction de transfert sous la forme d'une somme de systèmes du second ordre :

$$H(z) = \sum_{r=1}^{R} \frac{c_{0r} + c_{1r} z^{-1})}{(1 + a_{1r} z^{-1} + a_{2r} z^{-2})} \quad [61]$$

la sommation des éléments correspond à une mise en parallèle des systèmes.

4.3. *Exemples de filtres*

Un exemple de filtre non-récursif, à réponse impulsionnelle finie est le filtre différenciateur :

$$y(n) = x(n) - K\, x(n-1) \quad [62]$$

avec $1 > K > 0$. Sa réponse impulsionnelle vaut 1 en 0, et $-K$ en 1. sa fonction de transfert vaut :

$$H(z) = 1 - K z^{-1} \quad [63]$$

et sa réponse en fréquence vaut :

$$H(e^{2i\pi\nu}) = 1 - K e^{-2\pi\nu} \quad [64]$$

En écrivant le module de la réponse en fréquence :

$$|H(e^{2i\pi\nu})| = (1 + K^2 - 2K\cos(2\pi\nu))^{1/2} \quad [65]$$

on constate que c'est un filtre passe-haut.

Un exemple de filtre récursif est le filtre récursif du premier ordre :

$$y(n) = x(n) + K\, y(n-1) \quad [66]$$

avec la condition initiale y(-1) = 0, et $1 > K > 0$. On obtient facilement la réponse impulsionnelle, pour $n \geq 0$ (nulle pour $n < 0$) :

$$h(n) = K^n \quad [67]$$

la fonction de transfert de ce système vaut :

$$H(z) = \frac{1}{1 - Kz^{-1}} \quad [68]$$

et sa réponse en fréquence vaut :

$$H(e^{2i\pi\nu}) = \frac{1}{1 - Ke^{-2i\pi\nu}} \quad [69]$$

En écrivant le module de la réponse en fréquence :

$$|H(e^{2i\pi\nu})| = \frac{1}{(1 + K^2 - 2K\cos(2i\pi\nu))^{1/2}} \quad [70]$$

on constate que c'est un filtre passe-bas.

Un autre exemple de filtre récursif est le filtre du second-ordre :

$$y(n) = x(n) + a_1 y(n-1) + a_2 y(n-2) \quad [71]$$

Si l'on considère que $y(-1) = y(-2) = 0$.

La fonction de transfert de ce filtre est :

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2}} \quad [72]$$

Ce filtre peut se comporter de deux façons distinctes, suivant les racines du dénominateur de sa fonction de transfert. Si ses racines sont réelles (c'est à dire si $a_2 \geq -a_1^2/4$), alors le système est équivalent à deux systèmes du premier ordre en série.

Si ses racines sont complexes, elle sont conjuguées (z_p, z_p^*), et la réponse en fréquence est :

$$H(e^{2i\pi\nu}) = \frac{1}{1 - 2Re(z_p)e^{-2i\pi\nu} + |z_p|e^{-4i\pi\nu}} \quad [73]$$

Le module de cette réponse en fréquence montre un maxima unique : le filtre récursif du second ordre est un résonateur digital.

5. Compléments sur la numérisation

5.1. *La chaîne du traitement numérique*

Le son est une variation de pression aérienne, variation perçue par l'oreille. En toute généralité le signal sonore est donc un champ acoustique dans lequel l'auditeur est immergé. Pour le traitement du signal sonore, on travaille en général sur un signal électrique analogue au signal acoustique, et capté par un ou plusieurs microphones. Ce signal électrique analogique peut être traité dans le domaine électrique ou électronique, ou bien dans le domaine numérique. Le traitement numérique du signal, en particulier du signal sonore, est ainsi l'enchaînement de cinq opérations :

1. transduction acoustico-électrique. Le signal acoustique, vibration aérienne, est converti en signal électrique, en général par un microphone. Le signal électrique est analogue au signal acoustique capté en un point du champ acoustique, aux propriétés du transducteur près. Ainsi le champ acoustique est transformé en 1 (mono) ou 2 (stéréo) signaux électriques monodimentionnels.

2. la numérisation. Le signal sonore analogique, ou plutot son analoque électrique, est converti en une suite de nombres (numérisation). Pour cela deux oprétations interviennent : l'échantillonnage et la quantification. La numérisation est souvent dénommée conversion analogique-numérique (CAN, en anglais Analog/Digital Conversion, ADC).

3. le traitement numérique du signal. Les processeurs de signal numérique, de plus en plus puissants, permettent toutes sortes de traitement sur la série de nombres qui représente désormais le son.

4. la reconstruction du signal électrique analogique, ou conversion numérique-analogique (CNA, en anglais Digital/Analog Conversion, DAC). A partir de la suite de nombres échantillonnés et quantifiés, il d'agit de reconstituer un signal électrique continu.

5. transduction éléctro-acoustique. Le signal électrique analogique est amplifié, puis transformé en signal acoustique, généralement par un haut-parleur. Le signal acoustique rayonne depuis le haut-parleur pour reconstituer un champ acoustique.

Bien sur, cette chaîne de traitement peut-être utilisée partiellement, pour l'analyse du signal (les traitements seront par exemple des estimations spectrales, ou de paramètres), pour la synthèse du signal (on calcule alors un signal numérique pour synthétiser un signal acoustique), ou pour l'analyse/modification/synthèse.

Avant de parler des traitements numériques du signal audio, il faut donc examiner les extrémités de la chaîne de traitement. Les aspects électroacoustiques ne seront pas envisagés ici. On se limitera donc aux effets des CAN et CNA.

5.2. *échantillonnage*

L'ordinateur ne peut travailler que sur des séquences de nombres. Le traitement digital résulte d'une double discrétisation du signal analogique.

Un signal discret s'obtient par échantillonnage du signal analogique à des instants t_n. Les instants d'échantillonnage sont en général choisis régulièrement espacés : $t_n = nT_e$ ou T_e représente le pas d'échantillonnage et $1/T_e = F_e$ la fréquence d'échantillonnage. Le signal digital, à temps discret, est une suite d'échantillons, dépendants de l'indice n, et obtenus par échantillonnage du signal à temps continu :

$$x(n) = x(nT_e) \qquad [74]$$

Par la suite l'indice n sera dénommé temps discret, par abus de langage (le temps discret est nT_e), en considérant une fréquence d'échantillonnage de 1, pour simplifier les expressions mathématiques.

Le théorème de Shannon, ou théorème d'échantillonnage se formule de la façon suivante : Si un signal $x(t)$ possède une transformée de Fourier dont le support est borné, alors il peut être reconstruit exactement et de manière unique à partir de ses échantillons si la fréquence d'échantillonnage est égale au moins au double de la plus haute fréquence présente dans le signal.

On doit donc choisir la fréquence d'échantillonnage F_e telle que : $F_e \geq \Omega$ si $\tilde{x}(\nu) = 0$ pour $|\nu| > \Omega$. La demi-fréquence d'échantillonnage se nomme fréquence de Nyquist : c'est la plus haute fréquence que l'on doit rencontrer dans le signal pour pouvoir l'échantillonner correctement.

Dans le cas des signaux musicaux, on traite généralement des signaux réels. Rappelons que dans ce cas, le spectre du signal obéit à la symétrie Hermitienne :

$$x^*(\nu) = x(-\nu) \qquad [75]$$

c'est à dire que le module du spectre est symétrique, et la phase antisymétrique :

$$|x(\nu)| = |x(-\nu)| \qquad \text{et} \qquad \arg\, x(\nu)| = -\arg\, x(-\nu) \qquad [76]$$

Ainsi, par exemple, une sinusoïde de fréquence f possède du point de vue spectral deux contributions impultionnelle aux fréquences f et $-f$. La bande de fréquence utile pour un échantillonnage à la fréquence F_e est la bande de fréquence comprise entre $-F_e/2$ et $F_e/2$, de largeur F_e.

Nous allons maintenant voir ce que signifie échantillonner correctement. Pour montrer le théorème d'échantillonnage, il faut passer dans le domaine spectral. Le signal échantillonné peut s'écrire :

$$x_e(t) = x(t) \times \sum_{n=-\infty}^{+\infty} \delta(t - nT_e) = x(t)s(t) \qquad [77]$$

On peut calculer le spectre du signal échantillonné par un calcul direct de la transformée de Fourier. En notant $x(n) = x_e(nT_e)$, la transformée de Fourier du signal échantillonné est obtenue par intégrale de Fourier, et vaut :

$$\tilde{x_e}(\nu) = \sum_{n=-\infty}^{+\infty} x(n)e^{-2i\pi\nu nT_e} \qquad [78]$$

avec la transformée de Fourier inverse :

$$x_e(nT_e) = x(n) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \tilde{x_e}(\nu)e^{2i\pi\nu nT_e}d\nu \qquad [79]$$

On remarque que la transformée de Fourier du signal échantillonné de l'équation 78 possède la forme d'une série de Fourier, et que le signal échantillonné de l'équation 79 lui même possède la forme des coefficients de Fourier. Ainsi, le spectre du signal échantillonnée est une fonction périodique, puisque c'est une fonction développable en séries de Fourier.

Ce spectre étant périodique, montrons qu'un période fondamentale du spectre correspond au spectre du signal continu. Dans la définition du signal échantillonné, intervient un peigne de Dirac. D'après la formule de la somme de Poisson, le peigne de Dirac $s(t)$ s'exprime comme :

$$s(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT_e) = \frac{1}{T_e}\sum_{m=-\infty}^{+\infty} e^{2i\pi mF_e t} \qquad [80]$$

soit :

$$x_e(t) = x(t)s(t) = \frac{1}{T_e} \sum_{m=-\infty}^{+\infty} x(t)e^{2i\pi m F_e t} \quad [81]$$

la transformée de Fourier du signal échantillonné vaut donc :

$$\tilde{x_e}(\nu) = \frac{1}{T_e} \sum_{m=-\infty}^{+\infty} \tilde{x}(\nu - mF_E) \quad [82]$$

L'effet de l'échantillonnage est de répliquer, de périodiser, le spectre du signal continu correspondant. La période du spectre périodisé est égale à la fréquence d'échantillonnage F_e. Pour le spectre du signal échantillonné, la formule de somme de Poisson peut s'écrire comme :

$$\tilde{x_e}(\nu) = \sum_{n=-\infty}^{+\infty} x(nT_e)e^{-2i\pi\nu nT_e} = \frac{1}{T_e} \sum_{m=-\infty}^{+\infty} \tilde{x}(\nu - mF_e) \quad [83]$$

Dans l'intervalle spectral fondamental (intervalle de Nyquist) $[-F_e/2, F_e/2]$ le spectre du signal échantillonné est donc égal au spectre du signal continu, si celui-ci ne possédait pas de fréquences au dessus de la fréquence de Nyquist. Si il en possédait, alors, la périodisation entraîne un phénomène de repliement du spectre du signal continu dans la bande de Nyquist.

Si le signal a été échantillonné sous les conditions du théorème de Shannon, il n'y a pas de repliement spectral (car le spectre du signal continu est nul en dehors de l'intervalle de Nyquist). On nomme "bande de garde" la différence $\Delta = F_e - 2f_{max}$, qui mesure l'espacement entre deux périodes du spectre de signal échantillonné. Ainsi, tout traitement du spectre du signal échantillonné dans la bande de Nyquist est équivalent à un traitement du spectre du signal continu, que l'on peut reconstruite exactement.

Evidemment si la fréquence d'échantillonnage devient très grande, les répliques spectrales duent à la périodisation sont rejetées à l'infini, et seul le spectre du signal continu reste.

Avant l'opération d'échantillonnage, il faut donc mettre garantir que le spectre du signal ne contient plus de fréquences au dessus de la fréquence de Nyquist. C'est la raison d'être du filtre anti-repliement, ou pré-filtre du système de numérisation. Idéalement, il s'agit d'un filtre passe bas, de réponse plate entre $-F_e/2$ et $F_e/2$, et nulle en dehors de la bande de Nyquist. Pratiquement, la réponse peut être rendue plate dans la bande passante, par égalisation digitale, mais il reste une certaine largeur de bande de transition, et l'atténuation n'est pas parfaite dans la bande d'atténuée.

5.3. *Reconstruction du signal analogique*

Pour reconstruire le signal continu, idéalement il faut isoler dans le spectre du signal échantilloné la période spectrale principale, celle qui est dans la bande de Nyquist. Ainsi, le reconstructeur analogique idéal est un filtre passe-bas idéal, de bande passante $-F_e/2, F_e/2$, de fréquence de coupure $F_e/2$. En effet, à partir des échantillons de signal, on peut voir la reconstruction du signal analogique comme un problème d'interpolation : il s'agit de reconstituer le signal entre les valeurs à temps discret des échantillons. Une façon d'interpoler est de filtrer passe-bas le signal.

Nous allons étudier la réponse inpulsionnelle du filtre passe-bas $h(t)$ de reconstruction analogique. Il faut d'abord reconstruire un signal analogique impulsionnel $y_e(t)$ à temps continu partir du signal échantillonné (noté y, après un éventuel traitement numérique) $y(n) = y(nT_e)$:

$$y_e(t) = \sum_{n=-\infty}^{+\infty} y(nT_e)\delta(t - nT_e) \qquad [84]$$

ce signal analogique est ensuite filtré par le filtre passe-bas interpolateur :

$$\begin{aligned} y_a(t) &= \int_{-\infty}^{+\infty} y_e(\tau)h(t-\tau)d\tau & [85] \\ &= \sum_{n=-\infty}^{+\infty} y(nT_e)h(t - nT_e) & [86] \end{aligned}$$

Donc l'interpolation est obtenue par sommation des réponses impulsionnelles pondérées par les échantillons. Dans le domaine spectral, on a :

$$\tilde{y_a}(\nu) = \tilde{h}(\nu)\tilde{y_e}(\nu) = \tilde{h}(\nu)\frac{1}{T_e}\sum_{m=-\infty}^{+\infty} \tilde{y}(\nu - mF_e) \qquad [87]$$

Pour une reconstruction parfaite, idéale, du signal analogique, il faut que le spectre du signal analogique reconstruit soit égal au spectre analogique de départ. Il faut donc isoler la période principale du spectre du signal échantillonné :

$$\tilde{y_e}(\nu) = \frac{1}{T_e}\tilde{y}(\nu) \qquad \text{pour} - \frac{F_e}{2} \leq \nu \leq \frac{F_e}{2} \qquad [88]$$

Ainsi le filtre de reconstruction idéal doit avoir comme gain $\tilde{h}(\nu) = T_e$ pour $-F_e/2 \leq \nu \leq F_e/2$, et $\tilde{h}(\nu) = O$ en dehors de la bande de Nyquist. Ceci montre le théorème d'échantillonnage : le signal $y(t)$ peut être reconstruit à partir de ces échantillons par

$$y(t) = \sum_{-\infty}^{\infty} y(nT_e)h(t - nT_e) \qquad [89]$$

avec la réponse impulsionnelle :

$$h(t) = \int_{-\infty}^{\infty} \tilde{h}(\nu)e^{2i\pi\nu t}d\nu = \int_{-F_e/2}^{F_e/2} T_e e^{2i\pi\nu t}d\nu = \frac{\sin(\pi t F_e)}{\pi t F_e} \qquad [90]$$

soit :

$$y(t) = \sum_{-\infty}^{\infty} y(nT_e)\frac{\sin[\pi f_e(t - nT_e)]}{\pi f_e(t - nT_e)} \qquad [91]$$

Malheureusement, une telle reconstruction n'est pas physiquement réalisable, car elle n'est pas causale. En effet, la réponse impulsionnelle $h(t)$ est infinie et bilatérale.

Une approximation pratique du reconstructeur idéal est le reconstructeur en escalier. Pour une telle reconstruction, la réponse impulsionnelle vaut :

$$h(t) = \begin{cases} 1, & \text{si} \quad 0 \leq t \leq T_e \\ 0, & \text{sinon} \end{cases} \qquad [92]$$

Dans ce cas, le gain du filtre passe-bas de reconstruction est donné (en calculant la transformée de Fourier de $h(t)$) par :

$$\tilde{h}(\nu) = T_e \frac{\sin(\pi\nu T_e)}{\pi\nu T_e} e^{-\pi i\nu T_e} \qquad [93]$$

notons que le lobe principal de ce gain a exactement une largeur égale à F_e, mais une attenuation de seulement d'environ $-3,9dB$ à la fréquence de Nyquist $(F_e/2)$. Ainsi, les réplications spectrales dues à l'échantillonnage ne seront pas filtrée par le reconstructeur en escalier : c'est un mauvais filtre passe-bas, mais il est causal, stable et très facile à mettre en oeuvre.

Pour obtenir un signal analogique fidèle au signal digital correspondant, il faut refiltrer le signal analogique en escalier par un filtre passe-bas analogique

qui supprime les périodes spectrales mal filtrées au niveau digital, donc de fréquence de coupure $F_e/2$. Le rôle de ce filtre est de supprimer les images périodisées de la période principale (filtre anti-image).

On peut aussi rajouter une égalisation numérique du signal avant le passage par le reconstructeur en escalier, pour compenser le gain qui n'est pas plat du reconstructeur dans la période principale. Alors ce filtre d'égalisation est le filtre numérique inverse du fitre reconstructeur dans la période principale. La combinaison (triple) d'égaliseur, reconstructeur en escalier et post-filtre anti-image permet de réaliser avec une très grande fidélité le convertisseur CNA idéal.

Ainsi un système de traitement numérique doit comprendre au minimum 5 éléments :

1. un filtre passe-bas anti-repliement, qui limite le spectre à la bande de Nyquist.
2. un échantillonneur et quantificateur (CAN).
3. un processeur de signal numérique.
4. un reconstructeur en escalier (précédé éventuellement d'un égaliseur numérique)
5. un filtre passe-bas anti-image, qui limite le spectre à la bande de Nyquist.

5.4. *Quantification*

Le signal analogique passe par un circuit échantillonneur-bloqueur pour être digitalisé. Sa valeur est maintenue constante pendant une période T_e, et les amplitudes du signal sont discrétisées.

Alors que le signal analogique est à valeurs réelles, le signal digital est à valeurs discrètes. En général, un codage binaire des amplitudes est utilisé, et le nombres de pas de quantification est noté en bits. Un mot de B bits donne un nombre de niveaux égal à 2^B. Le convertisseur est caractérisé par sa gamme pleine-échelle R, qui est divisée en 2^B niveaux de quantification, chacun distant d'un pas de quantification ou résolution de $Q = R/2^B$.

En pratique, les valeurs de R sont de quelques volts. La quantification se fait soit par troncature, soit par arrondis sur la valeur du niveau de quantification le plus proche de la valeur mesurée. Pour un convertisseur bipolaire, la valeur quantifiée $x_Q(nT_e)$ est comprise entre -R/2 et R/2. L'erreur de quantification

est l'erreur que l'on commet en utilisant le signal quantifié au lieu du signal réel. Elle est définie par :

$$e(nT_e) = x_Q(nT_e) - x(nT_e) \qquad [94]$$

la valeur ce cette erreur pour un échantillon est comprise entre deux niveaux de quantification, donc $-Q/2 \leq e \leq Q/2$. L'erreur maximum est donc (en valeur absolue) $Q/2$, ce qui est une suréstimation de l'erreur réelle. L'erreur est en moyenne nulle, et l'erreur quadratique moyenne vaut :

$$\bar{e^2} = \frac{1}{Q}\int_{-Q/2}^{Q/2} e^2 de = \frac{Q^2}{12} \qquad [95]$$

soit la valeur RMS (root mean square) absolue, $\sqrt{\bar{e^2}} = Q\sqrt{12}$. On peut supposer en première approximation que l'erreur est uniformément répartie dans l'intervalle $[-Q/2, Q/2]$. Si on considère que R et Q représentent la dynamique du signal et du bruit, le rapport signal sur bruit de la quantification est donné par :

$$RSB = 20\log_{10}(\frac{R}{Q}) = 20\log_{10}(2^B) = B\log_{10}(2) \simeq 6,0206B \qquad \text{dB} \qquad [96]$$

Ainsi la dynamique que l'on peut représenter est d'environ 6 dB par bit. Pour une quantification sur 8, 12, 16 et 20 bits, le nombre de niveaux est 256, 4096, 65636 et 1048576, qui correspondent à environ 48, 72, 96 et 120 dB de dynamique.

En général, les statistiques du bruit de quantification sont très complexes. Néanmoins, sous la condition grandes amplitudes et large bande, on peut considérer que le bruit est blanc, stationnaire, non corrélé avec le signal, et que sa densité de probabilité est uniformément réparti sur l'intervalle $[-Q/2, Q/2]$. Ce cas représente une situation ou toutes les valeurs du quantificateur sont utilisées, et ou il y a de nombreux passages par chaque niveau (en particulier 0). Comme modèle, on peut alors considérer que le signal quantifié est la somme du signal original et d'un terme de bruit :

$$x_Q(n) = x(n) + e(n) \qquad [97]$$

Ce bruit est uniformément réparti sur l'intervalle, et doit donc avoir la densité de probabilité :

$$p(e) = \begin{cases} \frac{1}{Q} & \text{si} \quad -Q/2 \leq e \leq Q/2 \\ 0, & \text{sinon} \end{cases} \qquad [98]$$

le terme $1/Q$ est nécessaire pour avoir :

$$\int_{-Q/2}^{Q/2} p(e)de = 1 \qquad [99]$$

Alors la puissance moyenne ou variance ce ce bruit vaut :

$$\sigma_e^2 = E[e^2(n)] = \int_{-Q/2}^{Q/2} e^2 p(e)de = \frac{Q^2}{12} \qquad [100]$$

Dans le cas où l'hypothèse des grandes amplitudes et large bande n'est pas vérifié, le bruit de quantification prend des propriétés statistiques très difficile à étudier : il peut être quasi-périodique et fortement corrélé avec le signal. Par contre l'effet de la quantification est facilement perceptible : c'est le bruit de granulation, comme des clics plus ou moins régulier dans le signal.

5.5. *Tremblotement (dither)*

Dans le cas des signaux de faible amplitude, ou a bande étroite (par exemple une sinusoïde), le bruit de granulation dû à la quantification devient très génant. Une technique assez simple pour le corriger est l'ajout d'un "tremblotement" (anglais "dither") au signal, afin de redonner un carractère blanc et uniforme au bruit de quantification.

Le dither est soit analogique, un bruit r analogique est ajouté avant la quantification, soit digital, un bruit numérique est ajouté au signal numérique, avant une re-quantification. Dans le cas de requantification, on change souvent la fréquence d'échantillonnage. Au signal d'entrée $x(n)$ est ajouté le signal de dither r, ce qui résulte en un signal $y(n) = x(n) + r(n)$, qui sera quantifié en $y_Q(n)$: on parle alors de dither non-soustractif. L'erreur ϵ contient deux termes, un terme de quantification et un de tremblotement :

$$\epsilon(n) = y_Q(n) - x(n) = y(n) + e(n) - x(n) = x(n) + r(n) + e(n) - x(n) = r(n) + e(n) \qquad [101]$$

Un bon choix de $r(n)$ rend les deux sources de bruit decorrélée, et ainsi la variance du bruit vaut

$$\sigma_\epsilon^2 = \sigma_e^2 + \sigma_r^2 = \frac{Q^2}{12} + \sigma_r^2 \qquad [102]$$

En pratique la densité de probabilité du dither est uniforme, triangulaire, ou gaussienne sur l'intervalle $[-Q/2, Q/2]$. Dans le cas uniforme, on a bien sur $\sigma_r^2 = Q^2/12$. On peut montrer que, pour du dither non-soustractif :

$$\sigma_\epsilon^2 = \begin{cases} \frac{Q^2}{12} & \text{sans tremblotement} \\ \frac{2Q^2}{12} & \text{tremblotement uniforme} \\ \frac{3Q^2}{12} & \text{tremblotement triangulaire} \\ \frac{4Q^2}{12} & \text{tremblotement gaussien} \end{cases} \qquad [103]$$

Soit un bruit de quantification qui double, triple, ou quadruple, et une perte de RSB de $10\log_{10} 2 = 3$, $10\log_{10} 3 = 4,8$ et $10\log_{10} 4 = 16$ dB. Des expériences ont montré que le dither triangulaire est le meilleur.

Une autre stratégie de tremblotement est le dither soustractif. Dans ce cas, après transmission ou stockage de $y_Q(n)$, on soustrait le dither au signal quantifié pour obtenir le signal de sortie $y_s(n)$, et l'erreur de quantification devient :

$$\epsilon(n) = y_s(n) - x(n) = y_Q(n) - r(n) - x(n) == r(n) + e(n) - r(n) = e(n) \qquad [104]$$

L'expérience montre que le meilleur type de dither est le dither soustractif uniforme, qui ne dégrade donc pas du tout le SNR, mais supprime le bruit de granulation dû à la quantification. Le dither non-soustractif triangulaire supprime le bruit de granulation, mais dégrade le RSB du signal quantifié.

5.6. *Suréchantillonnage*

Dans le domaine fréquenciel, et sous l'hypothèse de large bande et grandes amplitudes, le bruit de quantification est supposé blanc, soit à spectre plat. Ainsi le bruit est uniformément réparti dans l'intervalle de Nyquist $[-F_e/2, F_e/2]$. La puissance moyenne du bruit est ainsi uniformément répartie en fréquence, donc avec une densité spectrale de puissance :

$$\tilde{e}(\nu) = \frac{\sigma_e^2}{F_e} \qquad \text{pour} \qquad \frac{-F_e}{2} \leq \nu \leq \frac{F_e}{2} \qquad [105]$$

La densité spectrale de puissance du bruit est donc inversement proportionnelle à la fréquence d'échantillonnage, et proportionnelle au pas de quantification. Ainsi, si on utilise un suréchantillonnage, par exemple d'un facteur entier, dans la bande utile du signal la puissance du bruit de quantification diminue, car elle est répartie uniformément sur tout l'intervalle de Nyquist.

Ainsi, suréchantillonner le signal est équivalent à diminuer le pas de quantification. En effet soit F_e^1 et F_e^2 deux fréquences d'échantillonage. Si on considère des pas de quantification Q_1 et Q_2, et une qualité équivalente, c'est à dire des densités spectrales de puissance du bruit égales :

$$\frac{\sigma_{e1}^2}{F_e^1} = \frac{\sigma_{e2}^2}{F_e^2} \qquad [106]$$

alors on a (supposons que $F_e^1 < F_e^2$) :

$$\sigma_{e2}^2 = F_e^2 \frac{\sigma_{e1}^2}{F_e^1} = \frac{F_e^2}{F_e^1}\sigma_{e1}^2 \qquad [107]$$

donc :

$$Q_2^2/12 = R^2 2^{-2B_2}/12 = \frac{F_e^2}{F_e^1} Q_1^2/12 = \frac{F_e^2}{F_e^1} R^2 2^{-2B_1}/12 \qquad [108]$$

soit :

$$2^{2(B_1-B_2)} = \frac{F_e^2}{F_e^1} \qquad \text{ou encore} \qquad B_1 - B_2 = 0.5 \log_2(\frac{F_e^2}{F_e^1}) \qquad [109]$$

Ainsi, chaque suréchantillonnage d'un facteur 2 permet de faire l'économie d'un demi-bit : par exemple si on suréchantillonne d'un facteur 4, pour une qualité égale de quantification on gagne 1 bit de quantification.

5.7. *Convertisseurs* $\Delta\Sigma$

Une autre façon de réduire le bruit de quantification est la mise en forme du bruit. L'idée est de suréchantillonner le signal, de transferer le bruit de quantification vers les hautes fréquences, puis de le filtrer avant de sous-échantillonner.

A partir du signal échantilloné à la fréquence F_e, le signal est suréchantillonné à une fréquence supérieure f_{e+}. Ainsi la largeur de bande du spectre numérique est augmentée. Un filtre de gain H_b de mise en forme du bruit

est appliqué au signal, pour filtrer le bruit de quantification dans la bande d'origine. Dans l'intervalle de Nyquist original, la puissance totale du bruit de quantification vaut :

$$\sigma_e^2 = \frac{\sigma_{e+}^2}{f_{e+}} \int_{-F_e/2}^{F_e/2} |H_n(\nu)|^2 d\nu \qquad [110]$$

Si le filtre de mise en forme du bruit est un filtre numérique d'ordre p, on peut montrer que le nombre de bits de quantification économisés est environ $p + 0,5$ si on suréchantillonne d'un facteur 2.

Une application de ce principe de suréchantillonnage est le convertisseur $\Delta\Sigma$. Le convertisseur opère à une fréquence élevée $F_e+ = LF_e$ et avec un faible nombre de bits (par exemple 1). Une paire de CNA/CAN permet de reconstruire le signal converti en analogique, de le soustraire au signal analogique de départ (partie Δ), puis de l'accumuler dans un intégrateur (partie Σ) qui moyenne l'entrée. La boucle permet de faire un filtrage passe-haut du bruit, en le rejetant ainsi den dehors de la bande de Nyquist utile (à la fréquence F_e). Le signal suréchantillonné, et quantifié sur peu de bits est ensuite sous-échantillonné, et quantifié sur un plus grand nombre de bits.

Etudions un modèle numérique de la conversion $\Delta\Sigma$ d'ordre 1, dans la partie suréchantillonnée. Le signal $x(n)$ d'entrée moins la sortie $y(n)$ du convertiseur (partie Δ) entre dans l'accumulateur de fonction de transfert H. Le bruit de quantification e est ensuite ajouté, pour former la sortie $y(n)$ du convertisseur, et la boucle continue pour l'échantillon suivant. En terme de transformée en Z, la fonction de transfert de l'accumulateur est (avec un retard pour le rendre calculable) :

$$H(z) = \frac{z^{-1}}{1 - z^{-1}} \qquad [111]$$

alors, l'équation récursive du convertisseur est :

$$H(z)(X(z) - Y(z)) + E(z) = Y(z) \qquad [112]$$

c'est à dire :

$$Y(z) = \frac{H(z)}{1 + H(z)} X(z) + \frac{1}{1 + H(z)} E(z) = z^{-1} X(z) + (1 - z^{-1}) E(z) \qquad [113]$$

ainsi, dans le domaine temporel :

$$y(n) = x(n-1) + (e(n) - e(n-1)) \qquad [114]$$

La sortie est le signal x retardé d'un échantillon, ajouté à la dérivée numérique du bruit de quantification. Le bruit de quantification est donc filtré passe haut, à la fréquence d'échantillonnage haute. Ensuite, ce signal numérique est sous-échantillonné, par un filtre de décimation, qui permet de retrouver la fréquence d'échantillonnage cible, et de reconstituer un nombre de niveaux de quantification plus élevé. L'effet du filtre passe-bas de décimation est de régénerer plus de niveaux sur le signal. Ce peut être, dans la version la plus simple un filtre moyenneur à la haute fréquence d'échantillonnage. Ensuite, ce signal moyenné est sous-échantillonné à la fréquence voulue. Le filtrage de décimation doit être à phase linéaire, pour ne pas altérer le signal, en général on choisi un filtre à réponse impulsionnelle finie (RIF). Par exemple le filtre passe-bas idéal pondéré par un fenêtre d'analyse pour en limiter la durée. Il existe des convertisseurs $\Delta\Sigma$ d'ordre plus élevé (2,3,4). mple un peigne périodique d'impulsion, ou bien un bruit blanc, ou des impulsions isolées. Le spectre de toutes ces sources possède une enveloppe spectrale plate. Ainsi le signal résultant aura une enveloppe spectrale fixée par le filtre. D'où le nom de synthèse par filtrage, ou synthèse soustractive.

6. Transformée de Fourier à court terme

Le signal sonore peut être analysé sans référence à un modèle de production. Nous allons considérer des représentations et analyse des sons qui ne font pas d'hypothèse sur la production, des représentations non-paramétriques. Pour représenter le signal sonore, on utilise des descriptions en temps et en fréquence, pour rendre compte de l'évolution dans le temps du contenu spectral des sons.

6.1. *Spectrographe*

Le premier spectrographe a été mis au point dans les années 40 aux Laboratoires Bell, et sa première réalisation commerciale, le "sonagraph" a connu un important succès. C'est un outil d'analyse du signal acoustique en temps et en fréquence. L'appareil délivre une analyse en trois dimension :

- l'abscisse représente le temps ;
- l'ordonnée la fréquence ;
- la noirceur du tracé indique l'intensité présente dans le signal au temps et la fréquence donné ;

Le principe de l'analyse, qui n'a pas varié seuls les moyens de le réaliser ayant évolué, est de filtrer le signal par un banc de filtres passe-bande. On mesure l'énergie présente à chaque instant d'analyse et à chaque fréquence d'analyse.

La largeur de bande du filtre employé (qui donne la précision fréquentielle désirée) est inversement proportionnelle à la précision temporelle que l'on peut escompter. On est donc amené à utiliser en gros deux types d'analyse :

1. l'analyse en bande étroite (typiquement 45 Hz) qui occulte les variations temporelles rapides du signal mais révèle finement sa structure fréquentielle (on distingue chaque harmonique d'un son périodique de parole avec précision) ;
2. l'analyse en bande large (typiquement 300 Hz), qui au contraire permet de bien visualiser les événements temporels, mais dont la résolution fréquentielle est faible ;

Le spectrographe est un moyen puissant d'analyse visuelle du signal acoustique, et certains experts en lecture de spectrogrammes de parole sont capable de lire les phrases analysées en regardant leurs spectrogrammes, avec un taux

de décodage acoustico-phonétique important (jusqu'à 95 % en multi-locuteur dans les meilleurs cas). Ceci laisse penser que l'information phonétique est bien conservée dans l'image spectrographique. Notons que l'information prosodique par contre est plus difficile à lire sur ce type de représentation.

Les spectrographes sont aujourd'hui réalisés à l'aide de la transformées de Fourier à court terme (en utilisant l'algorithme FFT). La transformée de Fourier à court terme a été définie pour considérer le spectre d'un signal sur une plage temporelle localisée.

A partir du signal à une dimension $x(\tau)$ on définit un signal à deux dimensions $x^w(t,\tau)$ qui représente à t fixé le signal x vu à travers la *fenêtre d'analyse* w centrée en t, dénommée *trame d'analyse*.

$$x^w(t,\tau) = x(\tau)w(t-\tau)$$

La transformée de Fourier à court terme est la transformée de Fourier du signal x^w, notée $\tilde{x}^w(t,\nu)$:

$$\tilde{x}^w(t,\nu) = \int x^w(t,\tau)e^{-2i\pi\nu\tau}d\tau$$

La représentation spectro-temporelle de $x(\tau)$ par $\tilde{x}^w(t,\nu)$ peut recevoir une interprétation en terme de filtrage linéaire. En reconnaissant la forme d'une convolution dans l'expression de $\tilde{x}^w(t,\nu)$, si on la considère comme une fonction de t, ce qui signifie que l'on fait glisser la fenêtre d'analyse, on peut réécrire l'analyse comme :

$$\tilde{x}^w(t,\nu) = w(t) * x(t)e^{-2i\pi\nu t}$$

ce qui permet d'interpréter la transformation de Fourier à court terme comme le filtrage linéaire (convolution temporelle par le filtre de réponse impulsionnelle $w(t)$) du signal $x_\nu(t)$ résultant de $x(t)$ modulé par $e^{-2i\pi\nu t}$: $w(t)$ porte ainsi le nom de filtre d'analyse. Les fenêtres d'analyse spectrale $w(t)$ peuvent être alors considérées comme réponses impulsionnelles de filtres passe-bas. Pour chaque instant d'analyse, le spectrogramme noté SX est obtenu comme le carré du module de l'analyse de Fourier à court terme :

$$SX(t,\nu) = |\tilde{x}^w(t,\nu)|^2$$

L'interprétation en terme de banc de filtres permet de mettre en lumière la largeur de bande due au filtre d'analyse. La durée et les caracteristiques spectrales de la fenêtre temporelle d'analyse fixent la largeur de bande effective du

filtre d'analyse. Un compromis courant en traitement de la parole est d'utiliser la fenêtre de Hamming.

La dynamique réellement representée par la noirceur du traçé est de l'ordre de 30 dB.

Le nombre de bandes d'analyse d'un analyse par FFT de signaux échantillonés et égal à la moitié du nombre de points de la FFT (typiquement 256 ou 512). Pour obtenir un tracé interprétable, on reconduit une analyse environ toutes les 1 ou 2 millisecondes.

6.2. *Vocodeur de phase*

La première application pratique de la TFCT à l'analyse-synthèse du son (de la parole en l'occurence) a été le vocodeur de phase. Dans ce système, le signal est synthétisé comme la somme des sorties d'un banc de N filtres passe-bande uniforme :

$$x(t) \simeq \sum_{n=1}^{N} x_n(t) \qquad [115]$$

avec pour chaque filtre une réponse impulsionnelle de la forme :

$$h_n(t) = w(t)e^{2i\pi\nu_n t} \qquad [116]$$

Les fréquences centrales d'analyse ν_n sont régulièrement réparties sur la bande passante du signal. Si B est cette bande, on a $\nu_n = n \times B/N$.

alors la sortie de chaque filtre vaut (en prenant un filtre causal, physiquement réalisable) :

$$x_n(t) = \int_{-\infty}^{t} x(\tau)w(t-\tau)e^{2i\pi\nu_n(t-\tau)}d\tau = e^{2i\pi\nu_n t}\int_{-\infty}^{t} x(\tau)w(t-\tau)e^{-2i\pi\nu_n\tau}d\tau \qquad [117]$$

Ainsi la sortie de chaque filtre peut être considéré comme une fréquence porteuse $e^{2i\pi\nu t}$ modulée par la TFCT complexe du signal. On remarque que cette interprétation "passe-bande" est légèrement différente de la forme "passe-bas" définie plus haut pour le spectrographe. Dans la forme passe-bas, le signal et l'exponentielle d'analyse sont synchrones (dépendent de la même variable

temporelle), et la fenêtre d'analyse est asynchrone. Dans l'interprétation passe-bande, la fenêtre d'analyse et l'exponentielle d'analyse sont synchrones, et le signal est asynchrone. La formule de synthèse est donc :

$$x(t) \simeq \sum_{n=1}^{N} x_n(t) = \sum_{n=1}^{N} e^{2i\pi\nu_n t} \int_{-\infty}^{t} x(\tau) w(t-\tau) e^{-2i\pi\nu_n \tau} d\tau \qquad [118]$$

Ce qui est finalement une version réduite (avec un nombre N de bandes d'analyse) de la formule de synthèse par transformée de Fourier inverse :

$$x(t) = \int_{-\infty}^{\infty} e^{2i\pi\nu t} \int_{-\infty}^{\infty} x(\tau) w(t-\tau) e^{-2i\pi\nu\tau} d\tau d\nu \qquad [119]$$

On peut réécrire la sortie des filtres en fonction du module et de la phase de la TFCT sous la forme :

$$x_n(t) = |\tilde{x}(t,\nu_n)| \ \exp(2i\pi\nu_n t + \varphi(\tilde{x}(t,\nu_n))) \qquad [120]$$

Donc, le signal de chaque bande est obtenu par modulation d'amplitude et de phase d'une porteuse sinusoïdale à la fréquence ν_n. En pratique, le module $|\tilde{x}(t,\nu_n)|$ évolue lentement dans le temps et est une fonction bornée. Ainsi, on peut sans perte de qualité sous-échantillonner ce module. Par contre, la phase $\varphi(\tilde{x}(t,\nu_n))$ n'a pas ces bonnes propriétés, en particulier ce n'est pas une fonction bornée et très régulière. Pour cette raison, on préfère représenter la phase par sa dérivée temporelle $\varphi'(\tilde{x}(t,\nu_n))$, qui elle est à la fois bornée et plus régulière. On peut reconstituer la phase, à une constante additive initiale près, par l'intégrale de sa dérivée :

$$\hat{\varphi}(\tilde{x}(t,\nu_n)) = \int_0^t \varphi'(\tilde{x}(t,\nu_n)) dt \qquad [121]$$

en espérant que cette différence d'une constante ne détériorera pas trop le signal de synthèse. La synthèse est donc réalisée par la somme d'un banc d'oscillateurs. Chaque composante est obtenue pas la modulation d'un oscillateur, en fonction du module et de l'intégrale de la dérivée de phase de la TFCT, d'où le nom de vocodeur de phase.

6.3. *Interprétation en banc de filtre de la TFCT*

Nous avons vu le spectrogramme comme une outil d'analyse visuelle du signal, et le vocodeur de phase comme la somme des sortie d'un banc de filtres. La

transformée de Fourier à court terme (TFCT) peut également s'appliquer pour la représentation et la modification des signaux numériques, et pas seulement des signaux continus. Soit un signal numérique $x(n)$, et le signal à court terme centré au point n, $x(n,m) = w(n-m)x(m)$, la TFCT de x (sous la forme passe-bas) est définie par :

$$\tilde{x}(n,\omega) = \sum_{m=-\infty}^{\infty} x(n,m)e^{-i\omega m} = \sum_{m=-\infty}^{\infty} w(n-m)x(m)e^{-i\omega m} \qquad [122]$$

Remarquons que cette somme infinie est en pratique limitée aux N échantillons pour lesquels la fenêtre d'analyse w est non nulle. Le signal peut être reconstruit par transformée de Fourier inverse de l'équation 122 :

$$x(n,m) = w(n-m)x(m) = \frac{1}{2\pi}\int_{\omega=-\pi}^{\pi} \tilde{x}(n,\omega)e^{i\omega m}d\omega \qquad [123]$$

donc, si la fenêtre est normalisée à $w(0) = 1$, en prenant $m = n$:

$$x(n) = \frac{1}{2\pi}\int_{\omega=-\pi}^{\pi} \tilde{x}(n,\omega)e^{i\omega n}d\omega \qquad [124]$$

L'équation 122 peut s'interpréter comme une équation de convolution discrète $w(n) * x(n)\exp(-i\omega n)$ donc comme le filtrage passe-bas du signal démodulé à la fréquence ω. La formule de synthèse 124 correspond alors à la sommation continue (intégration) d'un ensemble d'oscillateurs commandés, dans le temps et pour chaque fréquence, par les sorties des filtres passe-bas d'analyse. On peut également envisager ce processus comme la sommation des sorties $f_\omega(n)$ d'un ensemble de filtres passe-bandes de réponses impulsionnelles $w(n)e^{i\omega n}$, obtenus lorsqu'on fixe la fréquence ω.

$$x(n) = \frac{1}{2\pi}\int_{\omega=-\pi}^{\pi} f_\omega(n)d\omega \qquad \text{avec} \qquad [125]$$

$$f_\omega(n) = \tilde{x}(n,\omega)e^{i\omega n} = (w(n) * x(n)e^{-i\omega n})e^{i\omega n} = w(t)e^{i\omega n} * x(n) \qquad [126]$$

Pratiquement, un nombre fini de filtres suffit. Considérons N filtres passe-bandes f_k choisi avec des fréquences centrales uniformément espacées, telles que $\omega_k = 2\pi k/N$; pour $k = 0, 1, \ldots, N-1$. Si le nombre de filtres est suffisant (soit $N \geq K$ ou K est la durée de la fenêtre d'analyse w), on peut reconstruire le signal comme :

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} f_k(n) = \frac{1}{N}\sum_{k=0}^{N-1} \tilde{x}(n,\omega_k)e^{i\omega_k n} \quad [127]$$

$$= \frac{1}{N}\sum_{k=0}^{N-1} e^{i\omega_k n}\sum_{m=-\infty}^{\infty} x(n,m)e^{-i\omega_k m} \quad [128]$$

Une variante de la méthode du banc de filtre est le vocodeur de phase. Nous avans vu que cette méthode de TFCT représente les échantillons spectraux à court terme par leurs amplitudes et phases, ce qui est pratique pour le codage ou la modification et a connu un grand succès. Pour l'implémentation numérique du vocodeur de phase, on calcule les parties réélles $a(n)$ et imaginaires $b(n)$ de la TFCT, puis le module et la phase par :

$$|\tilde{x}(n,\omega_k)| = \sqrt{a^2+b^2} \quad [129]$$

$$\varphi'(\tilde{x}(n,\omega_k)) = fraca'b - b'aa^2 + b^2 \quad [130]$$

$$= \frac{1}{F_e}[\varphi(\tilde{x}(n+1,\omega_k)) - \varphi(\tilde{x}(n,\omega_k))] \quad [131]$$

$$= \frac{1}{F_e}\frac{[a(n+1)-a(n)]b(n) - [b(n+1)-b(n)a(n)]}{a^2(n)+b^2(n)} \quad [132]$$

6.4. *Interprétation par blocs de la TFCT*

L'interprétation en banc de filtres de la TFCT revient à découper le plan temps-fréquence en bandes fréquencielles. En pratique, l'analyse de Fourier à court terme est implémentée à l'aide d'algotithmes de transformée de Fourier rapide (FFT). Il faut donc pouvoir écrire les formules d'analyse et de synthèse sous la forme de transformée de Fourier discrête (TFD). Nous allons voir que c'est une autre façon d'interpréter la transformation de Fourier à court terme, qui est considérée comme une analyse par blocs temporels. Soit le signal à court terme obtenu par positionnement d'une fenêtre d'analyse à la date n sur le signal $x(m)$:

$$x(n,m) = w(n-m)x(m) \quad [133]$$

Si la fenêtre est de support fini, les seules valeurs non nulles de ce signal à court terme sont celles pour lesquelles $w(n-m) \neq 0$. Ainsi le même signal à court terme est obtenu en définissant l'origine des temps à la date n, par le changement de variable $l = m - n$:

$$x(n,m) = w(-l)x(l+n) \quad [134]$$

La transformée de Fourier à court terme de ce signal prend alors la forme de la TFD du signal positionné en n :

$$\tilde{x}(n,\omega_k) = \sum_{l=O}^{N} w(-l)x(l+n)e^{-i\omega_k l} \quad [135]$$

Par transformée de Fourier discrète inverse, on peut écrire une relation analogue à 123 :

$$x(n,l) = w(-l)x(l+n) = \frac{1}{N}\sum_{k=0}^{N-1} \tilde{x}(n,\omega_k)e^{i\omega_k l} \quad [136]$$

Le signal $x(n,l)$ peut être considéré comme un bloc temporel, que l'on peut échantillonner à la cadence de R échantillons, soit $x(sR,l) = w(-l)x(l+sR)$. A partir de cette série de signaux à court terme, on retrouve le signal x, en sommant par rapport à s (à un terme de fenêtre constant près), en faisant le changement de variable inverse $l + sR = m$:

$$\sum_{s=-\infty}^{\infty} w(-l)x(l+sR) = \sum_{s=-\infty}^{\infty} w(sR-m)x(m) = x(m)\sum_{s=-\infty}^{\infty} w(sR-m) \quad [137]$$

Chaque bloc, à s fixé, donne un spectre de Fourier à court terme. La sommation avec recouvrement (en s) des transformées inverses de chaque spectre discrêt à court terme redonne le signal initial à un instant m :

ce qui donne comme formule de synthèse par blocs :

$$x(m) = \frac{\sum_{p=-\infty}^{\infty} x(pR,m)}{\sum_{p=-\infty}^{\infty} w(pR-m)} = \frac{\sum_{p=-\infty}^{\infty} \frac{1}{N}\sum_{k=0}^{N-1} \tilde{x}(pR,\omega_k)e^{i\omega_k m}}{\sum_{p=-\infty}^{\infty} w(pR-m)} \quad [138]$$

Cette interprétation montre que le signal peut être représenté par une double somme en temps et en fréquence de composantes, qui sont localisées en temps et en fréquence. On rencontre souvent le terme de synthèse OLA (overlap-add ou recouvrement-addition) pour la formule de synthèse par blocs. Une variante de la méthode par bloc fait intervenir un filtre interpolateur f, passe-bas pour la resynthèse. La formule de synthèse devient alors :

$$x(m) = \sum_{p=-\infty}^{\infty} f(pR - m)x(pR, m) \qquad [139]$$

L'avantage de cette méthode est de pouvoir modifier les signaux à court terme avant resynthèse.

6.5. *Modification et reconstruction*

Une des utilisation les plus importantes de la TFCT est la modification du signal, par exemple le filtrage, l'interpolation, la compression, le débruitage etc. En général, une étape de modification est insérée entre analyse et synthèse. Le problème de la validité des modifications se pose donc. Il faut des contraintes spécifiques pour qu'une fonction du temps et de la fréquence $f(n, k)$ soit la TFCT $\tilde{x}(n, \omega_k)$ d'un signal $x(m)$. Par exemple, la modification de $\tilde{x}(n, \omega_k)$ ne donne pas forcément un signal valide par TFCT inverse. Cependant des méthodes de projection existent pour modifier le signal dans l'espace de la TFCT et resynthétiser correctement un signal temporel modifié. Une de ces méthodes a été proposée dans. Soit $\tilde{y}(sR, \omega)$ la TFCT modifiée d'un signal. On peut donc rechercher le signal $x(n)$ le plus proche, au sens de la distance des moindres carrés de cette TFCT modifiée. Soit d cette distance dans le domaine spectral :

$$d = \sum_{s=-\infty}^{\infty} \frac{1}{2\pi} \int_{\omega=-\pi}^{\pi} [\tilde{y}(sR, \omega) - \tilde{x}(sR, \omega)]^2 d\omega \qquad [140]$$

Dans le domaine temporel, par le théorème de Parseval, on peut écrire :

$$d = \sum_{s=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} [y(sR, k) - x(sR, k)]^2 \qquad [141]$$

Pour minimiser cette distance, on met à zéro la dérivée partielle de d par rapport à $x(k)$. On calcule par transformée de Fourier discrète inverse les échantillons temporels de la TFCT modifiée, et comme on a $x(sR, k) = w(sR - k)x(k)$, un calcul simple donne le résultat :

$$x(k) = \frac{\sum_{s=-\infty}^{\infty} w(sR - k)y(sR, k)}{\sum_{s=-\infty}^{\infty} w^2(sR - k)} \qquad [142]$$

Cette formule de synthèse est très proche de la formule de synthèse par bloc de l'équation 138.

6.6. *Applications de la TFCT*

La visualisation du spectre évolutif du signal par le spectrogramme a été la première application de l'analyse de Fourier à court terme. Cette application est toujours particulièrement importante pour comprendre le fonctionnement des instruments de musique. Certain musiciens ont même vu dans le spectrgramme un moyen universel de notation des sons concrets, électroniques ou instrumentaux. Il faut bien admettre que cet espoir a été déçu : autant le spectrogramme permet de visualiser et d'analyser en détail des documents bien enregistrés et relativement simples, autant il ne rend pas compte de la finesse de la perception (et de la cognition) pour des sons trop riches. Par exemple la visualisation d'un extrait orchestral est trop riche et confuse pour permettre de distinguer sur le tracé toutes les finesses que l'on entend.

L'analyse-synthèse de Fourier permet de nombreuse applications de transformation du son. Le principe est d'analyser le son sous forme d'une série de trames d'analyse, avec un taux d'échantillonnage donné de ces trames. Ensuite une transformation dans le domaine spectral est appliquée au trames, éventuellement de façon variable dans le temps. Les trames sont finalement recombinées, éventuellement à un taux d'échantillonnage différent de celui d'analyse, pour former le signal de synthèse. Parmi les applications de l'analyse de Fourier à court terme, on peut citer :

1. Modification de l'échelle temporelle du signal. Il s'agit de ralentir ou d'accelérer le son, sans changer son contenu fréquenciel. Donc de simuler par exemple un geste instrumental ou une prononciation plus lente ou plus rapide. Ce problème est un problème ancien, en particulier pour la radio, ou il est nécessaire d'avoir des séquences de durées très exactement fixées.

2. Modification de l'échelle fréquencielle du signal. Il s'agit de changer la bande fréquencielle du signal, de la comprimer ou de l'étendre. Cette modification entraîne un changement du timbre du son, mais pas de changement de l'échelle temporelle.

3. Transposition, ou modification de la hauteur fondamentale du son. Dans un contexte de son plutôt monophonique, il s'agit de modifier la fréquence fondamentale (donc la hauteur perçue (anglais "pitch") du son). Cette modification ne doit pas affecter la durée du son, elle doit en être indépendante. De même, pour une modification réaliste, l'enveloppe spectrale du son doit être préservée. Sinon, le timbre du son change. Par exemple

en parole, une telle modification du timbre entraîne la perte de l'identité des voyelles ou des consonnes, et l'identité des locuteurs.

4. Filtrage. Une application évidente de la TFCT est le filtrage, à travers l'interprétation en banc de filtres. Ce filtrage peut être dépendant du temps. Ainsi la TFCT est un outil très puissant pour en quelque sorte "sculpter les son" dans la mesure ou il permet de d'implémenter un banc de filtre variable avec une haute qualité.

5. Synthèse croisée, morphing. Il s'agit de croiser, d'hybrider, plusieurs sons, en utilisant par exemple l'enveloppe spectrale d'un son et l'excitation d'un autre. Ce type d'effet était connu sous le nom de "vocodeur", car il vient du codage de la parole (acronyme pour l'anglais "voice coder"). C'est ce type d'effet que l'on utilise pour faire parler une guitare ou pour mélanger des sources sonores.

6. Décomposition du son. En plus du filtrage linéaire (évoluant dans le temps), la TFCT permet de réaliser des décomposition assez fine du son. Par exemple la décomposition périodique-apériodique du son peut être réalisée par TFCT. Dans ce cas il s'agit de séparer la composante périodique (ou harmonique, parfois appelée "déterministe") du son de la composante apériodique (ou bruitée, parfois appelée "stochastique"). Cette décomposition possède de l'intéret pour l'analyse des sons instrumentaux, qui sont souvent un mélange de bruit (transitoires, bruit de souffle) et de sons quasi-périodiques. Par exemple, on peut isoler le bruit de choc du marteau et celui de la résonance de la corde dans un son de piano.

7. Suivi de partiels ou d'harmoniques. L'analyse-synthèse par TFCT peut servir de base à d'autres techniques plus spécifiques (paramétriques) d'analyse-synthèse du son. Par exemple, pour les sons quasi-harmoniques, on peut construire des repésentations sinusoïdale, en pistant les partiels ou les harmoniques du son. Ainsi on obtient une représentation paramétrique du son qui permet elle même de nouveaux effets, ou une meilleure qualité pour des effets classiques.

6.7. *Représentation sinusoïdale*

Pour beaucoup de son musicaux, l'excitation e est périodique. On peut donc considérer le signal comme une somme de composantes sinusoïdales. La représentation sinusoïdale généralise la décomposition du signal comme somme de segments sinusoïdaux. Elle est basée sur le modèle source filtre. Le signal s'écrit comme le filtrage d'une excitation $e(t)$ par un filtre de réponse impulsionnelle $h(t, \tau)$, évoluant dans le temps :

$$s(t) = \int_0^t h(t, t-\tau)e(\tau)d\tau \qquad [143]$$

L'excitation $e(t)$, est exprimée sous forme d'une somme de sinusoïdes, et l'action du filtre est représentée par sa fonction de transfert $H(t, \nu)$:

$$e(t) = \sum_{l=1}^{L(t)} a_l(t)\cos(\int_{t_l}^t \omega_l(\tau)d\tau + \varphi_l) \qquad \text{et} \qquad H(t,\nu) = M(t,\nu)e^{i\phi(t,\nu)} \quad [144]$$

Il faut noter que le nombre de segments sinusoïdaux de l'excitation $L(t)$ varie dans le temps, ainsi que les amplitudes a_l, et les fréquences ω_l. Les phases initiales φ_l dépendent de l'instant d'apparition de la sinusoïde t_l. le signal $s(t)$ résultant du modèle complet s'écrit :

$$s(t) = \sum_{l=1}^{L(t)} A_l(t)\cos(\Psi_l(t)), \quad \begin{cases} A_l(t) = a_l(t)M(t, \nu_l(t)) \\ \Psi_l(t) = \int_{t_l}^t \omega_l(\tau)d\tau + \varphi_l + \phi(t, \nu_l(t)) \end{cases} \qquad [145]$$

Cette analyse est basée sur la notion de piste ("track") sinusoïdale, les composantes de la sommation dans la formule de synthèse 145. Le nombre $L(t)$ de pistes varie dans le temps, et chaque piste existe donc pendant un certain temps, et doit être déterminé par un algorithme de suivi de composantes. Dans le cas des signaux discrêts, l'équation 145 se réduit à :

$$s(n) = \sum_{l=1}^{L} A_l(n)\cos(\omega_l n + \theta_l) \qquad [146]$$

Il faut donc estimer le nombre de composantes, et leurs amplitudes, phases et fréquences. La phase d'analyse est basée sur la TFCT. Soit $\tilde{x}_r(k)$ le k-ième échantillon de la transformée de Fourier discrête, pour la trame numéro r (la fenêtre d'analyse de N échantillons est non nulle entre O et $N-1$) :

$$\tilde{x}_r(k) = \sum_{n=0}^{N-1} w(n)x(n + rH))e^{-i\omega_k n} \qquad \text{pour } r = 0,\, 1,\, \ldots \qquad [147]$$

où $w(n)$ est une fenêtre d'analyse temporelle, H l'avance en nombre d'échantillons de la fenêtre d'analyse, et N la durée d'une trame. A partir de ce spectre complexe à court terme, on peut calculer les spectres d'amplitude et de phase

de $\tilde{x}_r(k)$. Pour chaque trame, les pics spectraux sont obtenus en cherchant sur le spectre d'amplitude tous les maxima locaux, et en éliminant ceux dont l'amplitude est inférieure à un seuil donné. La position des pics donne les fréquences ω_l et les amplitudes A_l des composantes sinusoïdales. Les phases θ_l de ces composantes sont calculées comme les phases de la TFCT à la fréquence ω_l. Pour chaque trame, un ensemble de L de pics spectraux est ainsi obtenu. Le signal de synthèse peut être obtenu par recouvrement-addition des signaux à court terme de l'équation 146. Dans ce cas, les pistes sinusoïdales ne sont pas explicites. Il est plus beaucoup plus économique de suivre explicitement les pistes sinusoïdales, et d'interpoler les paramètres de synthèse le long de ces pistes. Le problème est donc maintenant le suivi de ces pics dans le temps. Il s'agit d'apparier, en amplitude, phase et fréquence, les pics estimés sur une trame avec ceux estimés sur la trame suivante. En effet le nombre de pic va varier d'une trame à l'autre. Pour résoudre ce problème, un algorithme détecte la "naissance", la "continuité" et la "mort" des composantes sinusoïdales d'une trame sur l'autre. Soit q pics détectés à la trame $r+1$ avec des fréquences $g_1, g_2, \ldots, g_q$. Il faut les apparier aux p pics de fréquences $f_1, f_2, \ldots, f_p$ de la trame prédédents, la trame r. Chacune de ces pistes f_i cherche "son" pic dans la trame $r+1$, en prenant le plus proche en fréquence. Un seuil est fixé sur la différence $|g_i - f_i|$ pour que l'appariement soit possible. Il y a alors 3 possibilités :

1. un appariement est trouvé : il y a **continuité** de la piste d'une trame sur l'autre.
2. aucun appariement n'est trouvé : il y a **mort** de la piste d'une trame sur l'autre, c'est à dire que l'amplitude du pic dans la trame $r+1$ est mise à 0.
3. les pics de la trame $r+1$ qui n'ont pas trouvé d'appariement dans la trame r sont considérés comme appartenant à une nouvelle piste. Il y a **naissance** d'une piste dans la trame r, avec une amplitude 0 dans cette trame r.

Après l'algorithme d'appariement esquissé au dessus, les pics pour une trame r donnée correspondent tous à des pics dans la trame $r+1$, certain avec une amplitude nulle (devant donc naître ou mourir). La synthèse entre les instants r et $r+1$ peut donc être réalisée par la formule 146. Cependant, comme les paramètres changent à chaque tame, une procédure d'interpolation doit être mise en oeuvre pour éviter des ruptures brutales d'une trame à l'autre. Les amplitudes sont interpolées linéairement entre les valeurs A_r et A_{r+1} aux borne de la trame , soit :

$$\hat{A}^r(n) = A^r + (A^{r+1} - A^r)(\frac{n}{H}) \qquad [148]$$

Les phases et fréquences sont interpolée par une fonction cubique :

$$\hat{\theta}^r(n) = \theta^r + \omega^r n + \nu n^2 + \gamma n^3 \qquad [149]$$

où les coefficients ν et γ sont déterminés en fonction de ω^r, ω^{r+1} θ^r, θ^{r+1}, par une procédure qui ne sera pas détaillée ici. La formule de synthèse pour la trame r est alors :

$$\hat{s}^r(n) = \sum_{l=1}^{L^r} \hat{A}_l^r(n) \cos(\theta_l^r n) \qquad [150]$$

La formule de synthèse sinusoïdale peut être considérée comme une forme réduite de synthèse par TFCT, qui n'utilise que les pics spectraux d'une analyse par TFCT à bande étroite. Pour les sons périodiques, les pistes sinusoïdales correspondent aux harmoniques du signal. Pour les sons non-périodiques, la représentation sinusoïdale reste valide si on utilise beaucoup de pistes, car les composantes sinusoïdales forment une approximation du le spectre de la TFCT. La représentation sinusoïdale délivre un signal perceptivement indiscernable de l'original, et peut donc être utilisé comme méthode de codage. Des modifications de fréquence fondamentale et de durée sont également possibles en séparant sur chaque segment sinusoïdal les contributions de l'excitation et du filtre.

7. Les sons du Français

7.1. *Production de la parole*

L'appareil vocal humain est constitué d'un ensemble d'organes susceptibles de produire une grande variété de sons. La parole proprement dite n'exploite qu'un sous ensemble de cette riche palette.

On peut décomposer l'appareil phonatoire en trois sous-ensembles fonctionnels :

- les poumons et la trachée-artère ;
- le larynx ;
- le conduit vocal ;

7.1.1. *Les poumons et la trachée-artère*

La source d'énergie nécessaire pour produire de la parole réside dans les muscles abdominaux et thoraciques. Les poumons comprimés par cette musculature, agissent de façon analogue à un soufflet et fournissent une pression d'air (pression subglottique). Cette pression est transformée en son à travers le larynx et le conduit vocal. La pression d'air à la sortie des poumons est d'environ 4 cm d'eau pour la voyelle la plus douce possible, et 20 cm d'eau pour une voyelle très aigue et forte ; La trachée-artère est le conduit (quasi-cylindrique) qui part des bronches pour aboutir au larynx. La trachée-artère mesure environ 12cm de long pour un diamètre de 1,5 à 2 cm.

7.1.2. *Le larynx*

Le larynx est l'ensemble de muscles, cartilages articulés, ligaments et muqueuses compris entre la trachée-artère d'une part et la cavitée pharyngée de l'autre. C'est également l'organe qui, en conjonction avec l'épiglotte permet d'obstruer ou non la trachée-artère et l'oesophage (permettant ainsi d'éviter que le bol alimentaire ne passe dans les poumons, et l'air dans le tube digestif).

En ce qui concerne la phonation, les cartilages et ligaments du larynx permettent d'actionner une paire de muscles, les cordes vocales (qui ne sont pas des cordes mais des bandes musculaires) dont l'ouverture porte le nom de glotte. La glotte mesure environ 18 mm de long, sur une hauteur moyenne de 3 mm. Lorsque les cordes vocales sont séparées, l'air circule librement dans la glotte et aucun son n'est produit à ce niveau là. Il est également possible de rapprocher les deux membranes et de les accoler afin de produire, sous l'action de la pression d'air sub-glottique délivrée par les poumons, une impulsion isolée ou une vibration. Le processus est le suivant (théorie myo-élastique de la vibration glottique) :

1. les cordes vocales sont accolées et obstruent ainsi le larynx ;
2. sous l'action de la pression sub-glottique, les deux bandes de muscle tendent à s'amincir et à s'effacer ;
3. les cordes s'écartent alors brutalement et laissent ainsi passer un jet d'air ;
4. cette bouffée d'air entraine une dépression sous la glotte et l'effet de Bernouilli joint à l'élasticité des muscles tendent à refermer la glotte ;
5. les cordes vocales sont accolées et obstruent ainsi le larynx. . . ;

C'est la vibration des cordes vocales qui fixe la hauteur mélodique du signal vocal (ou pitch), on peut en effet suivant la contraction des muscles du larynx contrôler très finement la périodicité du processus décrit plus haut. Les cordes vocales possèdent de plus plusieurs modes de vibration différents, suivant le degré de leur accolement : de ces modes résultent les différents "registres" (registre de poitrine, registre de tête ou de fausset). D'ordinaire dans la voix parlée un même registre est systématiquement employé (généralement le registre de poitrine chez le locuteur masculin et le registre de tête ou de poitrine chez le locuteur féminin). Les cordes vocales peuvent produire des impulsions isolées et des sons périodiques sur une étendue d'environ 3 octaves. Le pitch moyen d'un locuteur masculin se situe entre 90 et 120 Hz, 150 à 300 Hz chez une locutrice et 350 à 400 Hz chez un jeune enfant.

7.1.3. *Le conduit vocal*

Le conduit vocal peut se subdiviser en trois cavités : le pharinx, la cavité orale, la cavité nasale. Le conduit vocal Le pharynx est la cavité limitée par le larynx d'une part, et le voile du palais ou velum et l'arrière de la langue, d'autre part. Le velum sert à dériver ou non l'onde acoustique en provenance du pharynx vers le conduit nasal. La cavité buccale commence au velum et s'achève par les lèvres. Elle contient la langue, ensemble complexe de muscles, et dépend de la mâchoire : c'est une cavité de géométrie variable. La cavité nasals se présente sous la forme de deux cavité fixes, les fosses nasales.

La conformation du conduit vocal dépend des organes articulateurs : machoires, lèvres, langue. Les sons émis dépendent principalement des paramètres articulatoires suivants :

- le point d'articulation, ou point ou la langue est le plus près du palais ;
- l'aperture ou section du conduit vocal au point d'articulation ;
- la labialisation ou forme des lèvres ;
- la nasalité ou dérivation de l'onde sonore vers le conduit nasal ;
- la latéralité ou passage de l'air de part et d'autre de la langue ;

Le conduit vocal est long d'environ 17 cm (chez l'homme adulte) et l'aire d'une section peut varier entre 0 (fermeture totale) et 20 cm2. La cavité nasale est longue d'environ 12 cm et possède environ 60 cc de volume.

7.1.4. *Production des sons par l'appareil vocal*

On peut distinguer trois types de sources sonores, qui peuvent se combiner ou intervenir individuellement :

1. la vibration des cordes vocales, source quasi-périodique, pouvant délivrer un signal (dit "voisé") arbitrairement long (dans les limites d'une expiration) ;
2. les bruits produits par un écoulement d'air turbulent dû à une constriction dans le conduit vocal (de même la durée n'en est limitée que par le souffle du locuteur) ;
3. les rapides occlusions du conduit vocal (lèvres, langue contre le palais ou les incisives par exemple), générant une impulsion acoustique. Ici la durée est brève.

Ces trois sources sonores élémentaires sont transformées par le conduit vocal, et se propagent à l'extérieur du système phonatoire par les ouvertures que du nez et de la bouche. Le conduit vocal assure une fonction de filtrage acoustique des signaux de source, le filtre mis en jeu évoluant avec la conformation des articulateurs. La cavité formée par le conduit vocal possède des résonances et des antirésonances, lors de l'utilisation de la dérivation nasale.

Les organes articulateurs peuvent donner naissance à des évolutions très rapides de la conformation du conduit vocal, et donc de ses propriétés acoustiques (par exemple en parole voisée, d'une période à l'autre les résonances peuvent avoir considérablement évolué). Un mouvement très rapide peut donner naissance à une impulsion acoustique.

7.2. *Phonèmes*

Chaque langue retient pour son fonctionnement un ensemble de sons, parmi ceux autorisé par l'appareil vocal. Les plus petites unités sonores distinctives utilisées pour parler sont dénomées phonèmes. Chaque énoncé parlé peut se décomposer en phonèmes : par exemple l'énoncé "onde" se décompose en /õ/, /d/. Les phonèmes sont des objets linguistiques : l'inventaire des phonèmes d'une langue s'appuit sur des critères distinctifs. Le phonème est la plus petite unité sonore qui, substituée à une autre, change le contenu linguistique d'un mot. Par exemple changer le premier son de /p/ de "peau" (/po/) en /b/ aboutit à un mot différent : "beau" (/bo/). On distingue donc les phonèmes /p/ et /b/. Par convention, les phonèmes sont notés entre des barres obliques (notation phonologique).

On utilise également une notation plus précise des différentes réalisations des phonèmes (ou allophones) : la transcription phonétique. Il faut distinguer la transcription phonétique, notation aussi précise que possible des sons (on distinguera par exemple le "r" apico-alvéolaire du "r" gutural), de la transcription phonologique, notation des unités linguistiques distinctives (un seul symbole "/r/", rendra compte des deux types de "r").

7.2.1. *Voyelles orales*

La première subdivision qui apparaŇt est liée au mode d'excitation du conduit vocal, et à la stabilité de ce dernier : c'est la séparation voyelles/consonnes. Les voyelles correspondent d'une part à une excitation quasi-périodique (donc assez longue) délivrée par les cordes vocales, et d'autre part à une conformation stable du conduit vocal, du moins lorsque la voyelle est prononcée isolement. Les voyelles sont des segments relativement stables et d'énergie assez élevée. On distingue les différentes voyelles grâce à l'emplacement des premiers formants. Les deux premiers formants permettent déjà de classer le système phonétique du français. On établit par un schéma classique la répartition des voyelles dans le plan F1, F2 (premier formant, deuxième formant) qui fait apparaitre le "triangle vocalique" (limites de F1 et F2 en fréquence pour les voyelles).

7.2.2. *Voyelles nasales*

Suivant que la dérivation nasale est ouverte ou non (grâce à l'abaissement du velum), les voyelles seront nasales ou seront orales. Les voyelles nasales se dinstinguent acoustiquement des voyelles orales par les anti-resonances introduites par les fosses nasales.

7.2.3. *Semi-voyelles*

Lorsque l'excitation glottique coexiste avec une évolution rapide du conduit vocal, on assiste à la naissance de semi-voyelle. Suivant la vitesse d'élocution les semi-voyelles seront perçues comme une entité indépendante ou comme une suite de phonèmes distincts.

7.2.4. *Fricatives*

L'excitation est un bruit de friction, le conduit vocal est ainsi perturbé en amont et en aval de la constriction, en première aproximation seule la partie aval

classe	phonème	exemple
voyelles orales	ɑ	pâte
	a	patte
	o	pot
	ɔ	pomme
	e	nez
	ɛ	mer
	ø	peu
	œ	peur
	ə	le (schwa)
	i	pis
	y	pied
	u	pou
voyelles nasales	õ	pont
	$\tilde{\epsilon}$	pain
	œ̃	brun
	$\tilde{a}$	pente
plosives voisées	b	bout
	d	doux
	g	goût
plosives sourdes	p	poux
	t	toux
	k	cou
fricatives voisées	v	voile
	ʒ	jeux
	z	zéro
fricatives sourdes	f	faon
	s	sang
	ʃ	chant
liquides	ʀ	rang
	l	lent
occlusives nasales	ɲ	agneau
	m	mot
	n	noix
	ŋ	camping
semi-voyelles	j	pied
	w	loi
	ɥ	lui

Tableau 3. *Phonèmes du français*

contribue à la "coloration" du signal de source. Les divers lieux de constriction (lieux d'articulation) déterminent les diverses fricatives : fricatives dentales (articulées entre les incisives supérieures et la lèvre inférieure), alvéolaires (aux alvéoles derrière les incisives) et post-alvéolaires (en arrière des alvéoles) ; Par ailleurs, les cordes vocales peuvent entrer en vibration en même temps que le bruit de friction, la fricative étant alors voisée, ou laisser passer l'air sans émettre de son, la fricative étant alors sourde. Les fricatives sont des phonèmes en général assez énergétiques (parfois autant que les voyelles) ;

7.2.5. *Plosives*

Elles résultent du relâchement d'une constriction dans le conduit vocal (ou explosion, d'ou le terme de plosives). Joint à une vibration des cordes vocales, on obtient une occlusive (ou plosive) voisée, sinon on obtient une occlusive sourde. Les lieux d'articulation peuvent être les lèvres, les incisives supérieures (ou les alvéoles) ou le velum. Un silence précède bien sur l'explosion pendant le temps d'occlusion, et (comme pour les fricatives) un petit temps est nécessaire avant le premier battement des cordes vocales, si la consonne est suivie par une voyelle (VOT : voice onset time) ;

7.2.6. *Nasales*

Elles sont assez proches des plosives voisées, mais l'abaissement du voile du palais entraine l'apparition d'anti-résonances dues à la dérivation nasale. D'autre part l'occlusion de la cavité buccale étant complète, le son est alors rayonné aux narines et l'on ne peut pas constater de silence dans ces phonèmes, seulement une baisse générale d'énergie par rapport aux voyelles ;

7.2.7. *Liquides*

Assez semblables aux semi-voyelles, elle résultent d'une excitation voisée et de rapides mouvements articulatoires, principalement de la langue. Le /R/ est assez souvent présent sous forme de deux variantes (ou allophones) suivant la façon dont il est articulé : par l'apex ou pointe de la langue, (/R/ apical) ou par la luette (/scr/ uvulaire). Ici encore on peut constater une baisse notable d'énergie par rapport au voyelles ;

Dans la parole continue, l'effet de coarticulation dû au mouvement ininterrompu des muscles de l'appareil vocal, entraine une perturbation mutuelle

des phonèmes voisins. En effet un phonème, ou groupe de phonèmes est considérablement affecté par son contexte d'apparition, en fonction de la vitesse d'élocution, des phonèmes qui suivent ou précèdent et de la prosodie.

7.3. *Traits distinctifs*

Les phonèmes peuvent être considérés comme des faisceaux de traits distinctifs. Ces traits permettent de mettre en évidence des classes de phonèmes qui partagent certaines propriétés acoustiques et articulatoires.

Un ensemble de 14 traits distinctifs permet de caractériser les phonèmes du Français :

a Général

1 ± vocalique (voc)
2 ± consonantique (cons)

b Mode d'articulation

3 ± voisé (voi)
4 ± continu (cont)
5 ± strident (str)

c Lieu d'articulation

6 ± coronal (cor)
7 ± antérieur (ant)
8 ± arrondi (arr)
9 ± nasal (nas)
10 ± latéral (lat)
11 ± ouvert (ouv)
12 ± fermé (fer)
13 ± avant (av)
14 ± arrière (ar)

Le premier trait général est l'opposition vocalique/consonantique : un son est vocalique lorsque les cordes vocales vibrent, et lorsque le conduit vocal est assez libre d'obstruction. Lorsqu'une obstruction se produit dans le conduit vocal, le son est consonantique. Seules les semi-voyelles ne sont ni vocaliques ni consonantiques. Les autres traits sont liés à l'excitation (mode d'articulation) ou au lieu d'articulation.

7.3.1. *Modes d'articulation*

Le premier trait est le trait de voisement (vibration des cordes vocales). Le trait continu est affecté aux phonèmes qui provoquent pas une fermeture totale du conduit vocal (plosives). Le trait strident se rapporte aux fricatives (génération d'un bruit d'écoulement dû à une turbulence).

7.3.2. *Lieux d'articulation*

La région articulatoire définit les traits relatifs au lieu d'articulation. Le trait coronal s'applique aux phonèmes utilisant la pointe de la langue (consonnes). Le trait antérieur indique que l'obstruction totale ou partielle du conduit vocal se produit vers l'avant de la bouche, pour les consonnes : c'est le cas des labiales, des dentales, des alvéolaires. Le trait latéral indique un passage de l'air sur les côtés de la langue, il est réservé au /l/ en Français. Le trait nasal indique le passage de l'air dans les cavités nasales, par abaissement du voile du palais. Le trait arrondi est spécifiquement labial, il est présent lorsque les lèvres s'arrondissent (protrusion des lèvres). En fonction de la position de la langue, on distingue les traits avant et arrière. Les traits ouvert et fermé indiquent la plus ou moins grande aperture du conduit vocal (par abaissement de la mâchoire et/ou de la langue.

7.4. ***Prosodie***

En quelque sorte superposée au flot de phonèmes, la prosodie articule le discours en distribuant accentuation et intonation. L'intonation est la mélodie, variation de hauteur mélodique dans le temps. Elle a une fonction distinctive (l'intonation distingue un énoncé interrogatif, affirmatif etc), syntaxique, expressive. L'accent met en valeur une entité suprasegmentale (syllabe, groupe de syllabes...) et possède donc une fonction contrastive, syntaxique et expressive. La distinction entre accentuation et intonation est plutôt d'ordre fonctionnel que d'ordre acoustique, les paramètres acoustiques en jeu dans les deux cas sont :

- des variations de la fréquence de vibration des cordes vocales ;
- des variations d'intensité ;
- des variations de durées phonémiques ce qui donne le rythme d'élocution ;

	voc	cons	voi	cont	str	cor	ant	arr	nas	lat	ouv	fer	av	ar
a	+	-	+	+	-	-	-	-	-	-	+	-	-	-
o	+	-	+	+	-	-	-	+	-	-	-	-	-	+
ɔ	+	-	+	+	-	-	-	+	-	-	+	-	-	+
e	+	-	+	+	-	-	-	-	-	-	-	-	+	-
ɛ	+	-	+	+	-	-	-	-	-	-	+	-	+	-
ø	+	-	+	+	-	-	-	-	-	-	-	-	+	-
œ	+	-	+	+	-	-	-	-	-	-	+	-	-	-
i	+	-	+	+	-	-	-	-	-	-	-	+	+	-
y	+	-	+	+	-	-	-	-	-	-	-	+	+	-
u	+	-	+	+	-	-	-	-	-	-	-	+	-	+
õ	+	-	+	+	-	-	-	+	+	-	+	-	-	+
œ̃	+	-	+	+	-	-	-	-	+	-	+	-	+	-
ã	+	-	+	+	-	-	-	-	+	-	+	-	-	-
b	-	+	+	-	-	-	+	-	-	-	-	-	-	-
d	-	+	+	-	-	+	+	-	-	-	-	-	-	-
g	-	+	+	-	-	-	-	+	-	-	-	+	-	+
p	-	+	-	-	-	-	+	-	-	-	-	-	-	-
t	-	+	-	-	-	+	+	-	-	-	-	-	-	-
k	-	+	-	-	-	-	-	+	-	-	-	+	-	+
v	-	+	+	+	-	-	+	-	-	-	-	-	-	-
ʒ	-	+	+	+	+	+	-	-	-	-	-	+	-	-
z	-	+	+	+	+	+	+	-	-	-	-	-	-	-
f	-	+	-	+	+	-	+	-	-	-	-	-	-	-
s	-	+	-	+	+	+	+	-	-	-	-	-	-	-
ʃ	-	+	-	+	+	+	-	-	-	-	-	+	-	-
ʀ	-	+	+	+	-	-	-	+	-	-	-	+	-	+
l	-	+	+	+	-	+	+	-	-	+	-	+	-	-
ɲ	-	+	+	+	-	-	-	-	+	-	-	+	-	-
m	-	+	+	+	-	-	+	-	+	-	-	-	-	-
n	-	+	+	+	-	+	+	-	+	-	-	-	-	-
j	-	-	+	+	-	-	-	-	-	-	-	+	+	-
w	-	-	+	+	-	-	-	+	-	-	-	+	+	+
ɥ	-	-	+	+	-	-	-	+	-	-	-	+	-	-

Tableau 4. *Traits distinctifs des phonèmes du français*

Une micro-mélodie est générée par l'enchainement des phonèmes, indépendament de toute action volontaire par le simple jeu de la réaction du conduit vocal sur la source laryngée dont la périodicité est ainsi légèrement modifiée.

8. Lecture de spectrogrammes

8.1. *Indices acoustiques*

Le spectrographe en bande large est généralement préféré en phonétique acoustique, car il met en valeur les formants. Des exemples d'analyse spectrographique en bande large pour les phonèmes du français sont donnés plus loin. On remarquera la bonne représentation des formants, qui permet de localiser et d'identifier les phonèmes. La lecture de spectrogramme doit délivrer deux types d'informations :

1. quel est le nombre de phonèmes, et où sont-ils ? (phase de segmentation).
2. quels sont les phonèmes ? (phase d'identification).

Ces deux phases utilisent des indices acoustique, formes-types que l'on peut lire sur les spectrogrammes. La nomenclature de ces formes-types peut-être très raffinée (si elle détaille les variations dues au contexte) et augmente avec l'expertise du lecteur. Les indices sont relatifs, dans une certaine mesure, au locuteur : un processus de normalisation prend place pour rendre compte du système vocalique particulier du locuteur, de ces habitudes articulatoires et prosodiques. Les niveaux supérieurs (système phonologique de la langue donnée, lexique, voire syntaxe) permettent souvent d'inférer des hypothèses sur la chaîne phonétique au cours de sa découverte. Voici quelques indices de base pour le français. C'est le mouvement et la position des formants (maxima du spectre, visibles sur les spectrogrammes comme les zones fréquencielles possédant le plus d'énergie) qui va permettre d'identifier les sons de la parole. Les quelques règles générales pour la relation entre articulation et formants sont :

1. un F1 bas signale une occlusion ;
2. un F2 bas signale une articulation labiale ;
3. un F2 très haut signale une articulation palatale ;
4. un F3 haut signale une articulation dentale ou alvéolaire ;
5. un F3 très bas signale une articulation rétroflexe ;

6. l'axe F1 est assimilable à l'axe haut/bas de la machoire (F1 bas=fermé, F1 haut=ouvert) ;

7. l'axe F2 est assimilable à l'axe avant/arrière de la langue (F2 haut=avant, F2 bas= arrière) ;

8.2. *Voyelles orales*

Les cordes vocales vibrent, les formants sont bien visibles, l'énergie du signal importante. Des mouvements formantiques dues à la coarticulation sont souvent apparents, mais les voyelles sont des segments assez stables. Elles peuvent ne durer que quelques battements des cordes vocales. Une fois la voyelle repérée, elle est identifiée en utilisant le triangle vocalique (qui doit être éventuellement adapté au locuteur particulier que l'on lit).

Le triangle vocalique est une représentation des voyelles en fonction de leurs deux premiers formants. La localisation de chaque voyelles dans le plan F1/F2 permet de discriminer les différentes voyelles. Les voyelles /a/, /i/ et /u/ forment les pointes du triangle vocalique, ce sont les voyelles "cardinales", ou extrèmes. Les axes F1/F2 peuvent recevoir une interprétation articulatoire simple et relativement robuste, de la manière suivante :

1. F1 représente la dimension haut/bas de la machoire. /a/ est la voyelle la plus ouverte (pour laquelle la machoire est la plus basse). Son F1 est haut. Au contraire /i/ et /u/ sont très fermées (la machoire est haute), leur F ! sont bas.

2. F2 représente la dimension antérieur/postérieur. "Antérieur" signifie que la langue est massée à l'avant de la bouche, et "postérieur" que la langue est massée à l'arrière. /i/ est la voyelle la plus antérieure (avec le F2 le plus haut) et /u/ la voyelle la plus postérieure (avec le F2 le plus bas).

8.3. *Voyelles nasales*

Seulement 3 ou 4 en français. Elles possèdent les mêmes caractères que les voyelles orales, mais 1 ou plusieurs anti-formants (c'est à dire creux marqués dans le spectre) apparaissent. En particulier dans la région du premier formant qui est souvent dédoublé par un anti-formant (on voit alors un formant nasal). Un autre indice important est leur durée, qui est beaucoup plus importante que celle des voyelles orales.

8.4. *Fricatives sourdes*

Le signal est un bruit (non-voisé), l'énergie est assez importante. Les fricatives sourdes sont assez faciles à localiser. Ce sont des segments assez stables, comme les voyelles, et souvent assez longs. Les trois fricatives sourdes françaises se distinguent par la force et l'emplacement du bruit. / ∫ / est la fricative la plus forte, avec un bruit assez grave (descendant vers 1.5 ou 2 kHz). /s/ est de force moyenne, souvent plus aigue. /f/ est la fricative sourde la plus faible, que l'on rencontre soit avec un bruit assez dispersé et plus grave que /s/, soit avec un bruit très aigu. Le lieu d'articulation de / ∫ / est post-alvéolaire, voire pré-palatal. Les formants des voyelles adjacentes illustrent ce mouvement : F1 monte, F2 descend et soit forme pince avec F3 qui monte, soit possède un mouvement parallèle a F3 qui descend. Pour /s/, dont l'articulation est apico-alvéolaire, F1 monte et F2 descend faiblement, ou ne bouge pas. Pour le /f/, labio-dentale, F1 et F2 montent.

8.5. *Fricatives voisées*

Elle ressemblent au fricatives sourdes correspondantes, mais avec en plus une vibration des cordes vocales, bien visible dans le grave du spectre (barre horizontale, dite barre de voisement). Comme les fricatives sourdes, les mouvements formantiques typiques des points d'articulation sont visibles dans les voyelles adjacentes.

8.6. *Plosives sourdes*

Les plosives sourdes sont repérables par un silence court dans le signal (occlusion ou tenue de la plosive), suivi d'une barre verticale ou barre d'explosion (relachement de la plosive), d'un court délai d'établissement du voisement (pendant lequel un bruit d'aspiration est parfois visible) et enfin du mouvement typique des formants en fonction du lieu d'articulation. Le /p/, plosive sourde labiale, est caractérisé par le mouvement montant de F1 et F2, ainsi que par une explosion assez basse fréquence (visible souvent sur le signal). Le /t/, plosive sourde apico-dentale, ou apico-alvéolaire, possède un F1 montant et un F2 stable ou légèrement descendant. Souvent beaucoup de bruit d'aspiration pendant le délai d'établissement du voisement. Le /k/, plosive sourde palatale possède le mouvement formantique typique de ce lieu d'articulation : F1 monte lentement, F2 descend et forme pince avec F3 qui monte. Une double explosion est souvent visible, sur le signal et/ou sur la barre d'explosion.

8.7. *Plosives voisées*

Les mouvements formantiques des plosives voisées correspondent à ceux des plosives sourdes ayant le même lieu d'articulation. Une barre de voisement est visible pendant la tenue, sur le spectrogramme et sur le signal. Par rapport aux plosives sourdes, la durée de tenue est plus courte pour les plosives voisées. Le délai d'établissement du voisement est aussi nettement plus court, ce qui semble un indice important perceptivement.

8.8. *Liquides*

Le /r/ possède au moins deux variantes (ou allophones) : l'allophone apico-alvéolaire ("r" roulé) et l'allophone uvulaire. On le remarque par une brusque diminution de l'énergie, mais sans tenue contrairement aux plosives. Le /r/ uvulaire est souvent dévoisé. Les formants varient peu au voisinage d'un /r/. Le /l/ est caractérisé par une brusque diminution d'énergie. Le voisement est continu, mais l'intensité des formants s'affaiblit. F1 descend et F2 monte légèrement pendant le /l/.

8.9. *Nasales*

Les nasales sont caractérisées par une baisse importante d'énergie, mais un voisement continu. Pendant la nasale, F1 est bas, avec éventuellement un dédoublement à cause d'un anti-formant. F2 est plus bas pour /m/ que pour /n/, avec éventuellement un dédoublement également. Les mouvements formantiques sont très brusques au passage entre le murmure nasal et les voyelles adjacentes : il y a comme une coupure brutale des formants due à l'occlusion. Le / ɲ/ est reconnaissable par le fort mouvement descendant de F2 et le fort mouvement montant de F1. Les mouvements formantiques des voyelles adjacentes reflètent le lieu d'articulation de chaque nasale. Articulation bilabiale pour /m/ avec F1 et F2 qui montent, alvéolaire pour /n/, avec F1 montant et F2 stable ou légèrement descendant, palatale ou uvulaire pour / ɲ/, avec montée lente de F1 et descente importante de F2.

8.10. *Semi-voyelles*

Ce sont des articulation très rapides entre voyelles. Les formants peuvent évoluer d'une centaine de Hertz d'un battement vocalique à l'autre. On distingue

donc les semi-voyelles par les mouvements obliques rapides de F1 et F2 dont les cibles sont données par le triangle vocalique en fonction des voyelles sous-jacentes aux semi-voyelles.

9. Les sons instrumentaux

Nous ne tenterons pas de définir la notion de "son musical" en toute généralité, dans le mesure ou la notion même de musique est si difficile à cerner. Les courants musicaux comtemporains ont montré que tout son peut être intégré dans une structure musicale pour devenir un son musical. Comme il serait trop ambitieux d'étudier tous les sons, on se restreindra aux sons "musicaux", au sens de son produit par un instrument musical de l'orchestre classique occidental.

Cet orchestre ne représente bien entendu pas tous les instruments, ou tous les modes d'utilisation des instruments de musique inventés par l'homme. Les sons instrumentaux peuvent être classés en fonction de l'instrument qui les a produits. Cette classification intuitive masque en fait qu'un même système mécanique peut produire en réalité des sons extrèmement variés : il suffit de penser au cas de la voix ou de l'orgue.

9.1. *Les familles d'instruments*

Dans la théorie musicale, la notion d'instrumentation et d'orchestration est venue tard. Le premier traité sur cette question date en effet du millieu du XIXème siècle : c'est le "traité d'instrumentation et d'orchestration" d'Hector Berlioz (1855). Le but de ce traité est double. Dans la partie instrumentation, il s'agit de présenter les différents instruments à la disposition du compositeur, leur caractère particulier, leur étendue mélodique et dynamique, ainsi que les spécificités de leur jeu (bonnes et mauvaises notes, facilité d'émission dans les différents registres, facilité d'émission des trilles etc.). Pour ce qui est de l'orchestration, l'auteur décrit les formations orchestrales les plus efficaces, et donne de nombreux exemples de dispositions instrumentales tirées des oeuvres classique et contemporaines (contemporaines de Berlioz, bien sur). Le traité ne manque pas d'ouverture sur la notion d'instrument, car dès la seconde page, on trouve : "Tout corps sonore mis en oeuvre par le Compositeur est un instrument de musique". Suit la description des moyens dont il dispose *pour le moment*. Le traité de Berlioz est toujours édité et utilisé par les étudiants en musique, ainsi que d'autre traités plus récents, comme celui de Nicolas Rimski-Korsakoff (Principes de l'orchestration (1891)), ou de Casela et Mortari (1948). Depuis

cette époque (1948) les possibilités d'instrumentation et d'orchestration on littéralement explosé, d'une part à cause des moyens électriques, mais aussi à cause des moyens de transport et de communication qui ont permis la diffusion (et même parfois la renaissance) des instruments, des techniques et des musiques anciennes et extra-occidentales.

La classification acoustique des instruments classiques occidentaux, correspond dans ses grandes lignes à la classification des traités d'orchestration. Dans cette classification, c'est le mode de production du son (partie "excitation" de l'instrument) qui prime. On distinguera donc :

1. les instruments à corde
   (a) cordes frottées : violon, alto, violoncelle, contrebasse, viole de gambe, ...
   (b) cordes pincées : clavecin, guitare, harpe, cythare viennoise, ...
   (c) cordes frappée : piano, cymbalum, clavicorde, ...
   (d) cordes sympathiques : tempura, harpe éolienne, ...
2. les instruments à vent
   (a) embouchure de flûte : flûtes traversières, flûtes à bec, flûtes à encoche, flûtes obliques, flûtes globulaires, ...
   (b) anches simples : clarinettes, saxophones,
   (c) anches doubles : hautbois, cor anglais, basson, sarrusophone,
   (d) anches lippales : trompettes, cors, trombones, serpents, ophicléïdes, tubas,
   (e) anches libres : accordéons, harmonicas,
   (f) orgue
   (g) voix
3. les instruments à percussion
   (a) à hauteur fixe : timbales, xylophone, vibraphone, jeu de cloches, cloches, wood blocks,
   (b) sans hauteur fixe : tambours, cymbales,
4. les voix humaines

9.2. *Propriétés temporelles des sons instrumentaux*

Si on considère un son instrumental dans la dimension amplitude/temps, l'oscillogramme, on observe des oscillations rapides et une forme globale du son décrite par son enveloppe temporelle. Sur cette enveloppe temporelle, on distingue 3 phases temporelles plus ou moins distinctes :

1. la phase d'attaque, ou transitoire d'attaque du son. A l'établissement du son quelques millisecondes de signal séparent le silence initial et la partie stable de l'enveloppe temporelle du son. Suivant l'instrument, cette attaque peut-être un transitoire bruité, un bruit de souffle, un bruit de choc, le bruit de bouche d'un tuyau, le raclement de la corde d'un violon etc.

2. la partie stationaire, ou tenue du son. Dans cette partie, le son évolue relativement peu (ou relativement moins). Dans beaucoup d'instruments, la tenue est assez brève, mais pour certains instruments, comme l'orgue la cornemuse ou la vielle à roue, ou certaines techniques, comme le souffle continu, la tenue peut durer très longtemps.

3. l'extinction ou, ou amortissement, ou transitoire de fin du son. Pour certains sons, comme les percussions relativement amorties, la phase d'extinction du son s'enchaîne directement à la phase d'attaque du son. Ainsi l'extinction peut être très longue. Pour d'autres sons, le transitoire de fin est très bref, comme le bruit d'étouffoir qui stoppe la vibration de la corde du piano ou du clavecin.

Cette décomposition en phases temporelle d'un son est relativement naturelle dans beaucoups de cas, (par exemple des instruments comme l'orgue, les vents, les cordes frottées etc.). Pour les instruments percussifs, la phase de tenue est souvent difficile à localiser, car le son décroît dès la fin de l'attaque.

L'enveloppe temporelle du son est en tout cas un paramètre extrêmement important : on peut transformer en percussion à peu près n'importe quel son en lui appliquant l'enveloppe d'une percussion. De même le transitoire d'attaque est une des caractéristiques les plus pertinentes du timbre d'un instrument.

Dans la technique de synthèse dite "par échantillonnage" des instruments, l'enveloppe est un paramètre particulièrement important. Pour ce type de synthèse, des sons naturels sont enregistrés, et découpés en phases temporelles. En général 3 (attaque, tenue, extinction) ou 4 phases (attaque, chute, tenue, extinction) sont distinguées. Pour synthétiser des sons de durées variables, on effectue un bouclage sur la phase de tenue : il est alors possible de produire des sons qui ont un timbre relativement naturel, et une durée spécifiée, à partir de

peu d'échantillons. Des instruments comme le piano ou l'orgue se prêtent particulièrement bien à la synthèse par échantillonnage, qui donne actuellement la meilleure qualité de synthèse pour le coût (mesuré en terme de calculs) le plus bas.

9.3. *Sons périodiques et apériodiques*

Les sons musicaux contiennent en général une composante périodique, que l'on peut analyser comme somme d'harmoniques, et une composante apériodique. La composante apériodique est faite de :

1. bruits transitoires
2. bruits continus
3. sons à partiels inharmoniques, ou multiphoniques
4. perturbations de la périodicité, ou apériodicité structurelle

La composante apériodique se rencontre soit pour le son tout entier, soit dans une de ses phases temporelles. Nous allons considérer ces différentes formes d'apériodicité, qui apparaissent tantôt comme des ornements mélodiques, tantôt comme des signatures du son, et tantôt comme des défauts.

Les bruits transitoires sont souvent caractéristique du son particulier d'un instrument. Dans les flûte par exemple, un bruit "de bouche" est souvent associé au premières millisecondes du son. Un transitoire d'attaque est également présent dans Les instruments à cordes frottées, dans les cordes pincées (clavecin, harpe) ou frappées (clavicorde, piano).

Certains instruments possèdent un bruit de souffle continu qui s'ajoute au son périodique. C'est le cas de certaines flûtes (comme le shakuhachi japonais ou la flûte de pan), ou des cordes frottées.

L'apériodicité d'une son n'est pas forcément liée à son caractère bruité ou aléatoire. Beaucoup d'instruments peuvent produire des sons apériodiques parfaitement déterministes. C'est par exemple le cas, de sommes de composantes sinusoïdales qui ne sont pas dans un rapport harmonique (c'est à dire qui ne sont pas multiples d'une même fréquence fondamentale). Parmis les sons inharmoniques, ont trouve par exemple les cloches, qui "s'accordent" en réglant les fréquences des différents partiels sur des consonances, les sons multiphoniques d'instruments à vent, les sons graves du piano, ou d'autres instruments à cordes très raides.

Une source d'apériodicité provient aussi des variations, déterministes ou non, de la périodicité d'un signal harmonique. En adoptant une terminologie qui vient du traitement de la voix, on appelera ces apériodicités des apériodicités structurelles. Comme exemple de variation périodique de la périodicité (qui introduit une apériodicité), on trouve le vibrato vocal ou instrumental. Les glissandos typiques de certains instruments (guitare avec bottle-neck par exemple) sont d'autres apériodicité déterministes. Des appériodicités structurelles aléatoires sont aussi présentes dans certains sont instrumentaux : jitter (perturbation aléatoire de la fréquence fondamentale) ou shimmer (perturbation aléatoire de l'amplitude des périodes).

Enfin, il faut noter que le contenu spectral du son évolue généralement dans le temps. Ainsi, un son de flûte par exemple, comprendra d'abord un transitoire de type bruit, puis une partie assez stationnaire, somme d'harmonique et de bruit de souffle, éventuellement une apériodicité structurelle comme le vibrato, et enfin un transitoire de fin.

9.4. *Enveloppe spectrale : le modèle source/filtre*

La production de beaucoup de sons instrumentaux, tout comme la production de la voix, montre des sources sonores relativement localisées, qui excitent des corps résonants. Par exemple on peut citer l'anche d'une clarinette et le corps de l'instrument, la corde du violon et la caisse etc. Cette relative indépendance de la source d'érnergie sonore et de sa transformation est à la base de la théorie acoustique de la voix, et de la modélisation de beaucoup de sons instrumentaux. On considère des termes de source, souvent non-linéaires, et un filtre linéaire qui transforme le signal de source.

Les composantes acoustiques de source et de filtre peuvent être étudiées séparément, car on peut en première approximation les considérer comme découplées. Du point de vue physique ce modèle est une approximation, dont l'avantage principal est la simplicité. Pour le traitement du signal, ce modèle acoustique peut être décrit en terme de systèmes linéaires, en négligeant l'intéraction source-filtre :

$$s(t) = e(t) * v(t) = [p(t) + r(t)] * v(t) \quad [151]$$

$$= [\sum_{i=-\infty}^{+\infty} \delta(t - iT0) * u(t) + r(t)] * v(t) \quad [152]$$

$$S(\omega) = E(\omega) \times V(\omega) = [P(\omega) + R(\omega)] \times V(\omega) \quad [153]$$

$$= [(\sum_{i=-\infty}^{+\infty} \delta(\omega - iF0)) \mid U(\omega) \mid e^{j\theta_{ug}(\omega)} + \mid R(\omega) \mid e^{j\theta_r(\omega)}] \times \mid V(\omega) \mid e^{j\theta_v(\omega)} \quad [154]$$

où $s(t)$ est le signal sonore, $v(t)$ est la réponse impulsionelle du résonateur, $e(t)$ est la source d'excitation, $p(t)$ est la partie périodique de l'excitation, $r(t)$ est la partie apériodique de l'excitation, $u(t)$ est l'onde d'excitation, $t0$ est la période fondamentle, $r(t)$ est la conposante bruitée de l'excitation, δ est la distribution de Dirac, et où $S(\omega)$, $V(\omega)$, $E(\omega)$, $P(\omega)$, $R(\omega)$, $U(\omega)$, sont les transformées de Fourier de $v(t)$, $e(t)$, $p(t)$, $r(t)$, $u_g(t)$, respectivement, et où $F0 = 1/T0$ est la fréquence fondamentale.

La composante de source $e(t), E(\omega)$ est un signal composé d'une partie périodique (vibration d'un excitateur caractérisée par F0 et la forme d'onde d'excitation) et d'une partie bruitée. Les différents sons ou phases de sons utilisent séparement ou conjointement les deux types de sources d'excitation.

Le résonateur est souvent une cavité acoustique (instruments à vents), ou un dispositif mécanico-acoustique (caisse de résonance des instruments à corde). Dans le modèle source-filtre, il assume le rôle du filtre, système passif et indépendant de la source. Sa fonction est de transformer le signal de source, par des phénomènes de résonance et d'anti-résonance. Les maxima du gain spectral du conduit vocal sont dénommés "formants spectraux", ou simplement “formants”. Les formants sont souvent assimilables au maxima spectraux visible sur les spectres acoustiques du son, puisqu'en général le spectre de source est globalement monotone (décroissant). Cependant, en fonction du spectre de la source, formants du spectre et résonances acoustiques du résonateur peuvent être décalés. De plus, dans certain cas, un formant de source est également présent. Des formants peuvent également être présents dans les sons ou phases de sons apériodiques, lorsqu'un résonateur est excité. Les spectres “à formants” sont typiques de certaines classes de son, ou d'instruments, comme :

la voix ici l'excitation est due au cordes vocale et le conduit vocal est le résonateur.

les cordes frottées ici l'excitation est due à la corde, et c'est la caisse de l'instrument qui forme le spectre en ajoutant ses propres résonances.

les anches ici l'excitation est due à l'anche couplée au corps de l'instrument, et le corps de l'instrument agit comme un tuyau qui forme le spectre en ajoutant ses propres résonances.

les percussions on peut considérer le sons produit comme la réponse impultionnelle d'un filtre composé. Ici l'excitation est une impulsion, et le filtre est le corps résonant.

Le modèle source filtre a été particulièrement développé pour les sons vocaux, mais s'applique également à certains sons instrumentaux. Pour la synthèse sonore, il conduit aux techniques de synthèse soustractive. Un filtre (évoluant dans le temps) va transformer un son de source riche : par exemple un peigne périodique d'impulsion, ou bien un bruit blanc, ou des impulsions isolées. Le spectre de toutes ces sources possède une enveloppe spectrale plate. Ainsi le signal résultant aura une enveloppe spectrale fixée par le filtre. D'où le nom de synthèse par filtrage, ou synthèse soustractive.