Notes de cours ISEN
Synthèse sonore

S. Tassart
Équipe Analyse-Synthèse
IRCAM, Institut de Recherche et de Coordination Acoustique–Musique
1 place Igor-Stravinsky, 75004 PARIS, FRANCE
tassart@ircam.fr

4 mars 1998

Introduction

La synthèse numérique de sons consiste essentiellement à synthétiser des échantillons à partir d'un *modèle de synthèse*, puis à convertir le signal numérique en signal analogique par l'intermédiaire d'un convertisseur numérique-analogique (DAC).

Les contraintes sont de différents ordres. D'abord, le modèle de synthèse utilisé doit être en adéquation avec la classe de signaux à synthétiser ainsi qu'avec le type de contrôle qui permet de commander la synthèse. Par exemple, les modèles dérivés des modèles source-filtre sont particulièrement indiqués en synthèse de parole ou bien, dans un domaine musical, pour la synthèse de sons évocant des chœurs orchestraux. Les modèles inspirés de celui de Karplus-Strong semblent plus adaptés pour la synthèse de sons impulsifs, tels que les instruments de percussion et quelques effets tels que les tables de réverbération numériques.

L'utilisation de la synthèse sonore par l'intermédiaire d'un instrument de musique, le synthétiseur, requiert en outre que la synthèse des sons ainsi que le contrôle se fasse en temps réel. Tous les modèles de synthèse ne sont pas forcément adaptés à la synthèse et au contrôle en temps réel. L'algorithme de synthèse doit être tout à la fois simple, efficace et rapide.

1 Synthèse par modèle de signaux

Le signal sonore à analyser, reproduire ou synthétiser consiste en un signal de pression acoustique enregistré à l'aide d'un ou plusieurs microphones et restitué par des haut-parleurs. La technologie actuelle dépend de plus en plus des techniques numériques d'enregistrement, d'archivage, de traitement, de diffusion et de synthèse. Les formats utilisés dans l'industrie audio-musicale consistent actuellement à échantillonner le signal analogique à 32kHz, 44.1kHz ou 48kHz et à quantifier le signal obtenu linéairement sur 16 bits. Cela revient à ne prendre en compte que des signaux audio dont la bande passante est limitée par la demi-fréquence d'échantillonnage (théorème de Shannon), et dont la dynamique est contenue dans 92dB. Aussi développerons nous uniquement l'aspect de modéles à temps discret.

Il est difficile d'extraire des caractéristiques communes à tous les types de signaux audio et musicaux. En conséquence, les modèles de représentation des signaux sont forcément vagues et généraux. Des modèles plus précis et plus rapides sont disponibles dès lors que la classe des signaux à analyser ou à synthétiser est spécifiée plus précisément (sons de cloches, de parole, d'intruments à vents ...).

1.1 Rappel sur les transformées

L'idée qui sous-tend toutes transformées consiste à décrire un signal sous un angle de vu différent afin de mettre en évidence des caractéristiques du signal autrement peu visibles. Les transformées de Fourier, par exemple, permettent de mettre en évidence les caractéristiques harmoniques en décomposant le signal sur une base de sinusoïdes. Cette section est largement inspirée de [Del91] en ce qui concerne les notations et les définitions.

1.1.1 Notations et convolutions

Dans tout ce qui suit, $(x_n)_{n\in\mathbb{Z}}$ et $(y_n)_{n\in\mathbb{Z}}$ désignent des signaux à temps discret, $X(f)$ et $Y(f)$ désignent leur Transformée de Fourier à Temps Discret, f désigne la fréquence normalisée sur l'intervalle $]-1/2, 1/2]$. Les opérations de convolution sont notées $*$ et $\circledast$ et sont respectivement définies dans le cas discret, dans le cas continu, et dans le cas vectoriel[1] par :

$$\forall n \in \mathbb{Z}, x_n * y_n = \sum_{k=-\infty}^{+\infty} x_k y_{n-k} \tag{1}$$

$$\forall f \in]-\frac{1}{2}, \frac{1}{2}], X(f) * Y(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} X(\nu) Y(f-\nu)\, d\nu \tag{2}$$

$$\forall n \in [0, \ldots N-1], x_n \circledast y_n = \sum_{k=0}^{N-1} x_k y_{n-k[N]} \tag{3}$$

1.1.2 Transformée de Fourier à Temps Continu (TFTC)

La *Transformée de Fourier à Temps Discret* est une transformation bijective qui associe à un signal à temps continu[2] une fonction[3] de la fréquence.

$$x(t) \longrightarrow X(f) = \int_{-\infty}^{+\infty} x(t) e^{-2j\pi t f}\, dt \tag{4}$$

Les propriétés de la transformée de Fourier sont résumés dans le tableau 1.

1. La notation $y_{n-k[N]}$ est à entendre, y indice $n-k$ modulo N.
2. les signaux acceptés sont les distributions tempérées.
3. Plus exactement une distribution, puisque les convergences sont à comprendre au sens des distributions.

propriétés	en temps	en fréquence
translation en temps	$x(t-t_0)$	$X(f)e^{-2j\pi t_0 f}$
translation en fréquence	$x(t)e^{2j\pi f_0 t}$	$X(f-f_0)$
linéarité	$ax(t)+by(t)$	$aX(f)+bY(f)$
convolution	$x(t)*y(t)$	$X(f)Y(f)$
produit	$x(t)y(t)$	$X(f)*Y(f)$
conjugaison	$x^*(t)$	$X^*(-f)$
retournement	$x(-t)$	$X(-f)$
conjugaison-retournement	$x^*(-t)$	$X^*(f)$
dérivation en temps	$x'(t)$	$2j\pi f X(f)$
dérivation en fréquence	$2j\pi t x(t)$	$X'(f)$

TAB. 1 – *Propriétés essentielles de la* TFTC

1.1.3 Transformée de Fourier à Temps Discret (TFTD)

La *Transformée de Fourier à Temps Discret* est une transformation bijective qui associe à un signal à temps discret [4] une fonction [5] définie sur l'intervalle $]-1/2, 1/2]$.

$$(x_n)_{n\in\mathbb{Z}} \longrightarrow X(f) = \sum_{n=-\infty}^{+\infty} x_n e^{-2j\pi n f} \qquad (\forall f \in]-\frac{1}{2}, \frac{1}{2}]) \qquad (5)$$

La transformation inverse s'obtient par intégration :

$$X(f) \longrightarrow x_n = \int_{-\frac{1}{2}}^{\frac{1}{2}} X(f) e^{2j\pi n f} \qquad (\forall n \in \mathbb{Z}) \qquad (6)$$

La variable f décrit le continuum des fréquences normalisées (continues), entre $-1/2$ et $1/2$, la fréquence $1/2$ correspondant à la fréquence de Nyquist. En règle générale, la transformée de Fourier n'a pas de raison d'être définie ailleurs que sur l'intervalle $]-1/2, 1/2]$. Toutefois, pour des raisons de compatibilité avec la définition de la transformée de Fourier à temps continu, il est commode de prolonger cette fonction par périodisation. Dans ce cas, on dira que la transformée de Fourier à temps discret est une fonction périodique de période unité. Pour un signal à temps discret issu de l'échantillonnage d'un signal à temps continu, la fréquence du signal continu se déduit de la fréquence du signal discret par multiplication par la fréquence d'échantillonnage. Le propriétés de la transformée de Fourier à temps discret sont résumées dans le tableau 1.1.3. On notera qu'un signal réel a nécessairement une transformée de Fourier a symétrie hermitienne : $X^*(f) = X(-f)$.

Il est important de noter que la TFTC est un outil théorique décrivant des caractéristiques *globales* d'un signal, défini de $n = -\infty$ à $n = +\infty$. Autrement dit, les caractéristiques locales du signal ne sont pas directement mises en évidence avec ce type de transformation.

4. Les séquences $(x_n)_{n\in\mathbb{Z}}$ acceptées sont les séquences à croissance lente, admettant un entier N tel que $(x_n)_{n\in\mathbb{Z}}$ soit équivalent à $|n|^N$ aux voisinage des infinis.

5. Là également, il s'agit en fait d'une distribution périodique.

propriétés	en temps	en fréquence
translation en temps	x_{n-n_0}	$X(f)e^{-2j\pi n_0 f}$
translation en fréquence	$x_n e^{2j\pi f_0 n}$	$X(f-f_0)$
linéarité	$ax_n + by_n$	$aX(f)+bY(f)$
convolution	$x_n * y_n$	$X(f)Y(f)$
produit	$x_n y_n$	$X(f) * Y(f)$
conjugaison	x_n^*	$X^*(-f)$
retournement	x_{-n}	$X(-f)$
conjugaison-retournement	x_{-n}^*	$X^*(f)$
dérivation en fréquence	$2j\pi n x_n$	$X'(f)$

TAB. 2 – *Propriétés essentielles de la* TFTD

On rencontre souvent dans la littérature la notation $X(e^{2j\pi f})$ à la place de $X(f)$ pour désigner la TFTC. Cette notation permet d'insister sur le fait que la TFTC est une fonction *périodique* de f, de période 1. Cette notation rappelle également que la TFTC correspond à l'évaluation de la *Transformée en* z (quand elle existe) notée traditionnellement $X(z)$, sur le cercle unité, c'est-à-dire pour $z = e^{2j\pi f}$. Toutefois, nous mettons en garde le lecteur sur la différence toute théorique qui existe entre la TFTC et l'évaluation sur le cercle unité de la transformée en z puisqu'il suffit que la TFTC ne soit pas continue pour que la transformée en z n'existe plus. En résumé la TFTC existe toujours [6] sans que nécessairement la transformée en z n'existe.

On rencontre également la notation $X(e^{j\omega})$, pour désigner la TFTC, auquel cas, il faut entendre ω au sens d'une pulsation normalisée, c'est-à-dire que la fréquence normalisée se déduit de la pulsation normalisée par division par 2π.

1.1.4 Lemmes

Pour introduire les parties qui suivent, nous allons avoir besoin de quelques résultats généraux.

Lemme 1. *Soit ω une racine N-ième de l'unité engendrant toutes les autres racines de l'unité (i.e. $\omega^k \neq 1$ pour $0 < k < N$ et $\omega^N = 1$), et Ω la matrice $N \times N$ constituée des éléments $\omega^{(i-1)(j-1)}$ (les coordonnées des matrices sont données pour $(i,j) \in [1,N]^2$). Alors nous avons les résultats suivants (δ_n désigne le symbole de Kronecker et Id la matrice unitée) :*

$$\sum_{k=0}^{N-1} \omega^{nk} = N \sum_{r=-\infty}^{+\infty} \delta_{n-rN} \quad \text{et} \quad \Omega^*\Omega = N Id$$

Preuve. Dans un premier temps, nous allons calculer $\sum_{k=0}^{N-1} \omega^{nk}$. Constatons d'abord que :

– si $\omega^n = 1$ alors $\sum_{k=0}^{N-1} \omega^{nk} = \sum_{k=0}^{N-1} 1^k = N$.

6. La transformée de Fourier définie au sens de la convergence des distribution pour toutes les séries à croissance lente, c'est-à-dire dominée à l'infini par les séries exponentielles [GW95].

– sinon, la relation suivante peut se simplifier par $1-\omega^n$:

$$0 = 1 - 1 = 1 - (\omega^N)^n = 1 - (\omega^n)^N = (1-\omega^n)\left(\sum_{k=0}^{N-1} \omega^{nk}\right)$$

Ce qui amène $\sum_{k=0}^{N-1} \omega^{nk} = 0$.

Il nous faut à présent caractériser le cas $\omega^n = 1$. Il est évident que si il existe un entier r telque $n = rN$, alors $\omega^n = 1$. En ce qui concerne la réciproque, écrivons la division euclidienne de n par N : $n = rN + p$, avec $0 \leq p < N$. Nous avons $1 = \omega^n = \omega^{rN}\omega^p = \omega^p$. Comme nous avons supposé que ω engendrait toutes les racines de l'unité, nous obtenons $p = 0$. Donc en résumé :

$$\sum_{k=0}^{N-1} \omega^{nk} = \begin{cases} N & \text{si } \exists r \in \mathbb{Z}, n = rN \\ 0 & \text{sinon} \end{cases}$$

Dans la somme $\sum_{r=-\infty}^{+\infty} \delta_{n-rN}$, tous les termes sont nuls et la somme est nulle, sauf si il existe un r telque $n - rN = 0$, auquel cas δ_{n-rN} est le seul élément non nul de cette somme qui devient alors 1. La première démonstration s'achève donc :

$$\sum_{k=0}^{N-1} \omega^{nk} = N \sum_{r=-\infty}^{+\infty} \delta_{n-rN} \tag{7}$$

Maintenant, considérons l'élément $a_{i,j}$ de la matrice produit $\Omega^*\Omega$:

$$a_{i,j} = \sum_{k=1}^{N} \omega^{-(i-1)(k-1)}\omega^{(k-1)(j-1)} = \sum_{k=0}^{N-1} \omega^{(j-i)k}$$

D'après le résultat précédant, $a_{i,j}$ vaut N quand $j-i$ est nul ou multiple de N, et 0 ailleurs. Or i et j variant tout deux dans l'intervalle $[1 \ldots N]$, $a_{i,j}$ ne peut être non-nul que si $j-i$ est nul, c'est-à-dire que $a_{i,j} = N\delta_{i-j}$, ce qui correspond à la matrice identitée multiplié par N. □

Lemme 2. *La transformée de Fourier (à temps discret) d'un peigne de Dirac (discret) est un peigne de Dirac (continu) de largeur inversement proportionnel à l'originale.*

Preuve. On note $(d_n^N)_{n\in\mathbb{Z}}$ le peigne de Dirac discret de largeur N :

$$d_n^N = \sum_k \delta_{n-kN}$$

La transformée de Fourier (à temps discret) de $(d_n^N)_{n\in\mathbb{Z}}$ ne se calcule pas directement parce que la somme ne converge qu'au sens des distributions. Par contre, nous pouvons écrire (les bornes de la somme ne sont réécrites que pour des raisons de commodité) :

$$d_n^N = \sum_k \delta_{n-kN} = \frac{1}{N}\sum_{k=0}^{N-1} e^{2j\pi\frac{k}{N}n} = \frac{1}{N}\sum_{k=-\lfloor\frac{N}{2}\rfloor}^{\lfloor\frac{N+1}{2}\rfloor} e^{2j\pi\frac{k}{N}n}$$

Or, $(e^{2j\pi\frac{k}{N}n})_{n\in\mathbb{Z}}$ a pour transformée de Fourier, quand $\frac{k}{N}\in]-\frac{1}{2},\frac{1}{2}]$, la distribution $\delta(f-\frac{k}{N})$, puisque :

$$\int_{-\frac{1}{2}}^{\frac{1}{2}}\delta(f-\frac{k}{N})e^{2j\pi fn}\,df = e^{2j\pi\frac{k}{N}n}$$

Donc par linéarité, $(d_n^N)_{n\in\mathbb{Z}}$ a pour transformée de Fourier (à temps discret) $\frac{1}{N}\sum_{k=-\lfloor N/2\rfloor}^{\lfloor(N+1)/2\rfloor}\delta(f-\frac{k}{N})$. Cette distribution est définie sur l'intervalle $]-1/2,1/2]$. Si nous adoptons une notation prenant en compte la périodisation de cette distribution en dehors de cet intervalle, nous obtenons pour transformée de Fourier de $(d_n^N)_{n\in\mathbb{Z}}$:

$$D^N(f)=\frac{1}{N}\sum_k\delta(f-k/N)$$

□

Lemme 3. *Soit $(u_n)_{n\in\mathbb{Z}}$ la suite définie par périodisation de la suite $(w_n)_{n\in\mathbb{Z}}$:*

$$u_n=\sum_{k\in\mathbb{Z}}w_{n-kR} \tag{8}$$

La suite $(u_n)_{n\in\mathbb{Z}}$ est constante si et seulement si $W(k/R)=\delta_k$ (pour $\frac{k}{R}\in]-\frac{1}{2},\frac{1}{2}]$), où $W(f)$ représente la transformée de Fourier (à temps discret) de $(w_n)_{n\in\mathbb{Z}}$.

Preuve. La démonstration repose essentiellement sur la réécriture de $(u_n)_{n\in\mathbb{Z}}$ sous forme d'un produit de convolution. On note $(d_n^R)_{n\in\mathbb{Z}}$ le peigne de Dirac de largeur R. $(u_n)_{n\in\mathbb{Z}}$ apparaît alors comme la convolution de $(d_n^R)_{n\in\mathbb{Z}}$ par $(w_n)_{n\in\mathbb{Z}}$. Les transformées de Fourier (à temps discret) sont donc reliées par la relation suivante :

$$v_n=d_n^R*w_n \qquad\Rightarrow\qquad V(f)=D^R(f)W(f)=\frac{1}{N}\sum_k W(f)\delta(f-k/R)$$

Nous constatons que $V(f)$ correspond à un spectre de raies, dont les raies sont toutes séparées de $1/R$ (fréquence normalisée). Maintenant, nous pouvons conclure : la suite $(u_n)_{n\in\mathbb{Z}}$ est constante si et seulement si $V(f)$ est nulle partout, sauf à la fréquence nulle, ce qui est équivalent à dire que $W(k/R)$ s'annule partout, sauf en 0. □

1.1.5 Transformée de Fourier Discrète (TFD)

La *Transformée de Fourier Discrète* est une transformation bijective qui associe à un vecteur $\boldsymbol{x}$ (dans le domaine temporel) un autre vecteur $\boldsymbol{X}$ (dans le domaine fréquenciel) de même taille. Elle est connue également sous le terme anglais *Discrete Fourier Transform*, ou bien son acronyme anglais DFT. Elle est définie par :

$$\boldsymbol{x}=(x_n)_{n=0\ldots N-1}\longrightarrow\boldsymbol{X}=(X_k)_{k=0\ldots N-1} \tag{9}$$

$$\forall k\in[0\ldots N-1],\ X_k=\sum_{n=0}^{N-1}x_ne^{-2j\pi\frac{nk}{N}} \tag{10}$$

Propriétés de la TFD : on obtient une notation vectorielle simple en utilisant $\omega = e^{-\frac{2j\pi}{N}}$, et Ω la matrice correspondante :

$$\boldsymbol{X} = \Omega \boldsymbol{x} \quad \text{et} \quad \boldsymbol{x} = \frac{1}{N}\Omega^* \boldsymbol{X} \tag{11}$$

Donc la transformation inverse s'obtient par sommation :

$$\forall k \in [0 \ldots N-1],\ x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k e^{2j\pi\frac{nk}{N}} \tag{12}$$

Les autres propriétés sont résumées dans le tableau 1.1.5. Il est remarquable de

propriétés	en temps	en fréquence
linéarité	$a\boldsymbol{x} + b\boldsymbol{y}$	$a\boldsymbol{X} + b\boldsymbol{Y}$
convolution	$x_n \circledast y_n$	$X_k.Y_k$
produit	$x_n.y_n$	$X_k \circledast Y_k$
conjugaison	$\boldsymbol{x}^*$	X^*_{N-k}
retournement	x_{N-n}	X_{N-k}
conjugaison-retournement	x^*_{N-n}	$\boldsymbol{X}^*$
modulation	$x_n e^{2j\pi n k_0/N}$	$X_{k-k_0[N]}$
permutation circulaire	$x_{n-n_0[N]}$	$X_k e^{-2j\pi k n_0/N}$

TAB. 3 – *Propriétés essentielles de la* TFD

noter que la modulation par un vecteur $e^{2j\pi nk/N}$ conduise à une permutation circulaire dans le domaine transformé. Si le vecteur original est réel, alors la symétrie hermitienne se traduit par $X_{N-k} = X^*_k$. En conséquence, la transformée de Fourier discrète d'un vecteur réel de $2N$ points conduit à seulement N valeurs significatives complexes (i.e. $2N$ valeurs significatives réelles). Il est possible d'arranger astucieusement les $2N$ termes réels du vecteur original de façon à utiliser la transformée de Fourier discrète (codé pour des nombres complexes) sur N points et retrouver les parties réels et imaginaires parmis le résultat donné sur N valeurs complexes.

Liens entre tftd et tfd : à partir des définitions précédantes, nous déduisons que la TFD correspond à l'évaluation aux fréquences $f = \frac{k}{N}$ de la TFTD du signal $(\ldots, 0, 0, x_0, x_1, \ldots x_{N-1}, 0, 0, \ldots)$:

$$\forall k \in [0 \ldots N-1],\ X_k = X(\frac{k}{N}) \tag{13}$$

Zero-padding : la technique du *zero padding* permet d'obtenir une évaluation fréquencielle du vecteur $(x_0, \ldots x_{N-1})$ sur plus de N points. En effet, d'après les remarques précédantes, le rajout de P zéros dans le vecteur conduit à l'évaluation de la TFTD sur $N + P$ points. Par exemple, la TFD du vecteur de $2N$ points $(x_0, \ldots x_{N-1}, 0, \ldots, 0)$ aboutit à l'échantillonnage fréquenciel du même signal $(\ldots, 0, 0, x_0, x_1, \ldots x_{N-1}, 0, 0, \ldots)$ sur $2N$ points au lieu de N points.

Dit autrement, pour évaluer un spectre à des fréquences qui ne tombent pas directement sur la TFD, la meilleur approximation consiste, non pas à interpoler directement entre les points déjà obtenus, mais à *zero-padder* le signal original en calculant une nouvelle TFD.

Si l'on note $(\Pi_n^{p,q})_{n\in\mathbb{Z}}$ le signal *porte* constitué de zéro partout sauf sur l'intervalle $[p,q]$ (où il est constitué de 1), alors la TFD de $(x_0,\dots x_{N-1})$ apparait comme l'échantillonnage sur N points de la TFTD du produit du signal $(x_n)_{n\in\mathbb{Z}}$ avec $(\Pi_n^{0,N-1})_{n\in\mathbb{Z}}$, qui correspond à la convolution de $X(f)$ avec la TFTD de la porte $(\Pi_n^{0,N-1})_{n\in\mathbb{Z}}$ (qui est un sinus cardinal). Cela signifie que la résolution fréquencielle maximale possible avec la technique du *zero-padding* est limitée par la résolution fréquencielle du sinus cardinal [7].

Transformée de Fourier Rapide : cette transformée, plus connue sous son acronyme anglais FFT (*Fast Fourier Transform*) n'en est en fait pas une, puisqu'il s'agit d'un algorithme *rapide* permettant de calculer efficacement la TFTD [Bel83]. Normalement, le calcul matriciel de la TFTD fait intervenir approximativement un nombre d'opérations proportionnel à N^2 (il s'agit en effet de multiplier une matrice $N\times N$ par un vecteur de taille N). Mais les structures caractéristiques des matrices Ω et Ω^* (pour certaines tailles) permettent de factoriser certains calculs. C'est en particulier pour des tailles multiples de 4, ou bien puissances entières de 2 qu'il est possible de mettre en évidence sur Ω une structure de *produit tensoriel* qui conduit à un algorithme rapide dont le nombre d'opération est proportionnel à $N\log N$.

1.2 Transformée de Fourier à Court Terme (TFCT)

La TFCT est une représentation théorique [Por80] qui permet de prendre en compte à la fois les propriétés locales (temporelles) du signal, et ses propriétés fréquencielles. La définition fait intervenir une *fenêtre d'analyse*, notée $(w_n)_{n\in\mathbb{Z}}$, permettant de prendre en compte la localisation des événements temporels. Le choix de cette fenêtre d'analyse est un degré de liberté qui modifie conjointement les résolutions temporelle et fréquencielle de la transformée. Nous noterons N la largeur, en nombre d'échantillons, de son support temporel, quand celui-ci est fini. Par définition, la TFCT de $(x_m)_{m\in\mathbb{Z}}$, notée $X^n(f)$, également connue sous son nom anglais *Short-Time Fourier Transform*, ou bien STFT est définie par :

$$X^n(f) = \sum_{m=-\infty}^{+\infty} x_m w_{n-m} e^{-2j\pi mf} \tag{14}$$

La somme sur laquelle porte la définition n'est réellement infinie que lorsque la fenêtre d'analyse est elle-même infinie. Dans la pratique le support temporel de la fenêtre est fini, et la somme ne porte que sur un nombre fini de termes, en l'occurrence N.

1.2.1 Interprétations

Le regroupement des trois termes dans l'expression de définition conduit à trois interprétations différentes, deux sous la forme d'un *produit de convolution*, et l'une sous la forme d'une *transformée de Fourier*, $(w_m)_{m\in\mathbb{Z}}$ jouant successivement le rôle soit de fenêtre temporelle, soit de filtre d'analyse.

7. Et plus précisément, la résolution fréquencielle dépend de la largeur du lobe principal du sinus cardinal.

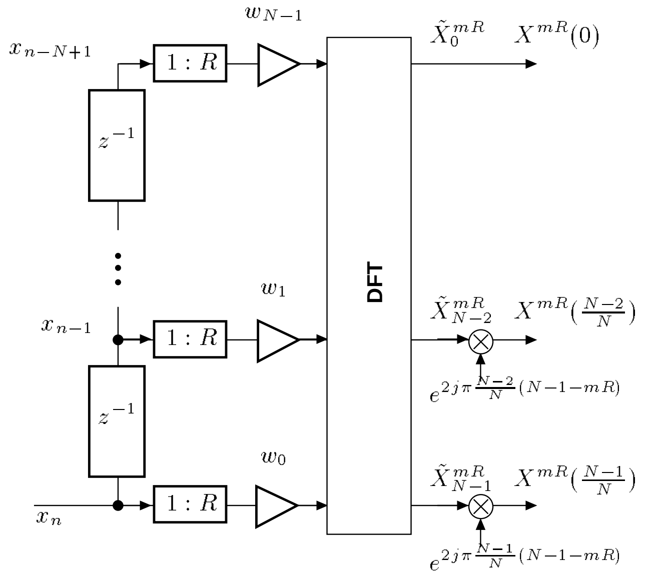


FIG. 1 – *Interprétation en transformée de Fourier.*

Interprétation en transformée de Fourier : Dans le cas où la fenêtre d'analyse est à support temporel fini ($N < \infty$), l'évaluation de la TFCT aux fréquences multiples de $1/N$ s'obtient par transformée de Fourier discrète du vecteur fenêtré $(w_{N-1}x_{n-N+1}, \ldots, w_0 x_n)$. Plus exactement :

$$\tilde{X}_k^n = \sum_{m=0}^{N-1} w_{N-1+m} x_{n-N+1+m} e^{-2j\pi\frac{km}{N}}$$

La TFCT se déduit de la transformée de Fourier discrète $\tilde{X}_k^n$ par modulation. Le schéma de principe est donné en figure 1.

$$X^n(\frac{k}{N}) = e^{2j\pi\frac{k}{N}(N-1-n)}\tilde{X}_k^n$$

Interprétation en banc de filtre : Il s'agit à présent de considérer la TFCT comme la convolution (i.e. le filtrage) de $(x_m)_{m\in\mathbb{Z}}$ par $\left(h_m^{(f)}\right)_{m\in\mathbb{Z}} = \left(w_m e^{2j\pi mf}\right)_{m\in\mathbb{Z}}$:

$$X^n(f) = e^{-2j\pi nf} \sum_{m=-\infty}^{+\infty} x_m \left(w_{n-m}e^{2j\pi(n-m)f}\right)$$

Les filtres $\left(h_n^{(f)}\right)_{n\in\mathbb{Z}}$ apparaissent alors comme la modulation du filtre original $(w_n)_{n\in\mathbb{Z}}$ par un porteuse à la fréquence f. On obtient alors le signal $(X^n(f))_{n\in\mathbb{Z}}$ par *démodulation* de la sortie de chaque filtre :

$$X^n(f) = \left(h_n^{(f)} * x_n\right) e^{-2j\pi nf}$$

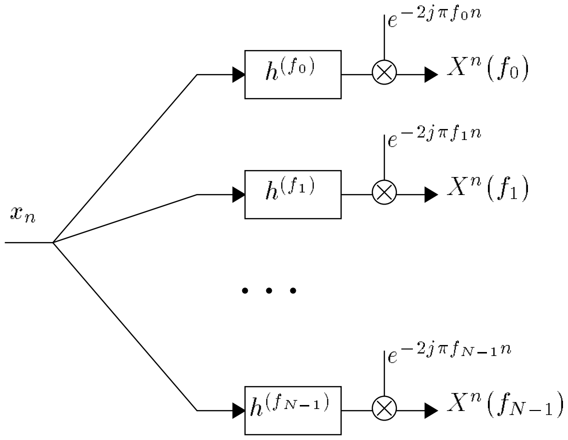


FIG. 2 – *Interprétation en banc de filtre.*

On désigne de *canal d'analyse* la procédure qui transforme le signal d'origine $(x_n)_{n\in\mathbb{Z}}$ en $(X^n(f))_{n\in\mathbb{Z}}$. Si on choisit un filtre passe-bas comme filtre d'analyse, alors chaque canal se décompose en une opération de filtrage autour de la fréquence f puis d'une opération de démodulation à la même fréquence. L'ensemble de la TFCT apparaît comme un *banc de filtre*, c'est-à-dire un ensemble de filtres en parallèle agissant simultanément sur $(x_n)_{n\in\mathbb{Z}}$. La figure 2 est à comparer avec la figure 1. À un terme multiplicatif près, la transformée de Fourier agit donc comme un banc de filtre passe-bande.

Interprétation en banc de filtres passe-bas : La troisième et dernière interprétation correspond à la dernière façon de regrouper les termes dans la définition : la convolution du signal $\left(s_m^{(f)}\right)_{m\in\mathbb{Z}} = \left(e^{-2j\pi m f} x_m\right)_{m\in\mathbb{Z}}$ avec $(w_m)_{m\in\mathbb{Z}}$:

$$X^n(f) = \sum_{m=-\infty}^{+\infty} \left(x_m e^{-2j\pi m f}\right) w_{n-m}$$

Vu sous cet aspect, chaque canal d'analyse correspond à la démodulation du signal $(x_n)_{n\in\mathbb{Z}}$ (permettant de descendre la partie du signal située à la fréquence f en *bande de base*), suivi du filtrage par le filtre d'analyse $(w_n)_{n\in\mathbb{Z}}$. En ce sens, chaque canal découpe une partie du spectre de $(x_n)_{n\in\mathbb{Z}}$, et le dépose en bande de base. Ce qui nous invite à constater que si le filtre d'analyse est un filtre passe-bas, alors la sortie de chaque canal est constituée d'un signal basse fréquence, traduisant les caractéristiques du signal original à la fréquence f.

1.2.2 Simplifications

Le fait d'effectuer tous les calculs sur un ordinateur nous contraint à simplifier quelque peu notre représentation. En pratique, il est impossible d'évaluer $X^n(f)$ pour toute fréquence de l'intervalle $]-1/2, 1/2]$: l'évaluation ne peut porter que sur un nombre fini de fréquences distinctes. En particulier, si le nombre

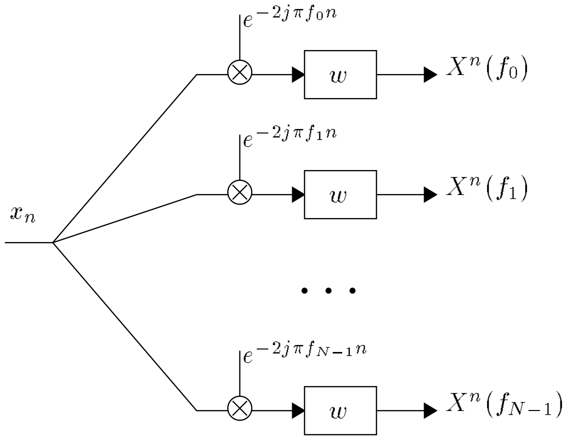


FIG. 3 – *Interprétation en banc de filtre passe-bas.*

de points de la fenêtre d'analyse est N fini, le théorème de Nyquist nous garanti la réversibilité de l'échantillonnage fréquenciel de $X^n(f)$ pour un minimum de N points, notés (f_k). Si de plus l'évaluation porte sur des fréquences équi-réparties $f_k = k/N$, alors l'obtention des $X^n(f_k)$ se fait simplement par transformée de Fourier discrète (figure 1). Naturellement, cette transformée de Fourier discrète est implémentée par une FFT. Chaque sortie de FFT est désignée alors sous le nom de *canal d'analyse*, ce qui trouve sa justification dans l'interprétation en banc de filtre.

Il est possible de développer des techniques spécifiques pour évaluer la transformée de Fourier à court-terme pour des fréquences non-équiréparties. Il existe par exemple de nombreuses variantes à l'algorithme de FFT permettant d'évaluer la transformée de Fourier à temps discret sur d'autres point que les $f_k = k/N$[Bel83]. Il est également possible d'utiliser une technique particulière, dite de *warped filters* pour évaluer cette même transformée de Fourier sur une échelle de fréquence modifiée [Cho82, EC97, KHL97]. Dans ce dernier cas, il peut s'avérer nécessaire d'évaluer $X^n(f)$ sur un nombre de point supérieur à N pour garantir la reconstructibilité de $X^n(f)$ à partir des $X^n(f_k)$. Il n'est pas nécessaire non plus de se restreindre à des fenêtres $(w_n)_{n\in\mathbb{Z}}$ de taille finie [Sas96].

La deuxième simplification traite du problème de la redondance des données en sortie de TFCT. En effet, la TFCT multiplie d'un facteur N (au moins 1024 en pratique) le débit de données, ce qui est incompatible avec la plupart des applications. La TFCT ne constitue évidemment pas un système de compression de l'information, mais elle aurait tout à gagner si il était possible d'optimiser conjointement l'échantillonnage fréquenciel et temporel, afin d'éliminer une partie de la redondance de la transformée. Dans ces conditions, le filtre d'analyse doit être localisé à la fois en temps et en fréquence, ce qui justifie l'utilisation de filtres d'analyse symétriques passe-bas. Nous avons montré au chapitre précédant que dans ce cas, la sortie de chaque canal est basse fréquence. Il est donc légitime de sous-échantillonner chaque canal par un facteur R, de telle façon que que le débit général ne soit multiplié que d'un facteur N/R. L'utilisation de fenêtres symétriques se justifie pour leur phase linéaire qui déforme peu la forme d'onde du signal entrant.

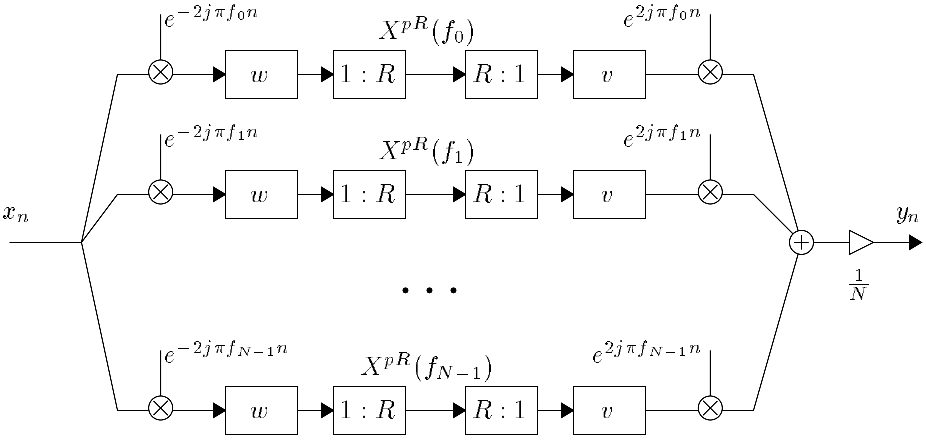


FIG. 4 – *Reconstruction en banc de filtre*

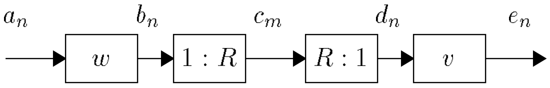


FIG. 5 – *Module de reconstruction*

En résumé, les éléments de la TFCT facilement accessibles en pratique sont les $\left(X^{pR}(f_k)\right)_{p\in\mathbb{Z}}$ pour $0 \leq k < N$.

1.3 Reconstruction à partir de TFCT

La TFCT constitue un système d'analyse. En sortie de chaque trame temporelle, l'utilisateur dispose d'informations de nature fréquencielle qu'il peut mettre à profit pour détecter des caractéristiques du signal telles que en traitement de la parole, la prosodie de la voix (i.e. les variations de hauteur dans le signal de parole), le degré de voisement, le degré d'inharmonicité, la position des formants, la fréquence des partiels ... La reconstruction du signal ne s'impose que dans le cas de systèmes d'*analyse-synthèse*, dont les multiples applications sont essentiellement tournées vers la transformations des signaux audio, par exemple, transformation du timbre, modification de la hauteur, interpolation entre deux timbres, modification de la prosodie ...

On peut imaginer un système dual de l'analyse pour reconstruire le signal à partir des information de la transformée de Fourier à court-terme. L'analyse avait essentiellement donnée lieu à deux interprétations. La synthèse donne donc lieu à deux interprétations, dites « méthodes de reconstruction ».

1.3.1 Filter Bank Summation (FBS)

Cette méthode de reconstruction [AR77] est la duale de l'interprétation en banc de filtre. Pour chaque canal de synthèse on procède au suréchantillonnage d'un facteur R, suivi d'un filtrage par le filtre de synthèse $(v_n)_{n\in\mathbb{Z}}$ puis enfin à la modulation à la fréquence porteuse f_k (voir figure 4). Ce schéma fait apparaître le bloc élémentaire décrit en figure 5, constitué du filtre d'analyse, suivi d'un bloc de sous-échantillonnage, puis celui de sur-échantillonnage, et enfin du filtre de synthèse. Ce bloc est décrit par les relations suivantes :

$$\left\{\begin{array}{lcl} b_n & = & \sum_k a_k w_{n-k} \\ c_m & = & b_{mR} \\ d_n & = & \sum_m c_m \delta_{n-mR} \\ e_n & = & \sum_p d_p v_{n-p} \end{array}\right. \Rightarrow \boxed{e_n = \sum_p a_p \left(\sum_m w_{mR-p} v_{n-mR}\right)} \tag{15}$$

Preuve. la preuve s'obtient par applications successives des différentes expressions de définitions :

$$\begin{aligned} e_n = \sum_p d_p v_{n-p} = \sum_p \sum_m v_{n-p} c_m \delta_{p-mR} = \sum_m v_{n-mR} c_m \\ = \sum_m v_{n-mR} b_{mR} = \sum_m \sum_p v_{n-mR} a_p w_{mR-p} \end{aligned}$$

□

La reconstruction s'obtient par utilisation de la relation (15) pour chacun des N canaux, puis par sommation sur tous les canaux :

$$a_p^k = x_p e^{-2j\pi p f_k} \qquad \text{et} \qquad e_n^k = \sum_p a_p^k \sum_m w_{mR-p} v_{n-mR}$$

$$y_n = \frac{1}{N} \sum_{k=0}^{N-1} e_n^k e^{2j\pi n f_k} = \frac{1}{N} \sum_p \sum_m x_p w_{mR-p} v_{n-mR} \sum_{k=0}^{N-1} e^{2j\pi(n-p)f_k}$$

Dans le cas où les fréquences sont équiréparties ($f_k = k/N$) le dernier terme se réduit à (en utilisant le lemme 1) :

$$\sum_{k=0}^{N-1} e^{2j\pi(n-p)f_k} = \sum_{k=0}^{N-1} \left(e^{\frac{2j\pi}{N}}\right)^{(n-p)k} = N \sum_r \delta_{n-p-rN}$$

Ce qui aboutit à la relation finale pour l'expression du signal reconstruit en utilisant δ_{n-p-rN} pour simplifier la somme sur p :

$$y_n = \sum_r x_{n-rN} \left(\sum_m w_{mR-n+rN} v_{n-mR}\right) \tag{16}$$

La reconstruction parfaite ($y_n = x_{n-pN}$, c'est-à-dire avec retard de pN échantillons) est atteinte si et seulement si :

$$\boxed{\forall r,\, \forall n,\ \sum_m w_{mR-n+rN} v_{n-mR} = \delta_{r-p}} \tag{17}$$

Si on note $(t_n^r)_{n\in\mathbb{Z}}$ la suite produit $(w_{rN-n}v_n)_{n\in\mathbb{Z}}$, alors la condition de reconstruction parfaite se traduit par une condition sur les suites périodisées $(t_n^r)_{n\in\mathbb{Z}}$:

$$\forall r,\ \sum_m t_{n-mR}^r = \delta_{r-p} \tag{18}$$

Autrement dit, toutes les suites périodisées sont nulles, sauf une unique qui est identiquement égale à 1. L'application du lemme 3 sur cette dernière relation aboutit, en notant $T^r(f)$ la tranformée de Fourier (à temps discret) de $(t_n^r)_{n\in\mathbb{Z}}$, à :

$$\forall r,\ \forall k,\ T^r(\frac{k}{R}) = \delta_k\delta_{r-p} \qquad \text{avec } -\frac{1}{2} < \frac{k}{R} \leq \frac{1}{2} \tag{19}$$

Dans le cas où les fenêtres sont à support fini et $R > N/2$, la condition de reconstruction parfaite aboutit essentiellement à une relation du type $\forall n,\ \sum_m w_{mR-n}v_{n-mR} = 1$, les autres sommes $r \neq 1$ s'annulant d'elles-mêmes car ne mettant en jeux que des termes nuls.

Quand les deux filtres d'analyse et de synthèse sont causaux et de support contenu dans l'intervalle $[0, N-1]$, alors la suite produit $(t_n^0)_{n\in\mathbb{Z}}$ est nécessairement identiquement nulle[8]. La condition de reconstruction parfaite ne peut donc être validée que pour une valeur p non nulle, qui plus est, positive pour assurer la causalité des calculs. Donc si les deux filtres sont causaux, alors nécessairement la reconstruction n'est possible qu'avec un retard proportionnel à N.

1.3.2 Overlap and add (OLA)

Cette méthode de reconstruction est le dual de l'interprétation en transformée de Fourier. Pour chaque trame une transformée de Fourier inverse permet de retrouver un vecteur de nature temporelle. La superposition par décalage successif de chacun des vecteurs permet de retrouver la continuité temporelle des échantillons.

Cette méthode devrait faire intervenir le dual d'une ligne à retard (une ligne en avance) et les termes symétrique du filtre d'analyse $v_0,\ v_{-1}\ \ldots$ Pour des raisons de causalité, on inverse le flux des données, on remplace les opérateurs d'avance par des opérateurs de retard, ce qui explique le sens de la ligne à retard dans le schéma de reconstruction OLA. Cela ne change rien, à part un retard incompressible[9] dans le système d'analyse-synthèse de $N-1$ échantillons.

8. Sauf éventuellement pour $t_0^0 = v_0w_0$ qui pourrait être le seul élément non nul de la séquence $(t_n^0)_{n\in\mathbb{Z}}$. Mais dans ce cas, cette suite est proportionnelle à une impulsion, la suite périodisée à un train d'impulsion. La condition de reconstruction impose alors à ce train d'impulsion d'être proportionnel à une suite constante ou nulle, ce qui ne peut être vérifié que si $R = 1$ (pas de suréchantillonnage) ou $v_0w_0 = 0$ ce qui en contradiction avec l'hypothèse de départ.

9. Remarquons au passage que la valeur de N, donnée initialement comme la largeur des supports des fenêtres d'analyse et de synthèse, est forcément surévaluée puisque le premier point d'une des deux fenêtres est nécessairement nul (voir la note 8). Cela induit un retard supplémentaire qui n'apparait pas explicitement dans le schéma d'analyse synthèse. Ceci explique l'apparante contradiction entre le retard incompressible de $N-1$ échantillons (mais en fait de N échantillons à cause de la condition de reconstruction parfaite) et le retard proportionnel à N échantillons du paragraphe précédent.

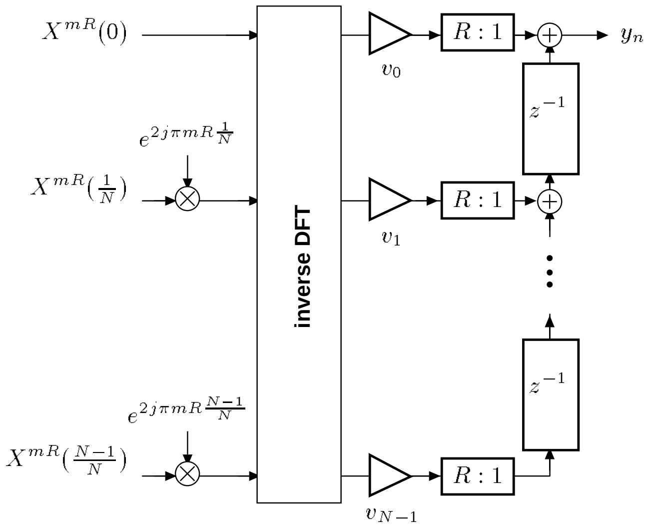


FIG. 6 – *Module de reconstruction par* overlap and add.

1.3.3 Équivalence OLA - FBS :

Les deux méthodes de reconstruction sont duales de deux interprétations équivalentes. Il est donc logique de s'attendre à ce qu'elle soient elles-mêmes équivalentes. Un calcul détaillé permet de se rendre compte de cette évidence.

Plaçons nous dans le cas $R = 1$. On considère la sortie des modules de reconstruction contrôlés par les mêmes $X^n(\frac{k}{N})$. Dans le cas FBS, la sortie s'obtient en sommant le résultat de la modulation du filtrage par v de chaque canal :

$$y_n^{\text{FBS}} = \frac{1}{N}\sum_k e^{2j\pi f_k n}\sum_p v_p X^{n-p}(f_k) = \frac{1}{N}\sum_k\sum_p v_p X^{n-p}(f_k)e^{2j\pi f_k n} \quad (20)$$

Dans le cas de la reconstruction par superposition-addition, on calcule dans un premier temps y_n^p, la sortie du bloc de transformée de Fourier discrète inverse à l'instant n :

$$y_n^p = \frac{1}{N}\sum_{k=0}^{N-1} X^n(\frac{k}{N})e^{2j\pi n\frac{k}{N}}e^{2j\pi\frac{kp}{N}} \quad (21)$$

La sortie du bloc s'obtient alors par superposition-décalage :

$$y_n^{\text{OLA}} = \sum_{p=0}^{N-1} v_p y_{n-p}^p = \frac{1}{N}\sum_{p=0}^{N-1}\sum_{k=0}^{N-1} v_p X^{n-p}(\frac{k}{N})e^{2j\pi(n-p)\frac{k}{N}}e^{2j\pi\frac{kp}{N}}$$
$$= \frac{1}{N}\sum_{p=0}^{N-1}\sum_{k=0}^{N-1} v_p X^{n-p}(\frac{k}{N})e^{2j\pi\frac{nk}{N}} \quad (22)$$

Dans le cas $R \neq 1$, les calculs sont un peu plus lourd. Pour la méthode dite FBS, le résultat du suréchantillonnage en sortie du $k^{\text{ème}}$ canal de synthèse donne :

$$e^{2j\pi f_k n}\sum_p\left(v_{n-p}\sum_m X^{mR}(f_k)\delta_{p-mR}\right)=\sum_m v_{n-mR}X^{mR}(f_k)$$

La sortie du bloc aboutit alors à :

$$y_n^{\mathrm{FBS}}=\frac{1}{N}\sum_k\sum_m v_{n-mR}X^{mR}(f_k)e^{2j\pi f_k n} \tag{23}$$

En ce qui concerne la reconstruction par superposition-décalage, la sortie du bloc de transformée de Fourier discrète inverse n'a pas changé (21). La sortie du bloc s'obtient encore par superposition-décalage, mais avec moins de termes que précédemment (avec la convention $v_p = 0$ quand celui-ci n'est pas défini) :

$$\begin{aligned}y_n^{\mathrm{OLA}} &= \sum_p v_{n-p}y_p^{n-p} = \sum_m v_{n-mR}y_{mR}^{n-mR}\\ &= \frac{1}{N}\sum_m\sum_{k=0}^{N-1} v_{n-mR}X^{mR}(\frac{k}{N})e^{2j\pi mR\frac{k}{N}}e^{2j\pi\frac{k(n-mR)}{N}}\\ &= \frac{1}{N}\sum_m\sum_{k=0}^{N-1} v_{n-mR}X^{mR}(\frac{k}{N})e^{2j\pi\frac{kn}{N}}\end{aligned} \tag{24}$$

De la même façon que nous avions montré que l'analyse pouvait s'interpréter indifférement en banc de filtre ou en transformée de Fourier (moyennant des hypothèses sur la taille de la fenêtre et les valeurs des fréquences f_k), nous montrons ici que la synthèse s'interprète selon les mêmes hypothèses indifférement en sommation du banc de filtre, ou en superposition-décalage.

On remarque dans l'interprétation en transformée de Fourier du schéma d'analyse-synthèse la simplification des termes de modulation et de démodulation $\exp(2j\pi kmR/N)$. En pratique ces multiplications ne sont pas effectuées. On tient toutefois compte de leur présence quand on opère des modifications (par exemple étirement temporel) sur le signal.

1.3.4 Exemple

On prend un pas d'avancement de 50% ($2R = N$), une fenêtre d'analyse de Hann ($w_n = \sin^2(\pi n/2R)$ pour $0 \le n < N$) et une fenêtre de synthèse rectangulaire ($v_n = 1$ pour $0 \le n < N$).

On vérifie aisément la condition de reconstruction. Considérons la séquence $(t_n^r)_{n\in\mathbb{Z}}$ définie par le produit $(w_{rN-n}v_n)_{n\in\mathbb{Z}}$. Pour $r \neq 1$, les deux fenêtres ne se recouvrent pas, donc $(t_n^r)_{n\in\mathbb{Z}}$ est nulle, ainsi que leurs versions périodisées. Dans le cas $r = 1$, il faut vérifier que la périodisation de $(t_n^1)_{n\in\mathbb{Z}}$ donne systématiquement 1. Or, une simple considération graphique permet de constater que seuls deux alias successifs de la fenêtre (t_n^1) se recouvrent. Calculons la somme de t_n^1 et t_{n+R}^1 :

$$\begin{aligned}t_n^1+t_{n+R}^1 &= \frac{1-\cos(\pi\frac{n}{2R})}{2}+\frac{1-\cos(\pi\frac{n+R}{2R})}{2}\\ &= 1-\frac{1}{2}\left(\cos(\pi\frac{n}{2R})-\cos(\pi\frac{n}{2R})\right)=1\end{aligned}$$

1.4 Reconstruction à partir d'une TFCT modifiée

Les sections précédentes ont montré que la TFCT était une représentation redondante du signal, notemment à cause de la multiplication du débit et du passage d'un espace monodimensionnel à un espace bidimensionnel. Il apparaît donc que la transformation (linéaire) est injective, sans être surjective. Donc parmi l'ensemble des X_k^{mR}, un certain nombre seulement sont des images *valides* d'un signal à temps discret.

La méthode de reconstruction précédente supposait implicitement qu'il existait un signal x_n dont X_k^{mR} était l'image. Aussi la méthode perd-elle sa cohérence dès lors que X_k^{mR} n'est plus l'image d'aucun signal à temps discret. Il est légitime de se poser la question de savoir à quoi ressemble la transformée de Fourier à court-terme $\hat{X}_k^{mR}$ d'un signal y_n construit à partir de X_k^{mR} quelconques.

Griffin et Lim dans [GL84] propose une méthode consistant à choisir le signal y_n de telle façon que la distance $|\hat{X} - X|$ soit minimale pour un critère quadratique. Cette méthode consiste donc à trouver une représentation temps-fréquence valide à partir d'une représentation temps-fréquence quelconque, puis de l'inverser. La représentation temp-fréquence valide est obtenue par *projection* [10] de la représentation originale sur l'ensemble de toutes les représentations temps-fréquence valides.

1.4.1 Critère quadratique

Si la densité de l'échantillonnage fréquenciel est suffisante, alors il est possible de reconstituer $X^{mR}(f)$ à partir des $X^{mR}(f_k)$. La minimisation porte alors sur le critère J, où $\hat{X}^{mR}(f)$ représente la transformée de Fourier à court-terme de $(y_n)_{n\in\mathbb{Z}}$:

$$J = \sum_m \int_{-\frac{1}{2}}^{\frac{1}{2}} |X^{mR}(f) - \hat{X}^{mR}(f)|^2 \, df$$

$X^{mR}(f_k)$ n'est pas une TFCT valide, $X^{mR}(f)$ n'en admet pas moins une transformée de Fourier inverse, notée $(x_p^{mR})_{p\in\mathbb{Z}}$:

$$\forall m, \ \ \forall p, \ \ x_p^{mR} = \int_{-\frac{1}{2}}^{\frac{1}{2}} X^{mR}(f) e^{2j\pi pf} \, df$$

La transformée de Fourier inverse de $\hat{X}^{mR}(f)$ est quant à elle d'après (14), à peu de chose près, $(y_p)_{p\in\mathbb{Z}}$:

$$\forall m, \ \ \forall p, \ \ \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{X}^{mR}(f) e^{2j\pi pf} = w_{mR-p}\, y_p$$

Par la relation de Perceval, la norme $\int |X - \hat{X}|^2 \, df$ est égale à celle calculée sur les échantillons : $\sum |x - wy|^2$. Le terme à minimiser devient alors une expression

10. Une projection minimise la distance d'un point à un ensemble.

simplement quadratique en $(y_p)_{p\in\mathbb{Z}}$:

$$J = \sum_m \sum_p |x_p^{mR} - w_{mR-p}\, y_p|^2$$
$$= \sum_p \left(y_p^2 \sum_m w_{mR-p}^2 - 2y_p \sum_m x_p^{mR} w_{mR-p} + \sum_m \left(x_p^{mR}\right)^2 \right)$$

La minimisation de cette expression par rapport à $(y_p)_{p\in\mathbb{Z}}$ s'obtient, par exemple, terme à terme. En l'occurence, on trouve :

$$\forall n \in \mathbb{Z},\ \ y_n = \left(\sum_m x_n^{mR} w_{mR-n} \right) \Big/ \left(\sum_m w_{mR-n}^2 \right) \tag{25}$$

1.4.2 Comparaison avec la reconstruction classique

Par extention, si les f_k sont équiréparties, x_p^{mR} se calcule par transformée de Fourier discrète inverse. Ainsi, la formule de reconstruction se trouve-t-elle être :

$$y_n = \frac{1}{N} \sum_m \sum_{k=0}^{N-1} \frac{w_{mR-n}}{\sum_p w_{pR-n}^2} X^{mR}(k/N) e^{2j\pi \frac{nk}{N}} \tag{26}$$

Cette expression est à comparer avec l'expression (23), pour laquelle on choisit $(v_n)_{n\in\mathbb{Z}}$ comme fenêtre de synthèse :

$$v_n = \frac{w_{-n}}{\sum_p w_{n-pR}^2} \tag{27}$$

Ce couple de fenêtres d'analyse/synthèse vérifie bien évidemment la condition de reconstruction parfaite[11] (17) :

$$\sum_m w_{mR-n+rN} v_{n-mR} = \left(\sum_m w_{mR-n+rN} w_{mR-n} \right) \Big/ \left(\sum_m w_{n-mR}^2 \right) = \delta_r$$

Finalement, cette méthode de reconstruction se révèle être un cas particulier de la reconstruction duale de la TFCT, où fenêtres d'analyse et de synthèse sont équilibrée par la relation (27). Cet effet d'équilibrage peut-être compris comme une tentative de répartition équitable des tâches d'analyse et de synthèse.

Cette méthode n'est évidemment pas directement causale. Il suffit cependant de retarder la fenêtre de synthèse (27) de N échantillons pour la rendre causale.

En outre, ce calcul permet de mettre en évidence que les expressions de reconstruction naturelle de la TFCT, précedemment exprimées sous la forme d'une estimation non-paramétrique de signal, peuvent prendre la forme d'une estimation paramétrique pour certains choix de fenêtres d'analyse et de synthèse.

11. On rappelle que N correspond à la largeur de la fenêtre d'analyse w. En conséquence, le terme $w_{mR-n+rN} w_{mR-n}$ est nécessairement nul dès que $r \neq 0$.

1.5 Filtrage par Transformée de Fourier à Court Terme

Dans le chapitre précédant, nous avons montré que la représention temps-fréquence de la transformée de Fourier à court terme permettait de reconstruire à l'identique le signal original. Il est tentant maintenant de manipuler cette représentation temps-fréquence pour obtenir par exemple une opération de filtrage dont le gabarit fréquenciel serait obtenu directement en atténuant ou amplifiant la sortie de chaque canal d'analyse, à la manière d'un égaliseur graphique. C'est ce problème que nous allons étudier dans cette section.

1.5.1 Gain variable

Chaque trame d'analyse est multipliée par un coefficient g^{mR} avant la phase de resynthèse. La résultat de cette modification s'obtient rapidement en adaptant la relation (15). Dans la figure 5, on multiplie c_m par g_{mR}. On obtient *in fine* la relation qui s'apparente à l'équation (16) :

$$y_n = \sum_r x_{n-rN} \left(\sum_m g_{mR} w_{mR-n+rN} v_{n-mR} \right) \tag{28}$$

1.6 Vocodeur de phase

Le vocodeur de phase est une application directe de la transformée de Fourier à court-terme. C'est une méthode d'analyse-synthèse qui suppose la présence d'une sinusoïde au plus dans chaque canal d'analyse. Supposons dans un premier temps que le signal x soit uniquement composé d'une sinusoïde à la fréquence f^0 : $\left(a \exp(2j\pi n f^0)\right)_{n\in\mathbb{Z}}$. La phase initiale de la sinusoïde est contenue dans le gain complexe a. Notons $W(f)$ la transformée de Fourier à temps discret de la fenêtre d'analyse. Il s'agit d'estimer l'amplitude et la fréquence de la sinusoïde.

$$\begin{aligned} X^n(f) &= a \sum_p w_{n-p} e^{2j\pi p f^0} e^{-2j\pi p f} \\ &= a \sum_p w_{n-p} e^{-2j\pi(n-p)(f^0-f)} e^{2j\pi n(f^0-f)} = aW(f^0-f) e^{2j\pi n(f^0-f)} \end{aligned} \tag{29}$$

Le module de $X^n(f)$ présente un pic spectral situé à la fréquence f^0. En pratique, la transformée de Fourier n'est connue que sur un nombre discret de points. La fréquence de la sinusoïde peut être détectée par exemple par interpolation quadratique entre les 3 plus proches voisins du pic spectral. On préfère en général une autre estimation de f^0. La différence de phase entre deux trame d'analyse nous donne en effet une autre indication précise sur f^0. Entre les instants mR et $(m+1)R$, la phase du canal k a évoluée de $2j\pi(f^0 - f_k)R$ radians. L'évaluation de cette différence de phase est en général difficile car celle-ci n'est connue qu'à 2π près. Mais si la fréquence d'analyse f_k du canal k est proche de f^0, et si le recouvrement est suffisant [12], la phase n'a pas le temps de faire plus d'un saut de π, et la différence de phase s'évalue alors simplement. Cette

12. Le recouvrement de deux fenêtres d'analyse est en pratique au minimum de 50%. Auquel cas, $|f^0 - f_k|$ est inférieur à $1/N$ qui, lui-même, est inférieur à $1/(2R)$. La différence de phase est alors inférieur à π, ce qui assure sa détermination unique.

estimation ne dépend ni de la nature de la fenêtre d'analyse ni de la phase initiale. Le *vocodeur de phase* tire parti de cet avantage en représentant les trames de données par le module de chaque canal d'analyse, et la *différence de phase* entre deux trames successives.

1.6.1 Analyse additive

Le modèle additif représente un signal x comme la superposition de plusieurs sinusoïdes de fréquence et d'amplitude lentement variables. L'analyse additive d'un signal correspond à l'estimation des fréquences, phases et amplitudes respectives de chacunes des sinusoïdes qui, dans le cadre du modèle additif, expliquent au mieux le signal observé. L'optimisation globale des paramètres, même pour un critère simple du type, moindre carré, est une tâche insurmontable. Il faut décomposer le signal sur différentes trames temporelles, et proposer une méthode d'estimation pour chacunes d'entre-elles. La technique du vocodeur de phase vue dans la section précédente est une technique d'estimation non-paramétrique des fréquences, phases, et amplitudes. Sur chaque trame d'analyse, on suppose les variations de fréquence et d'amplitude suffisement lente pour que le vocodeur de phase fournisse des estimations correctes.

La véritable détection des sinusoïdes se fait en utilisant un premier crible à seuil sur le module de la TFCT, puis un second tenant compte des résultats de la détection aux instants passés (et légèrement futurs). Vient ensuite la procédure d'appariement des *partiels*. Cette procédure doit prendre en compte la continuité des trajectoires de chaque sinusoïde, ainsi que la possibilité de *naissance* et de *mort* de ces trajectoires.

La synthèse s'obtient en général par interpolation linéaire pour l'amplitude de chaque sinusoïde. En ce qui concerne la phase, d'une trame sur l'autre le vocodeur donne les phases initiale et finale, ainsi que les fréquences initale et finale. Connaissant ces 4 paramètres, la phase est reconstruite par interpolation cubique. Cependant, la encore, la connaissance de la phase à 2π près donne un degré de liberté supplémentaire au problème. On choisit le polynôme de degré trois respectant les contraintes (à 2π près), et minimisant les variations de fréquences. Un critère de moindre carré[13] conduit en une détermination analytique des coefficients du polynôme codant la variation de la phase.

1.6.2 Synthèse additive

Le probème consiste ici en la synthèse et la superposition de sinusoïdes lentement variables en amplitude et en fréquence. Dans le cas de fenêtres d'analyse et de synthèse vérifiant la condition de reconstruction parfaite (17), l'introduction dans le système de synthèse des coefficients complexes obtenue en (29) permet de reconstruire une sinusoïde parfaite. La fenêtre d'analyse étant passe-bas, peu de coefficients dans l'expression (29) restent significativement non-nuls.

La synthèse d'une sinusoïde par FFT-inverse consiste donc :

- à choisir les coefficients $W(f^0 - k/N)$ dans une table au pré-calculée. Si la table est suffisement suréchantillonnée, et si les coefficients sont rangés habilement, la consultation de la table est immédiate.

13. Si on note φ la phase d'une sinusoïde au milieu d'une trame, alors la minimisation de la variation de fréquence suivant le critère des moindre carré revient à minimiser $\int |\ddot{\varphi}|^2$.

- à les moduler chacun par $a\exp(2j\pi mRf^0)$. La multiplication directe par $\exp(-2j\pi mRk/N)$ est inutile car se simplifiant avec un terme conjugué en entrée du bloc de transformée de Fourier discrète inverse,
- superposer-décaller le résultat.

Dans l'application de synthèse additive que l'IRCAM a développé, 7 points centrés autour de $W(f^0)$ suffisent pour générer une sinusoïde avec une bonne précision. La génération de R points demande :

- une consultation de table d'onde de 7 points alignés (quasi-immédiat),
- la génération dans un accumulateur de $a\exp(-2j\pi mRf^0)$ (peut également être tabulé puis interpolé),
- la multiplication vectorielle du premier vecteur par le contenu de l'accumulateur.

Le résultat de cette multiplication est stockée dans un tampon. La transformée de Fourier étant linéaire, on procède de même pour toute les sinusoïdes à générer. L'opération finale consiste en une transformée de Fourier inverse sur N points, et la superposition du résultat à la trame de synthèse précédente décallée.

L'intérêt de la méthode vient du nombre d'opération nécessaire pour générer une sinusoïde. Comme les opérations de transformée de Fourier (algorithme rapide en $N\log N$ opérations) et de superposition-décallage ($N-R$ additions) sont communes pour toute les sinusoïdes générées, asymptotiquement, une sinusoïde sur R points nécessite uniquement :

- une consultation de table,
- le calcul de $a\exp(-2j\pi mRf^0)$,
- 7 multiplications complexes,
- et autant d'additions,

alors que classiquement une sinusoïde compte pour :

- R multiplications,
- autant d'additions,
- R consultations de table d'onde,
- plus la gestion de l'index de la table d'onde.

Au total, le bénéfice est en faveur de la TFCT dès que le nombre de sinusoïde est supérieur à quelques unités (de l'ordre de 10 à 20, suivant le choix de N et R). Le gain en coût de calcul est incomparable quand le nombre de sinusoïdes s'accroît.

L'algorithme présenté ne permet pas facilement de faire évoluer, dans un bloc, la fréquence d'une sinusoïde. L'astuce consiste à faire évoluer par « bloc » la fréquence de la sinusoïde, en jouant sur la phase de la sinusoïde pour minimiser les effets d'interférences destructives lors de l'opération de superposition.

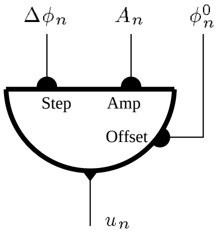


FIG. 7 – *Schéma d'un oscillateur*

1.7 Oscillateur numérique

La manipulation pratique des signaux musicaux conduit souvent à représenter les signaux dans leur zone de stabilité comme une somme pondérée de signaux harmoniques. C'est le principe de l'analyse harmonique et des premiers types de modèles de synthèse musicale, dites de synthèse additive, utilisés dès le début des années 60.

D'un point de vue pratique, cela amène les ingénieurs à étudier et à réaliser un élément fondamental : l'oscillateur. D'un point de vue externe (fig. 7), il s'agit d'un composant contrôlé par deux paramètres principaux : la période d'oscillation $\Delta\phi_n$ et l'amplitude d'oscillation A_n ainsi par un paramètre auxilliaire : la phase initiale (offset de phase) ϕ_n^0. La réalisation la plus efficace et la plus simple de ce système fait intervenir dans l'équation d'évolution du système un état interne, ϕ_n, dit de phase courante ou d'index de phase.

$$\left\{ \begin{array}{rcl} \phi_{n+1} & = & \phi_n + \Delta\phi_n \\ u_n & = & A_n \sin(\phi_n + \phi_n^0) \end{array} \right. \tag{30}$$

Si le calcul du sinus n'est pas optimisé, cette réalisation reste trop coûteuse pour être utilisable. On gagne énormément à précalculer le sinus dans une *table d'onde* et à consulter systématiquement la table d'onde au lieu de calculer explicitement la fonction sinus à chaque échantillon. Dans ce cas, les équations d'évolution du système sont décrites par l'algorithme suivant, où f est la fonction qui permet d'accéder à la table d'onde, `TableSize` est la taille de la table d'onde, `Step`, `Amp` et `Offset` sont les paramètres de contrôle et `Index` l'état interne du système.

- $\texttt{Index} \leftarrow \texttt{Index} + \texttt{Step} \bmod \texttt{TableSize}$,
- Détermination à partir de la valeur de `Index` et celle de `Offset` du défaut d'index ρ et des index entiers n_1, n_2, … n_p pour la réalisation de l'interpolation sur p points,
- Interpolation à partir de $f(n_1)$, $f(n_2)$, … $f(n_p)$ et ρ. Postmultiplication par `Amp`.

La table d'onde correspond à l'échantillonnage d'une période d'onde élementaire. Il n'est pas nécessaire de se restreindre à une table d'onde sinusoïdale. Toute forme d'onde peut a priori être utilisée. Toutefois la table d'onde est censée reproduire une période d'un signal périodique continu quelconque, donc il faut que le début et la fin de la forme d'onde coïncident afin que la continuité du

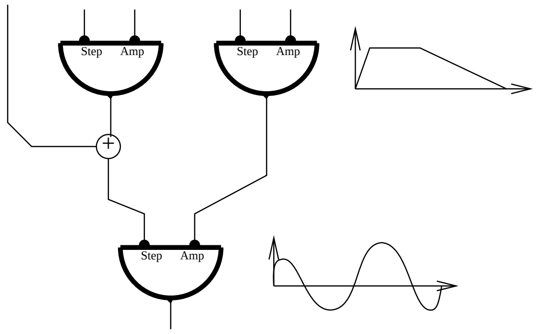


FIG. 8 – *Réseau d'oscillateurs ou patch*

signal généré soit préservée. En pratique, les formes d'ondes couramment utilisées sont les triangles, les carrés, les dents de scies, les crénaux et les sinusoïdes.

Ici l'interpolation a pour fonction de reconstruire le signal continu à partir de sa version échantillonnée, c'est à dire un rôle de conversion numérique-analogique. Les mêmes choix technologiques qu'en conversion numérique-analogique se présentent afin d'améliorer la qualité de la synthèse : soit on augmente la résolution de la synthèse en utilisant des interpolateurs d'ordre croissant (pas d'interpolation, interpolation linéaire, quadratique ou cubique), ce qui correspond exactement à des bloqueurs d'ordre croissant (0, 1, 2, 3), soit on sur-échantillonne[14] la table d'onde d'un facteur 2, 4 ou 8. Le second choix est le plus souvent préféré, parce que à performances égales moins cher à réaliser.

1.8 Réseau d'oscillateurs, notion de patch

Très rapidement l'aspect modulaire des oscillateurs a conduit les utilisateurs à considérer des réseaux d'oscillateurs dont les sorties pouvaient servir de paramètres de contrôle à d'autres oscillateurs, par exemple pour contrôler l'évolution temporelle de la fréquence fondamentale du son ou de son volume sonore (fig. 8).

Les premiers logiciels musicaux à exploiter le concept de *patch* furent dès 1960 les logiciels de la gamme **Music I** jusqu'à **Music V**. **CSound** est l'équivalent en C de **Music V**. Pendant les années 70, les synthétiseurs analogiques du commerce ont utilisé ce même concept de modularité et d'oscillateurs plurifonctionnels (VCO : Voltage Control Oscilator). Actuellement certains synthétiseurs continuent d'utiliser explicitement ce concept. Par exemple l'utilisateur dispose sur le DX7 de Yamaha 6 oscillateurs selon 16 configurations différentes. La technologie actuelle limite la taille des patches à 8 oscillateurs. En effet, il n'est pas raisonnable de limiter la polyphonie d'un synthétiseur à moins de 16 voix, ce qui implique l'allocation d'au plus $16 * 8 = 128$ oscillateurs en même temps. Pour donner un ordre d'idée, un ordinateur personnel **DX33** supporte sans trop de difficultés 8 oscillateurs à 8 kHz; quant aux nouveaux ordinateurs **MacIntosh PowerPC**, ils supportent facilement la charge de 8 oscillateurs à 44.1 kHz. On comprend que 128 oscillateurs puissent être une limite pour des architec-

14. On retrouve le sur-échantillonnage en bout de chaîne numérique sur la plupart des lecteurs CD audio-numériques.

tures spécialisées conventionnelles telles qu'on les trouve sur les synthétiseurs actuellement sur le marché.

1.9 Synthèse par modulation de fréquence

Les sons naturels sont en général très fournis en partiels (ou en harmoniques) et il convient de trouver des méthodes pour créer à partir d'un minimum d'oscillateur des sons complexes. La première solution consiste à utiliser une table d'onde constituée d'une période complexe, synthétique ou naturelle et riche en harmoniques. C'est à peu de chose près le principe des *échantillonneurs*. Les inconvénients liés à cette solution sont multiples : peu ou pas de contrôle du timbre, difficulté de réaliser soi-même des tables d'onde, difficulté de prendre en compte les bruits, impossibilité de synthétiser des signaux non périodiques (percussifs, bruités ou dissonants). Une autre solution consiste à distordre les signaux simples à travers des dispositifs non-linéaires. Une dernière solution consiste à intermoduler le contrôle de période de plusieurs oscillateurs entre-eux. Il s'agit de la modulation de fréquence.

1.9.1 Origines

La modulation de fréquence est à l'origine une technique de transmission de signal analogique qui utilise la phase d'une sinusoïde autour d'une fréquence porteuse comme support de l'information. Elle a été introduite en 1936 par E. H. Armstrong.

$$x(t) = \cos(\omega_p t \; + \; I(t)) \tag{31}$$

Au début des années 70, Chowning a le premier tenté d'utiliser et de contrôler ce type de modulation à des fins de synthèses sonores. Il a constaté que si le signal transmis était sinusoïdal, alors les caractéristiques spectrales du signal modulé étaient connues.

$$x(t) = \cos(\omega_p t \; + \; \beta \sin(\omega_m t)) = \sum_{k=-\infty}^{+\infty} J_k(\beta) \cos(\omega_p + k\omega_m) t \tag{32}$$

À partir des propriétés des fonctions de Bessel de premières espèces J_k, on déduit les propriétés suivantes :

- Le signal FM est constitué d'un partiel central à la fréquence porteuse $\frac{\omega_p}{2\pi}$, et disposés symétriquement autour de la fréquence centrale d'une infinité de partiels situés aux fréquences $\frac{\omega_p + k\omega_m}{2\pi}$. Autrement dit, il se constitue autour de la fréquence centrale un ensemble de partiels également espacés, et dont l'espacement vaut la fréquence de la modulante.
- La répartition d'énergie dans les différents partiels est entièrement déterminée par l'indice de modulation β. Autrement dit, l'enveloppe spectrale dépend uniquement de la valeur de l'indice de modulation.
- La répartition en fréquence des différents partiels est entièrement déterminée par le rapport de la fréquence de la porteuse à la modulante $\frac{\omega_p}{\omega_m}$.

- Les fonctions de Bessel J_k tendent vers 0 quand k tend vers l'infini. En conséquence, l'energie est concentrée autour de la fréquence porteuse. En particulier, pour les faibles valeurs de β, seuls les deux premières fonctions, J_0 et J_1, ne sont pas négligeable (et valent respectivement 1 et 1/2).

1.9.2 Modulation de fréquence

Quelques cas particuliers dans la synthèse par modulation de fréquence appliquée à la musique sont susceptibles de nous intéresser. Quand la fréquence de modulation est un multiple simple de la celle de la modulante ($\omega_m = r\omega_p$), le signal obtenu est la somme de partiels dont les fréquences sont toutes multiples d'une fréquence fondamentale. Donc le signal obtenu est un signal harmonique, dont la fréquence fondamentale est un sous-multiple de la fréquence de la modulante.

$$\left\{ \begin{array}{lcl} \omega_p & = & p\,\omega_0 \\ \omega_m & = & m\,\omega_0 \end{array} \right. \Rightarrow x(t) = \sum_{k=-\infty}^{+\infty} J_k(\beta)\cos(p+km)\omega_0 t \qquad (33)$$

Dans le cas où la fréquence de la porteuse est exactement le double de celle de la modulante, $m = 2p$, alors le signal est harmonique, sa fréquence fondamentale est la fréquence de la modulante, mais toutes les harmoniques paires sont absentes. Il apparaît qu'il s'agit d'une caractéristique commune à tous les sons de clarinettes. En conséquence ce paramétrage constitue un moyen relativement simple pour synthétiser des sons dont la consonnance rappelle celle des sons de clarinettes.

Dans le cas où la fréquence porteuse et la fréquence modulante ne sont pas dans un rapport simple (par exemple un rapport irrationnel ou dont le dénominateur est trop grand), le signal obtenu est inharmonique. Le degré d'inharmonicité se contrôle en modifiant faiblement le rapport de la porteuse à la modulante autour d'un rapport simple. La synthèse par modulation de fréquence apparaît donc comme une méthode très efficace pour générer et contrôler des sons plus ou moins inharmoniques tels que la plupart des sons de percussions, de cloches et même dans une certaine mesure les sons de pianos [15].

Rappelons en outre que la synthèse s'effectue dans le domaine numérique. Donc non seulement les partiels de fréquence négative se replient sur la partie positive du spectre, mais les partiels situés au-delà de la fréquence de Nyquist se replient également dans le spectre réel. En raison de cet effet d'*aliasing*, des partiels harmoniques mal atténués dans la partie haute du spectre sont susceptibles de se replier en partiels inharmoniques avec ceux déjà présents dans le spectre, et contribuer à une distortion désagréable du son.

En résumé, en synthèse par modulation de fréquence :

- l'indice de modulation β contrôle l'enveloppe spectrale du son, c'est-à-dire le timbre,
- le rapport de la porteuse à la modulante affecte la répartition spectrale des partiels et en particulier le degré d'inharmonicité,

15. Le piano est en effet un instrument faiblement inharmonique.

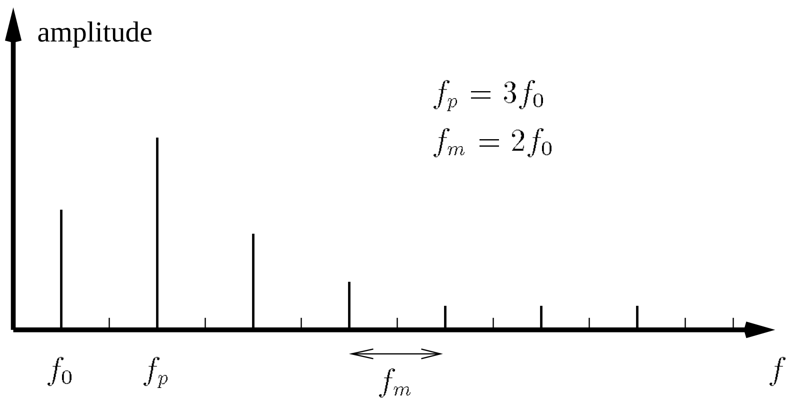


FIG. 9 – *Exemple de spectre issu d'une synthèse par modulation de fréquence*

- si les fréquences de la porteuse et de la modulante sont modifiées en proportions égales, la fréquence fondamentale du signal résultant est modifiée dans les mêmes proportions (si tant est que le signal est harmonique et possède une fréquence fondamentale).

L'inconvénient essentiel de ce type de synthèse, c'est qu'il n'existe pas à nos jours de méthodes d'analyse convaincantes qui permettent d'extraire automatiquement les paramètres de la synthèse, c'est à dire, l'indice de modulation et le rapport de la porteuse à la modulante. En conséquence, la création des sons pour la synthèse FM se fait essentiellement par expérience ou au coup par coup. Par exemple, on sait que la disparition des harmoniques impaires donne un aspect plus "timbré" au son.

1.10 Contrôle MIDI

Il est notoire que la perception des sons repose en très grande partie sur la qualité de l'attaque. Pierre Schaeffer a montré dans les années 50 qu'après suppression des premières dixièmes de seconde de l'attaque d'un son de piano, plus personne n'était capable de reconnaître le son du piano. C'est pourquoi le contrôle s'intéresse beaucoup à la qualité et au contrôle de l'enveloppe temporelle, et en particulier celle de l'attaque.

Dans le cas de la synthèse par modulation de fréquence, l'amplitude du signal et l'indice de modulation sont tout deux contrôlés par une enveloppe temporelle. Cette enveloppe se décompose traditionnellement en quatre phases (fig. 10), respectivement (en franglais) l'*attaque*, le *decay*, le *sustain*, et le *release*. L'attaque et le decay correspondent aux transitoires d'attaque du signal et sont décrits par deux temps caractéristiques : un temps de monté et un temps de descente et un niveau d'amplitude maximal. Le sustain correspond à la partie tenue (ou entretenue) de l'instrument et est caractérisé par un niveau d'amplitude. Enfin la dernière phase correspond aux résonnances libres de l'instrument et est décrit par un temps de descente.

La très grande majorité des intruments électroniques sont désormais contrôlés par l'intermédiaire d'un protocole de communication musical : le MIDI. Deux événements principaux s'échangent constamment en MIDI : NOTE-ON et NOTE-

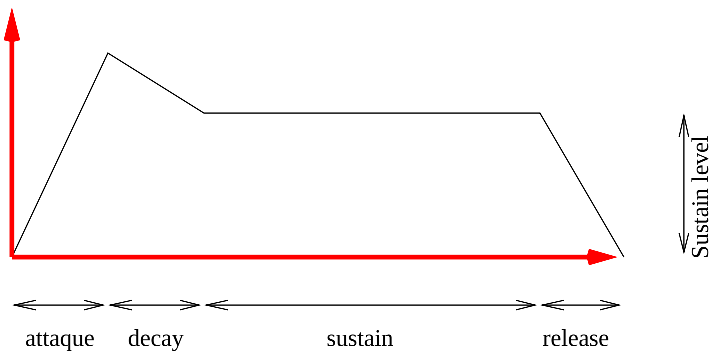


FIG. 10 – *Décomposition de l'enveloppe temporelle*

OFF. NOTE-ON signale le début de la production d'une note (sur un clavier, l'appui sur une touche). NOTE-OFF signale l'arrêt de la production d'une note (sur un clavier, le relachement d'une touche). NOTE-ON déclenche successivement les trois premières phases de l'enveloppe (i.e. l'*attaque*, le *decay* et le *sustain*). La dernière phase *release* est déclenchée par l'événement NOTE-OFF. Dans le cas des sons non entretenus, comme par exemple les sons de percussions, la partie de *sustain* est absente et la phase de *release* est déclenchée sans attendre NOTE-OFF juste après la phase de *decay*.

En modulation de fréquence, l'enveloppe temporelle de l'oscillateur associé à la porteuse contrôle effectivement les caractéristiques temporelles du son synthétique. Par contre, l'enveloppe temporelle du second oscillateur contrôle l'évolution temporelle du timbre du son synthétique. L'évolution de l'indice de modulation selon le schéma *attaque*, *decay*, *sustain*, *release* contribue à générer un spectre très riche pendant l'attaque, stable pendant la phase tenue et une phase de résonnance classique où tous les partiels (ou harmoniques) décroissent progressivement à des rythmes différents.

Il est aisé de remarquer que la synthèse par modulation de fréquence s'implémente simplement par un patch connectant deux oscillateurs l'un sur l'autre. Comme nous l'avons fait remarquer au chapitre 1.7, les tables d'ondes ne sont pas nécessairement sinusoïdales. Utiliser une table d'onde non-sinusoïdale produit en général un fort enrichissement du spectre dans les aigüs.

1.11 Synthèse par distorsion d'amplitude

L'utilisation d'une non-linéarité est une autre méthode pour générer des partiels à partir d'une sinusoïde. Il est naturel de tenter de contrôler les partiels générés en utilisant les relations de transformations trigonométriques. Les polynômes de Chebyshev T_k permettent en l'occurence de transformer les expressions trigonométriques. Ils sont définis par la relation de récurence suivante :

$$T_0 = 1,\, T_1 = X,\, \text{et} \;\; \forall k \in \mathbb{N},\; T_{k+1} - 2XT_k + T_{k-1} = 0 \qquad (34)$$

La propriété qui permet de transformer les expressions trigonométrique est la suivante :

$$\forall k \in \mathbb{N},\ T_k(\cos\theta) = \cos(k\theta) \tag{35}$$

Maintenant considérons le filtre non-linéaire qui associe à l'entrée e la sortie s par la relation $s = T_k(e)$. D'après la relation précédante, il apparaît qu'excité par une sinusoïde d'amplitude unitaire, le filtre génère une autre sinusoïde de fréquence k fois plus élevée.

Par contre évidemment la relation précédante n'est pas linéaire, et on ne peut pas la généraliser pour des sinusoïdes non unitaires. Par exemple pour le cinquième ordre, on a :

$$T_5(a\cos\theta) = a^5\cos 5\theta + 5a^3\left(a^2-1\right)\cos 3\theta + 5a\left(a^2-1\right)\left(2a^2-1\right)\cos\theta \tag{36}$$

Cette relation montre que la répartition d'énergie entre les partiels dépend essentiellement de l'amplitude du signal en entrée. Pour des faibles amplitudes, le filtre se comportement linéairement et ne génère pas de partiels. Pour des valeurs d'amplitude supérieures, les harmoniques 3 et 5 apparaissent progressivement. En conséquence cette technique apporte un contrôle relativement simple du timbre directement par l'intermédiaire du niveau d'entrée du filtre.

Il est également possible d'intermoduler deux sinusoïdes. Similairement au cas de la synthèse par modulation de fréquence, cette opération génère des partiels aux fréquences multiples entières de celles présentes dans le signal original :

$$T_k(\alpha\cos(a) + \beta\cos(b)) = \sum\nolimits_{\substack{|k_1|\le k,\\ |k_2|\le k}} K_{k_1,k_2}(\alpha,\beta)\cos(k_1 a + k_2 b) \tag{37}$$

Toutefois, dans le cas général, il est compliqué d'exprimer $K_{n_1,n_2}(\alpha,\beta)$ explicitement. Le contrôle reste complexe. De plus, il n'existe pas de méthodes convaincantes d'analyse de signaux pour ce modèle.

2 Synthèse par modèle physique

2.1 Simulations numériques

2.1.1 Motivations

La synthèse par simulation numérique est à l'origine un outil permettant de valider un modèle en comparant son comportement par rapport au comportement naturel du système mécano-acoustique étudié. Dans ce cas, il est plus judicieux de parler de simulations de systèmes physiques plutôt que de synthèses par modèle physique bien que les deux termes soient acceptés. Il est important de noter que la finalité première des simulations n'est pas de réaliser un instrument de musique virtuel, mais uniquement de valider quelques propriétés du modèle. En conséquence les contraintes de réalisation ne sont pas, dans un premier temps, ni des contraintes de contrôle, ni des contraintes de temps réel. Il s'agit d'aboutir à un modèle qui possède le plus de similitudes possibles avec le modèle naturel, afin de pouvoir identifier le modèle à la réalité physique et acoustique du système.

Différents types de conclusions peuvent être tirées des simulations numériques. Elles peuvent par exemple valider un modèle complet et complexe d'un système réel. Le modèle obtenu peut tenter d'expliquer certains comportements inattendus du système physique. Plus important, les simulations permettent de *prévoir* certains comportements du système réel. La prévision du comportement d'un système est un atout majeur dans les tâches, par exemple, de contrôle des systèmes. Les simulations permettent également de tirer des enseignements sur le soin à apporter à la fabriquation physique du système, tant dans la nature des matériaux à utiliser que dans le soin nécessaire aux opérations d'assemblage. En tout état de cause, il s'agit là d'une démarche très générale, qui tient à la fois du génie scientifique en ce qui concerne l'élaboration du modèle physique, que des sciences de l'ingénieur, en ce qui concerne l'exploitation industrielle du modèle. Ce n'est en aucun une démarche spécifique à l'acoustique, et encore moins à l'acoustique musicale.

En conséquences les techniques utilisées pour la réalisation pratique des simulations sont des techniques très générales, et très largement utilisées, à la fois dans le milieu scientifique et dans le milieu industriel. En général les outils utilisés sont très puissants et précis. En revanche ils sont très coûteux et ne répondent pas exactement aux contraintes citées en début d'exposé. Ils ne tiennent pas non plus compte des quelques degrés de liberté que nous nous accordons en raison des finalités perceptives du signal obtenu. Entre deux méthodes différentes qui génèrent deux signaux différents mais perçus identiquement par notre oreille, il est clair que nous choisirons la méthode la plus simple et la plus rapide. Toutefois, la synthèse par modèle physique complet est souvent un préalable obligatoire avant de pouvoir obtenir un modèle physique efficace, simple et convaincant.

2.1.2 Réalisation

La démarche générale consiste à considérer dans un premier temps le système d'équations différentielles ordinaires (EDO) qui régit le système. La solution algébrique d'un tel système n'est pas en général accessible. Il convient en conséquence de déduire de ce système continu un second système discret dont les solutions sont, *à peu de chose près*, l'échantillonnage (spatial ou temporel) de la solution continue du système original. On dit que l'on propose un *schéma de discrétisation* pour le système continu.

Il faut bien être conscient, au moment de l'exploitation des simulations, que le schéma de discrétisation n'est qu'un substitut au système original continu; ce n'est pas en toute rigueur l'échantillonnage de la solution continue (ou exacte) du système original. Aussi la prudence est-elle de rigueur afin de ne pas confondre les artefacts du schéma de discrétisation (typiquement les phénomènes de dispersion et de stabilité numérique) avec ceux directement liés au modèle physique.

Sans rentrer dans les détails, on peut distinguer deux grandes approches différentes parmis les schémas de discrétisation habituels : les schémas aux différences finies, et les schémas aux éléments finis.

2.1.3 Les éléments finis

Le schéma aux éléments finis (ou approche variationnelle) consiste à projeter la solution continue de l'EDO sur un ensemble de fonctions de dimension finie (par exemple l'ensemble des fonctions polynômes par morceaux de degré inférieur ou égal à n). Dans ce cas, l'erreur de calcul est évaluée par la norme fonctionnelle de la différence de la solution calculée à la solution exacte (norme au sens fonctionnelle, c'est à dire l'intégrale d'une certaine combinaison linéaire entre la valeur absolue de la fonction et de ses dérivées successives). C'est donc une approximation globale, par opposition au schéma aux différences finies qui constituent une approximation locale.

Le principe des éléments finis s'applique de la façon suivante. Dans un premier temps, une fonction d'énergie (à valeurs réelles positives) J est définie sur un espace de fonction[16] V de telle façon que la solution de l'EDO réalise le minimum de J. Ensuite, on calcule la *dérivée* de la fonction J. La dérivée de J au point u, notée $\frac{\partial J}{\partial v}(u)$, est une application linéaire de V dans $\mathbb{R}$ qui représente la linéarisation de J au point u. Ce calcul de variation donne son nom à la méthode dite variationnelle. En réalisant un minimum local, La solution de l'EDO annulle la dérivée $\frac{\partial J}{\partial v}$. Maintenant on restreint V à un sous-espace vectoriel V_p de dimension finie p (qui dépend par exemple du maillage ou de l'ordre des polynômes approximants ...). La projection v_p de la solution dans l'espace V_p annulle (dans V_p) la dérivée de J. Cela revient à chercher v_p la solution dans V_p qui vérifie $p_{/V_p}(\frac{\partial J}{\partial v}(v_p)) = 0$, ce qui aboutit en général à un système linéaire de p équations à p inconnues.

2.1.4 Exemple d'application des éléments finis

Considérons l'EDO en v suivante avec des conditions aux limites nulles :

$$\begin{cases} -v''(x) + c^2(x)v(x) = f(x), \quad 0 \leq x \leq 1, \\ v(0) = v(1) = 0 \end{cases} \tag{38}$$

Pour cette EDO, on définit la fonction d'énergie J par l'application suivante :

$$\forall v \in V,\ J(v) \stackrel{\text{déf}}{=} \frac{1}{2}a(v,v) - f(v), \tag{39}$$

avec f la forme linéaire et a la forme bilinéaire suivantes :

$$\begin{cases} a : V \times V \to \mathbb{R}, & a(u,v) = \displaystyle\int_0^1 (u'v' + c^2uv)\,dx \\ f : V \to \mathbb{R}, & f(v) = \displaystyle\int_0^1 fv\,dx \end{cases} \tag{40}$$

L'application linéaire $v \mapsto a(u,v) - f(v)$ n'est autre que la linéarisation de J au point u, i.e. la dérivée de J en u. On vérifie que la solution s de l'EDO originale annule la dérivée, c'est-à-dire :

$$\forall v \in V,\ a(s,v) - f(v) = 0 \tag{41}$$

Il suffit maintenant de restreindre l'espace V au sous-espace envisagé et de résoudre dans ce sous-espace l'équation (41) qui se ré-écrit matriciellement, une fois qu'une base a été définie :

$$\forall \vec{v} \in \mathbb{R}^p,\ \ \vec{s}^{\,t}.A.\vec{v} = \vec{f}^{\,t}.\vec{v} \tag{42}$$

16. L'espace V peut par exemple représenter l'ensemble des fonctions continues et dérivables par morceaux.

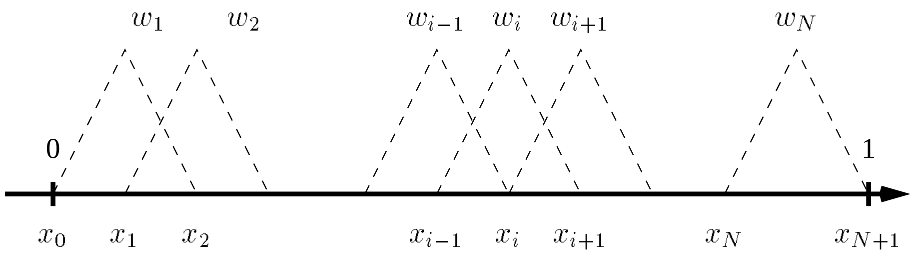


FIG. 11 – *Base de fonctions pour les éléments finis*

c'est à dire :

$$A\vec{s} = \vec{f} \tag{43}$$

On considère par exemple le maillage régulier de l'intervalle $[0, 1]$ de pas h, et l'ensemble des fonctions polynômiale de degrée 1 définit par morceaux sur le maillage (c'est-à-dire affines par morceau). On note ce sous-espace $V^{1,h}$. On choisit une base $\{(w_p)_{p\in[1,N]}\}$ de cet espace (dont on pourrait facilement obtenir une formulation explicite, voir la figure 11) de façon à avoir :

$$\forall v \in V^{1,h}, \quad v = \sum_{i=1}^{N} v(x_i) w_i \tag{44}$$

Il ne reste plus qu'a évaluer A et $\vec{f}$ dans ce système de coordonnées, c'est-à-dire, à calculer $a(w_i, w_j)$ et $f(w_i)$ pour tous les vecteurs de la base et à résoudre l'équation (43).

On suppose par exemple que c^2 et f sont constant. Dans ce cas $f(w_i) = \frac{f}{h}$ et :

$$\left\{ \begin{array}{rcl} a(w_i, w_{i+1}) & = & \frac{1}{h}(-1 + \frac{c^2}{6}h), \\ a(w_i, w_i) & = & \frac{1}{h}(2 + \frac{2c^2}{3}h), \\ a(w_i, w_j) & = & 0 \text{ sinon} \end{array} \right. \tag{45}$$

2.1.5 Les différences finies

Le schéma aux différences finies consiste à établir un *maillage* de l'espace, puis d'évaluer une approximation de la solution sur les nœuds du maillage. Il n'est pas nécessaire d'utiliser un maillage régulier de l'espace, c'est à dire qu'il est tout à fait envisageable de spécifier en certaines regions intéressantes de l'espace ou du temps une résolution plus grande (ou le contraire). Une fois le maillage établi, il est possible de quantifier l'erreur commise localement sur l'évaluation de la solution en les nœuds du maillage. La méthode repose essentiellement sur des développements de Taylor, en conséquence, l'erreur est en général majorée par la valeur absolue d'une dérivée de la solution, donc l'erreur est d'autant plus petite que la solution est lisse (autrement dit, c'est une approximation basse fréquence).

2.1.6 Exemple d'application des différences finies

Le schéma le plus simple et le plus couramment utilisé reste le schéma aux différences centrées. Par exemple pour l'EDO de l'équation (38), sur un maillage régulier de pas h, le schéma aux différences centrées correspond à substituer $v(x)$ par v_n et $v''(x)$ par $(v_{n-1} + 2v_n - v_{n+1})/h^2$. Cela correspond à faire des développements de Taylor à l'ordre 2 et à négliger les termes d'ordre supérieurs. L'EDO précédante se traduit encore par un système linéaire :

$$A.s = f \tag{46}$$

2.1.7 Limitations des méthodes numériques

Les modèles des systèmes acoustiques reposent en général sur des équations différentielles partielles (EDP) incluant des dérivées spatiales et temporelles. Il est souhaitable d'aboutir à un schéma numérique dont la solution à un instant donné se calcule en fonction de la solution aux instants précédants (i.e. un schéma explicite) par opposition au schéma implicite. Les méthodes numériques des EDP, pour un schéma explicite donné, imposent des conditions sur les pas du maillage temporel et spatial afin que l'erreur de discrétisation ne diverge pas avec le temps. Ce sont en général des conditions du même type que la relation d'incertitude d'Heisenberg, qui restreignent le raffinement simultanné de la résolution du maillage temporel δt et de celui du maillage spatial δx :

$$m \leq \frac{\delta t}{\delta x} \leq M \tag{47}$$

L'optimalité, c'est-à-dire, le cas où la dispersion numérique est la moins sensible, s'obtient en général en maximisant ce rapport $\frac{\delta t}{\delta x}$. La dispersion numérique se traduit en général par un décallage vers le grave des fréquences de résonnances de la structure simulée par rapport à la structure originale.

Il est également nécessaire d'évaluer la robustesse de l'algorithme vis-à-vis des conditions initiales et des erreurs de calculs. Toutes les variables sont représentées dans les calculateurs soit en virgule fixe soit en virgule flottante. Il découle de cette représentation des erreurs d'arrondi et des erreurs de calculs. Tant que le système n'est pas chaotique (ou du moins tant que le système discrétisé n'est pas chaotique) ces erreurs ne sont pas sensibles. La robustesse des schémas numériques est importante pour pouvoir correctement interpréter les résultats des simulations.

Il est également important de noter qu'il n'existe pas à proprement parler de méthode d'analyse pour de tels modèles physiques, c'est à dire que connaissant la structure du modèle, il est difficile à partir des grandeurs accessibles à la mesure de remonter aux paramètres qui ne sont pas accessibles à la mesure. Autrement dit, il n'est pas possible d'extrapoler systématiquement à partir des observations des paramètres de contrôle qui conduisent à une synthèse aussi proche que voulue de l'original.

Enfin, la structure du modèle discret ne conduit en général à aucune symmétrie particulièrement remarquable. En conséquence, si la vectorisation du code est toujours possible (en associant un processeur de calcul à chaque région de l'espace simulé), la factorisation du code ainsi que la mise au point d'algorithmes rapides ne sont pas toujours aisées, et sont même dans la plupart des cas impossibles. La charge de calcul dépend en principe de la résolution du maillage spatial, mais comme la résolution du maillage temporel (la fréquence d'échantillonnage) dépend également linéairement de la résolution du maillage spatial, la charge de calcul requise pour synthétiser une seconde de son est proportionnelle au carré du nombre de maille pour les modèles monodimensionnels[17] (cordes, tubes ...). Cela conduit à des modèles numériques assez lents et pas inutilisables en condition de temps réels.

2.1.8 Exemple de restriction d'utilisation

Prenons par exemple la discrétisation de l'EDP correspondant à l'équation des ondes par un schéma aux différences finies centrées (les conditions initiales sont mises de coté afin de simplifier la présentation) :

$$\frac{\partial^2 y}{\partial t^2} - c^2 \frac{\partial^2 y}{\partial x^2} = f(x,t) \tag{48}$$

Le schéma aux différences finies centrées appliqué à cette équation aboutit aux substitutions suivantes (n représente l'index de temps et k représente l'index d'espace) :

$$\left\{ \begin{array}{ccc} C & = & c\frac{\delta t}{\delta x} \\ S^2 & = & 1 - C^2 \\ y(k\delta x, n\delta t) & \rightarrow & y_n^k \\ f(k\delta x, n\delta t) & \rightarrow & f_n^k(\delta t)^{-2} \\ \dfrac{\partial^2 y}{\partial x^2} & \rightarrow & (y_n^{k-1} - 2y_n^k + y_n^{k+1})(\delta x)^{-2} \\ \dfrac{\partial^2 y}{\partial t^2} & \rightarrow & (y_{n-1}^k - 2y_n^k + y_{n+1}^k)(\delta t)^{-2} \end{array} \right. \tag{49}$$

$$\text{et donc : } y_{n+1}^k = 2S^2 y_n^k + C^2(y_n^{k+1} + y_n^{k-1}) - y_{n-1}^k + f_n^k \tag{50}$$

Si on note, Y_n le vecteur $(y_n^0, y_n^1, \ldots, y_n^K)$, alors l'équation de récurence précédante s'écrit :

$$\begin{bmatrix} Y_{n+1} \\ Y_n \end{bmatrix} = \Gamma \begin{bmatrix} Y_n \\ Y_{n-1} \end{bmatrix} + \begin{bmatrix} F_n \\ 0 \end{bmatrix} \tag{51}$$

$$A = \begin{pmatrix} 2S^2 & C^2 & & 0 \\ C^2 & 2S^2 & \ddots & \\ & \ddots & \ddots & C^2 \\ 0 & & C^2 & 2S^2 \end{pmatrix}, \ \Gamma = \begin{pmatrix} A & -\mathrm{Id} \\ \mathrm{Id} & 0 \end{pmatrix} \tag{52}$$

La non-divergence de cette équation de récurence est assurée par l'inclusion des valeurs propres de la matrice Γ à l'intérieur du cercle unité. La condition $c\delta t \leq \delta x$ est suffisante pour assurer la stabilité numérique du modèle discret.

17. Pour des modèles bidimensionnels tels que ceux nécessaires pour représenter des membranes, des tables d'harmonie etc. la charge de calcul requise pour synthétiser une seconde de son est proportionnelle au produit des résolution spatiales et de la résolution temporelle

2.1.9 Simulations de la vibration d'une corde avec raideur

Comme cité dans un chapitre précédant, les simulations numériques tirent leur source directement de l'étude de la physique du système et proposent des schémas numériques adaptés aux situations.

Prenons par exemple le cas de la simulation de la vibration d'une corde avec terme de raideur. L'EDP qui régit le système est :

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2} + l_0^2 c^2 \frac{\partial^4 y}{\partial x^4}$$

Les conditions initiales et les conditions aux limites de cette EDP (qui correspondent à la situation d'une corde encastrée) sont données par :

$$\begin{aligned} y(x,0) = f(x);\ \frac{\partial y}{\partial t}(x,0) &= 0 \\ y(0,t) = y(L,t) &= 0 \\ \frac{\partial y}{\partial t}(0,t) = \frac{\partial y}{\partial t}(L,t) &= 0 \end{aligned}$$

L'utilisation d'un schéma aux différences finies pour cette équation conduit à l'équation de récurence suivante, en posant $m^2 = (l_0 N/L)$ et $r = c\frac{\partial t}{\partial x}$ où N est le nombre de point du maillage spatial régulier :

$$y_{n+1}^{i} = 2(1 - r^2 - 3m^2 r^2) y_n^i - y_{n-1}^i + r^2(y_n^{i+1} + y_n^{i-1}) - m^2(y_n^{i+2} + y_n^{i-2} - 4y_n^{i+1} - 4y_n^{i-1})$$

L'optimalité vis à vis de la dispertion numérique s'obtient en prenant $r = \frac{1}{\sqrt{1+4m^2}}$. Toutefois la dispersion numérique liée à ce schéma de discrétisation reste audible. Il est en fait nécessaire d'utiliser un autre schéma de discrétisation, dit implicite, beaucoup moins pratique à utiliser afin que la dispersion numérique ne se manifeste qu'à un niveau inaudible à nos oreilles.

2.2 Synthèse modale

Les modèles de synthèse dites de synthèse modale se situent à mi-chemin entre les modèles purement physiques et les modèles de signaux. L'IRCAM développe un environnement de synthèse modale baptisé Modalys. Les objets sonores sont représentés par une collection de sous-structures interagissantes et vibrantes. Chacune d'entre-elle est caractérisée par un jeu de données modales et représente des systèmes acoustiques tels que des chevalets, des caisses de résonnance, des tubes acoustiques, des cloches, des membranes de timbales, des cordes, etc.

L'intérêt principal de cette méthodologie vient de la simplicité de la représentation des structures et à la diversité des méthodes d'analyse qui sont disponibles. En effet, les données modales qui caractérisent les sous-structures peuvent provenir soit de fonctions intégrées qui calculent analytiquement les informations modales à partir de caractéristiques géométriques d'objets simples (cordes, membranes, tubes acoustiques ...), soit de mesures expérimentales sur des objets physiques, soit d'une quelconque autre source pour peu que le jeu de données modales ait un sens acoustique.

$$F_1 = F_2 = f(X_1, X_2)$$

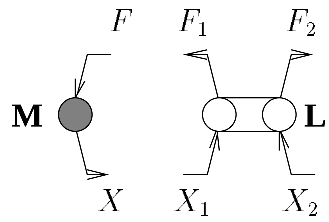


FIG. 12 – *Élément principaux du langage Cordis-Anima*

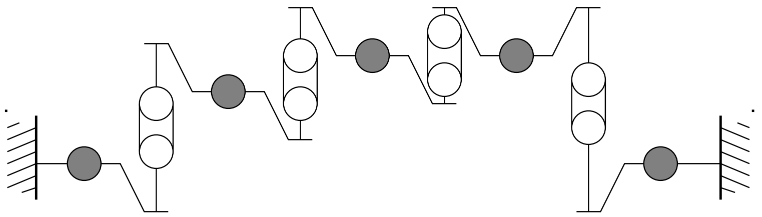


FIG. 13 – *Réseau représentant une corde sous Cordis-Anima*

Le couplage des structures s'obtient à partir de sous programmes d'interactions particuliers (par exemple : A-frappe-B, A-pince-B, A-colle-B ...). Le son est obtenu en plaçant virtuellement des micros sur une des structures vibrantes.

2.3 Modèle virtuel

Le système de synthèse très particulier développé à actuellement à l'ACROE, à Grenoble sous la direction de Claude Cadoz, connu sous le nom de Cordis-Anima, se situe également à mi-chemin entre les simulations numériques purement physiques et les modèles de signaux. Ce système se veut être l'analogue des modèles des guides d'onde pour des cellules représentant la loi fondamentale de la dynamique et les principaux types de forces et de frottements. Chaque cellule correspond soit à un point (une masse ponctuelle), soit à une liaison entre deux points (fig. 12). Chaque cellule correspond également à un algorithme. Pour la masse ponctuelle, l'algorithme utilisé est le le suivant : $X_n = 2X_{n-1} - X_{n-2} + \frac{F_{n-1}}{M}$; pour un ressort : $F_n^2 = -F_n^1 = -K(X_n^2 - X_n^1)$; pour un frottement visqueux : $F_n^2 = -F_n^1 = -Z(X_n^2 - X_{n-1}^2 - X_n^1 + X_{n-1}^1)$.

Sans rentrer dans les détails, le mouvement d'une corde serait dans ces conditions reproduit par le réseau proposé en figure 13. La raideur de la corde peut être introduite dans le modèle en superposant de nouveaux éléments de liaison entre les éléments de masses.

2.4 Modèles de guide d'onde

2.4.1 Guide d'onde acoustique sans perte

Contrairement aux modèles précédants où la démarche consistait à approcher au mieux des solutions inaccessibles au calcul algébrique et exactes du système

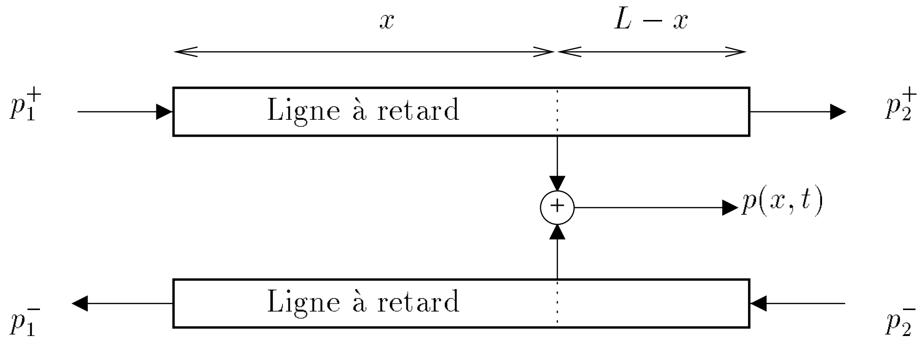


FIG. 14 – *Modèle de guide d'onde sans perte*

complet, les modèles de guides d'onde discrétisent directement la solution exacte dans le cas de modèles simples. Les modèles plus complexes sont fabriqués par concaténation ou par raffinement successif du modèle de guide d'onde original. On doit la théorie des guides d'onde, appliquée à la synthèse des instruments de musique au professeur J.O. Smith, de Stanford au début des années 1980. Les nouveaux synthétiseurs du commerce dits instruments virtuels (VL1 de Yamaha, Korg Prophecy, Technics WSA1 ...) sont en fait des réalisations industrielles basées sur la théorie des guides d'onde.

Le modèle acoustique le plus simple dont les solutions exactes sont facilement accessibles correspond à la propagation du premier mode transverse d'une onde acoustique dans un tube monodimensionnel (guide d'onde acoustique) sans perte. Le système est régi par l'équation des ondes (48). Dans un tel système, il est toujours possible de décomposer la solution par la superposition d'une onde dite aller et d'une autre onde dite retour. Si les réflexions terminales ne sont pas prises en compte, le système se comporte comme un retard pur pour chacune des ondes aller et retour. Le retard correspond au temps nécessaire à un paquet d'onde pour se propager d'un bout à l'autre du guide d'onde à la célérité c. L'implémentation pratique de ce modèle correspond au schéma de la figure 14. La prise en compte des conditions aux limites correspond à introduire des coefficients des réflexions et de transmittances aux extrémités du guide d'onde. L'implémentation des conditions aux limites se fait conformément à la figure 15 avec le coefficient R de réflexion et T de transmittance. D'un point de vue numérique, quand le retard introduit par le tube est proportionnel à la période d'échantillonnage des signaux alors la solution calculée est exactement égale à la discrétisation de la solution originale.

Physiquement, les coefficients de réflexions sont toujours inférieurs en module à 1. Dans le cas contraire, c'est à dire si l'onde réfléchie est amplifiée à chaque réflexion, cela signifie que les conditions aux limites rajoutent de l'énergie au système au lieu d'en dissiper comme c'est le cas habituellement. Numériquement, cela correspond à utiliser un système qui est susceptible de diverger.

2.4.2 Systèmes avec pertes

La prise en compte de modèles complets, incluant par exemple des pertes visco-thermiques ou le rayonnement dans l'espace du champs acoustique par le pavillon, n'aboutit pas à une solution analytique explicite. Toutefois l'utilisation

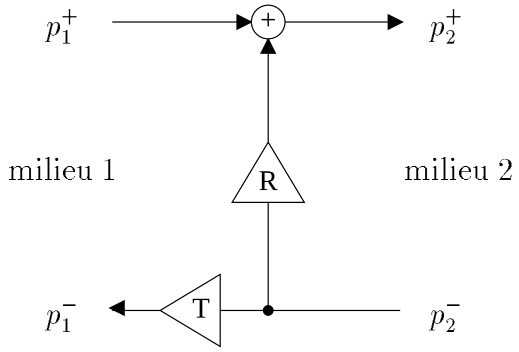


FIG. 15 – *Modèle de conditions aux limites, vues du milieu 2*

de transformées de Fourier et des développement limités permettent d'obtenir des expressions linéaires. Donc au premier ordre, tout s'interprète comme des interconnections de guides d'onde sans pertes, et au second ordre, par des filtres dynamiques, c'est-à-dire par des termes qui dépendent de la fréquence. Or il est connu que la multiplication par un terme dependant de la fréquence dans le domaine spectral (de Fourier) correspond à une convolution dans le domaine temporel, c'est à dire, en terme de traitement de signal à un filtrage.

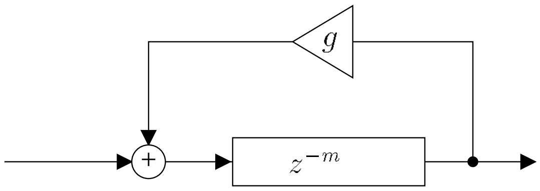
Les modèles physiques traitées en traitement de signal, sont en fait des modèles qui reproduisent la structure fondamentale du système acoustique en terme de propagation d'ondes, et qui approximent le système physique dans ses détails fréquenciels.

2.4.3 Modèle de Karplus-Strong

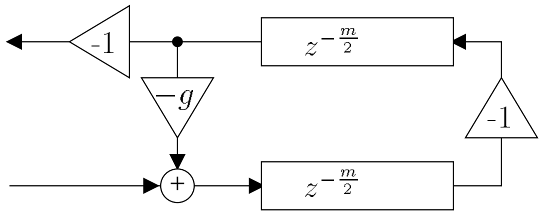
Le modèle de Karplus-Strong est une implémentation particulière des guides d'ondes sans pertes. Cela correspond à observer l'onde retour dans un guide d'onde acoustique de longueur $cT_e\frac{m}{2}$ avec comme conditions aux limites une réflexion totale à une extrémité du guide, et un coefficient de réflexion g à l'autre extrémité du guide (voir figure 16).

Ce modèle fut utilisé à l'origine pour simuler le fonctionnement d'une corde. En effet, en négligeant la constante de raideur de la corde, en ne considérant que le mode de transmission transverse des ondes (c'est-à-dire en négligeant les ondes de pression qui se propagent longitudinalement dans la corde, ainsi que le couplage des deux modes qui se propagent sur le plan transverse de la corde), on aboutit au même système d'équations que celui qui régit la propagation des ondes acoustiques dans le milieu aérien, à l'exception de la célérité de l'onde qui dépend à présent de la tension de la corde. Les conditions aux limites correspondent pour une extrémité à une réflexion totale c'est à dire à une fixation rigide de la corde sur le sillet et pour l'autre extrémité ($g \approx 1$) à une réflexion presque totale sur le chevalet qui n'est pas complètement rigide et qui transmet une partie de la vibration à la table d'harmonie (et réciproquement).

L'utilisation pratique du modèle consiste à initialiser les deux lignes à retard en fonction des conditions initiales de position et de vélocité de la corde. Pour une corde pincée, la position initiale est une position sans vélocité initiale. Cela correspond sur le modèle éclaté à des valeurs initiales dans les deux lignes à



modèle compact



modèle éclaté

FIG. 16 – *Modèle de Karplus-Strong*

retard égales. Dans le modèle compact, cela correspond à une initialisation antisymétrique de la ligne à retard. Pour les cordes frappées, la positions initiale est la position au repos. Donc les valeurs initiales des deux lignes à retard du modèle éclaté sont opposées. Cela correspond à une configuration symétrique pour les valeurs initiales de la ligne à retard du modèle compact.

Dans les deux cas, l'entrée du système reste habituellement à zéro. Cela correspond physiquement à un découplage, artificiel, de la table d'harmonie et du chevalet avec la corde. Physiquement, le couplage entre la corde et le système résonateur est très faible et donc négligeable.

D'un point de vue numérique le système s'apparente à une structure bien connue dit de filtre en peigne. La fonction de transfert entrée-sortie de ce système est $H(z) = \frac{z^{-m}}{1-gz^{-m}}$.

$$\left|H(e^{-j\omega})\right|^{-2} = 1 + g^2 - 2g\cos(m\omega) \tag{53}$$

La réponse impulsionnelle du filtre en peigne est un peigne de Dirac : $\forall k \geq 0$, $h_{m(k+1)} = g^k$, zéro ailleurs. La réponse en fréquence présente $m/2$ sommets situés en rapports harmoniques, répartis sur tout le spectre, qui correspondent à m pôles (complexes conjugués) répartis régulièrement sur le cercle de rayon $\sqrt[m]{g}$. Ce filtre atténue essentiellement toutes les fréquences, exceptées celles multiples entières de $\frac{F_e}{m}$ (F_e : fréquence d'échantillonnage) qui ne sont que faiblement atténuées.

Une analyse modale du système physique correspondant aboutit aux mêmes conclusions. Le système physique dispose d'une infinité de modes, de fréquences multiples de la fréquence fondamentale $\frac{c}{2l}$. La limitation du nombre de mode dans le modèle numérique correspond à l'hypothèse de limitation de la bande du signal d'entrée : le signal en entrée est supposé être à bande limitée et donc ne possède pas d'énergie au-dela de la fréquence de Nyquist (i.e. la demi-fréquence d'échantillonnage), donc les modes supérieurs du système ne sont pas excités, donc il n'est pas utile de les représenter dans le modèle numérique.

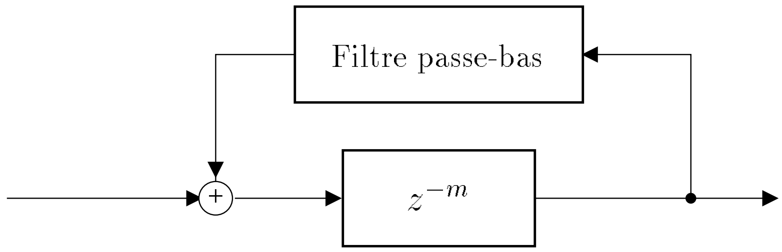


FIG. 17 – *Modèle complet de Karplus-Strong*

Pour comprendre les conséquences de l'initialisation de l'état de la ligne à retard, il suffit de constater que tout se passe comme si aux instants $-m, -m+1, \ldots, -1$, les valeurs initiales de la ligne à retard étaient séquentiellement entrées dans le filtre. Donc l'observation du déplacement du chevalet correspond exactement à ce que l'on obtiendrait en sortie du filtre s'il était excité, sur une période de m échantillons par un signal correspondant au déplacement initial (et à la vélocité initiale) du système. Cela correspond donc à exciter certains modes du filtre.

2.4.4 Limitations et amélioration du modèle

Dans un premier temps, il faut noter qu'il reste idéaliste que de supposer que les conditions aux limites n'introduisent pas de pertes dans le système. En général, les pertes qu'introduisent les conditions aux limites se manifestent à haute fréquence. En effet, les mesures expérimentales de coefficients de réflexion (en fait d'impédance mécanique) sur le rayonnement des instruments à vent ou le mouvement du chevalet pour les instruments à corde présentent toujours le même profil passe-bas. Cela conduit à remplacer dans le modèle (voir figure 17) le coefficient g par un filtre passe-bas. Le système numérique, dissipatif à autre fréquence, le plus simple à obtenir à partir de ce modèle s'obtient en utilisant le plus simple des filtres passe-bas : $\frac{1+z^{-1}}{2}$. On obtient alors le modèle complet de Karplus-Strong. Ce modèle très simple permet d'obtenir des synthèses de sons de cordes très convaincant.

Il n'est pas toujours possible de pouvoir supposer que les pertes aux limites se manifestent forcément comme le filtre $\frac{1+z^{-1}}{2}$. Pour obtenir un système plus proche de la réalité, il faut faire des mesures acoustiques des coefficients de réflexion sur un instrument et en déduire le profil du filtre passe-bas requis.

Dans tous les cas, le rajout d'un filtre passe-bas dans la boucle de rétro-action a pour effet d'atténuer tout à la fois le gain et la fréquence des résonnances[18] (fig. 18). On constate alors une certaine inharmonicité dans la répartition des modes de résonnance du système. Cette inharmonicité correspond physiquement à l'effet de la réactivité du chevalet (ou du pavillon) à haute fréquence.

Il existe en général d'autre sources de pertes, en particulier des pertes réparties sur toute la géométrie du système oscillant. Par exemple dans les cordes, les pertes se manifestent également dans le terme de raideur qui empèche la corde de se plier indéfiniment; dans les tubes acoustiques, les pertes se manifestent

18. On entend ici que les pôles du filtre se trouvent tous d'une part déplacés vers l'intérieur du cercle unité, d'autre part décalés d'un angle qui ramene le pôle vers fréquences plus graves.

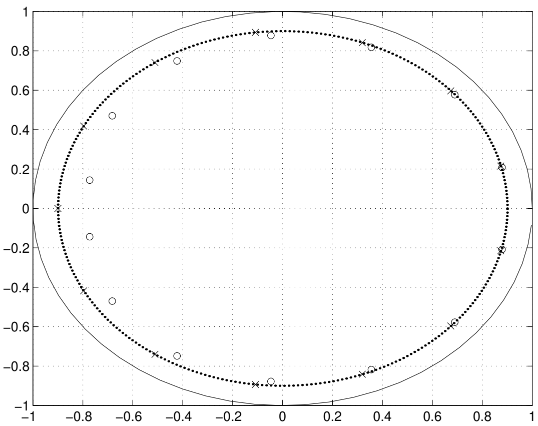


FIG. 18 – *Déplacement des pôles*

sous formes viscothermiques, qui traduisent des effets de frottement visqueux et d'échange thermique important entre les couches du fluide en oscillation dans le tube et les parois rigides et froides du tube (voir chapitre 2.4.8). Là également, les pertes influencent la propagation des ondes essentiellement à haute fréquence. Il est très couteux en temps de calcul de tenir compte du caractère réparti des pertes dans un modèle de guide d'onde, puisque cela revient à répartir m filtres à l'intérieur de la ligne à retard z^{-m}. En approximation faibles pertes, il est toutefois possible de factoriser tous les termes de perte répartis sur tout le guide d'onde en un seul terme de perte localisé en bout de guide d'onde. Dans ce cas, cela revient à modifier le filtre passe-bas évoqué précédemment par un nouveau terme tenant effectivement compte des pertes réparties.

En résumé, le choix habile du filtre passe-bas dans la boucle de rétro-action permet de reconstituer l'effet des pertes dans le guide d'onde et de simuler ainsi l'amortissement plus rapide des partiels de haute fréquence par rapport à ceux de basses fréquence que l'on observe dans la plupart des système acoustique en régime libre.

La seconde limitation du modèle relève de la quantification de la longueur de la ligne à retard. En effet, une fois la fréquence d'échantillonnage choisie dont le choix est en général imposé par des critères technologiques (32kHz, 44.1kHz ou 48kHz), la longueur du tubes (ou de la corde) est quantifiée. Par exemple à 32kHz, le pas d'échantillonnage spatial pour un tube dans les conditions standarts de pression et de température est de l'ordre du centimètre. C'est l'expression d'une relation d'incertitude. Ce niveau de précision est incompatible avec les exigences des facteurs d'instrument ainsi qu'avec la précision de notre perception auditive. Pour palier cet inconvéniant, il est nécessaire d'implémenter numériquement des dispositifs interpolateurs qui permettent de simuler des *retards fractionnaires*.

2.4.5 Retards fractionnaires

Nous ne rentrerons pas dans les détails d'implémentation des retards fractionnaires. Dans les grandes lignes, pour un retard réel d, cela consiste à réaliser un filtre dont la réponse en fréquence est idéalement $H_d(e^{j\omega}) = e^{-jd\omega}$.

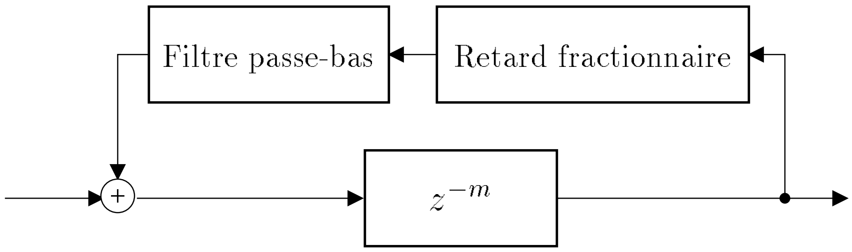


FIG. 19 – *Modèle étendu de Karplus-Strong*

Parmi toutes les directions qui s'offrent à nous pour approximer ce filtre, nous n'en retiendrons que deux : les filtres interpolateurs de Lagrange et les filtres passes-tout. Comme il est inacceptable pour un modèle physique que le retard introduise de l'énergie dans le système, le gain en fréquence du filtre doit constamment être inférieur au gain unitaire.

Les filtres interpolateurs de Lagrange dérivent des fameux polynômes interpolateur de Lagrange. Par exemple le filtre passe-bas de transformée en z $\frac{1+z^{-1}}{2}$ correspond au filtre interpolateur de Lagrange d'ordre 1 réalisant le retard d'un demi-échantillon. En effet, ce filtre réalise la demi-somme de deux échantillons successifs à chaque échantillon, donc on conçoit facilement que ce filtre réalise l'interpolation linéaire centrée entre deux échantillons, ce qui correspond à retarder l'ensemble du signal d'un demi-échantillon. Une interpolation quadratique entre 3 échantillons conduirait à un filtre linéaire d'ordre 2. Les propriétés de ces filtres sont les suivantes :

- c'est un filtre à réponse impulsionnelle finie, donc les transitoires sont très courts et peu audibles,
- ce filtre donne lieu à un implémentation stable et efficace,
- il est possible de modifier en cours d'utilisation le retard,
- le filtrage réalisé est passe-bas, c'est à dire que l'interpolation est optimale à basse fréquence,
- l'optimisation est mauvaise à haute fréquence, quelque soit l'ordre choisi,
- le retard obtenu est de moins bonne qualité qu'avec les filtres passe-tout.

Les retards fractionnaires passe-tout découlent de la réflexion suivante : le filtre idéal est passe-tout, donc il est judicieux d'approximer ce filtre idéal dans l'espace des filtres passe-tout. De plus, les filtres passe-tout sont facilement caractérisables puisque leur transformée en z est toujours de la forme $z^{-n}P(z)/P(z^{-1})$, où P est un polynôme d'ordre n. Par exemple pour l'ordre 1, on obtient :

$$H^d_{\mathrm{ap}}(z) = \frac{a + z^{-1}}{1 + az^{-1}} \text{, avec } a = \frac{1-d}{1+d} \quad \Rightarrow \quad H^d_{\mathrm{ap}}(z) = \frac{(1-d) + (1+d)z^{-1}}{(1+d) + (1-d)z^{-1}}$$

Les propriétés remarquables de ces filtres sont les suivantes :

- implémention très efficace,

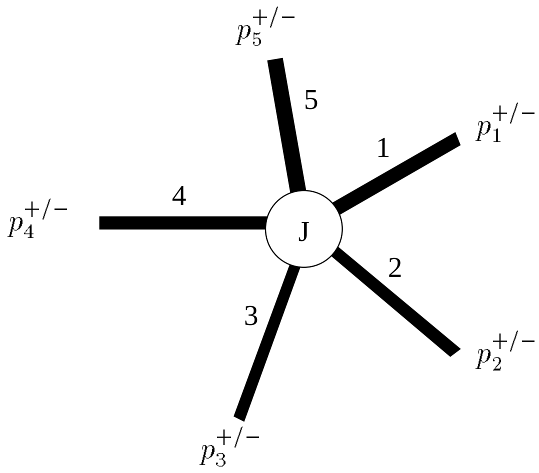


FIG. 20 – *Jonction de n guides d'onde*

- retard de grande qualité,
- difficulté pour faire varier la valeur du retard d en cours d'utilisation sans provoquer d'artefacts,
- le retard ne peut varier que dans un intervalle restreint de valeurs.

En résumé, dans le cas où le retard (la longueur du guide d'onde) doit être précis mais invariant temporellement, les filtres passe-tout constituent une solution idéale. Par contre, dans le cas où le retard est assujéti à varier dans le temps, alors les filtres interpolateurs de Lagrange restent un compromis très acceptable. Dans tous les cas, le modèle qui tient compte de toutes les améliorations présentées précédemment est celui présenté en figure 19.

2.4.6 Jonctions de N guides d'onde

Le problème consiste à trouver une représentation adéquate afin de coupler plusieurs guides d'onde entre eux (voir figure 20). On attend d'une représentation acceptable qu'elle fournisse des principes d'assemblage similaire aux fameuses lois de Kirchoff en électricité.

Nous avons vu au chapitre précédent comment les conditions aux limites pouvaient s'exprimer en terme de coefficients de réflexions et coefficients de transmissions (fig. 15). Une telle analyse peut être menée du point de vue du milieu 2; elle peut être également menée à partir du milieu 1. Au total, la jonction de deux guides d'ondes se caractérise par une liste de coefficients de transmission et de réflexions correspondant aux interfaces entre chaque milieu (i.e. entre chaque guide d'onde). L'extrapolation du raisonnement au cas de la jonction de n guides d'onde conduit à une matrice $n \times n$ dont le terme (i, j) correspond au coefficient de transmission du milieu i dans le milieu j (et le terme diagonal (i, i) correspond au coefficient de réflexion du milieu i sur lui-même). La formulation explicite de la matrice dépend de la nature des signaux utilisés.

Les guides d'ondes permettent de représenter tout système physique dont une des grandeurs physiques est régie par l'équation des ondes. On a vu que l'élongation longitudinale d'une corde répondait en première approximation à cette équation. C'est également le cas des signaux de débit acoustique et de pression acoustique pour la propagation des ondes acoustiques planes dans un guide onde acoustique. De façon plus étonnante, le signal de puissance $\sqrt{pu}$ satisfait également cette équation.

Dans le cas des guides d'ondes acoustiques, les lois qui permettent d'obtenir les équations qui régissent l'interface de n guides d'onde sont celles qui concernent le débit acoustique et la pression acoustique :

$$\begin{aligned} \forall i \in [1,n], \quad P = p_i^+ + p_i^- \\ \sum_{i=1}^{n} (u_i^+ + u_i^-) = 0 \end{aligned} \tag{54}$$

Ces équations, dans le cas des jonctions de deux guides d'ondes conduisent aux expressions suivantes en notant $r = \frac{R_2^2 - R_1^2}{R_2^2 + R_1^2}$, coefficient de réflexion partielle (R_1 et R_2 correspondent respectivement aux rayons des tubes 1 et 2) :

$$\begin{aligned} \begin{bmatrix} p_1^- \\ p_2^+ \end{bmatrix} = \begin{pmatrix} -r & 1+r \\ 1-r & r \end{pmatrix} \begin{bmatrix} p_1^+ \\ p_2^- \end{bmatrix} \\ \begin{bmatrix} u_1^- \\ u_2^+ \end{bmatrix} = \begin{pmatrix} -r & 1-r \\ 1+r & r \end{pmatrix} \begin{bmatrix} u_1^+ \\ u_2^- \end{bmatrix} \end{aligned} \tag{55}$$

Il est facile, quoique fastidieux, de calculer l'expression de la jonction de n guides d'onde acoustiques. Il apparaît que la jonction est alors caractérisée par $n-1$ coefficient de réflexion partielle.

Il est important de noter que les lois utilisées (eq. 54) correspondent à des jonctions sans pertes, c'est-à-dire que la jonction est neutre au niveau du bilan énergétique. En conséquence, les algorithmes de jonction ne peuvent pas introduire d'instabilités numériques qui conduisent les calculs à diverger exponentiellement. Il est possible au prix d'un complexité très grandement accrue d'introduire des termes de perte dans les jonctions connectant plus de deux guides d'onde. Par contre, il faut prendre garde de ne pas introduire dans le système des jonctions réactives qui ré-introduisent artificiellement de l'énergie dans le système, et peuvent conduire les algorithmes à exploser numériquement.

2.4.7 Liens avec les filtres en treillis

Les chapitres précédants ont montré que les guides d'ondes était une alternative efficace pour simuler les systèmes régis par l'équation des ondes. De plus, ce type de synthèse s'insère très facilement dans la théorie du traitement du signal, parque les réseaux linéaires de guides d'ondes peuvent être représentés par des filtres en treillis (fig. 21). Or les filtres en treillis se prètent fort bien à l'analyse auto-régressive avec l'algorithme de Levinson.

Structurellement, un treilli élémentaire (figure 21) correspond à un guide d'onde élémentaire, avec une condition aux limites. Le passage du guide d'onde au treilli se fait par une opération simple appelée normalisation. La normalisation dépend de la nature des signaux véhiculés dans le guide d'onde, mais il existe toujours une normalisation qui permet de passer d'un guide d'onde (sans pertes) à un filtre en treillis et réciproquement. Cette correspondance entre les

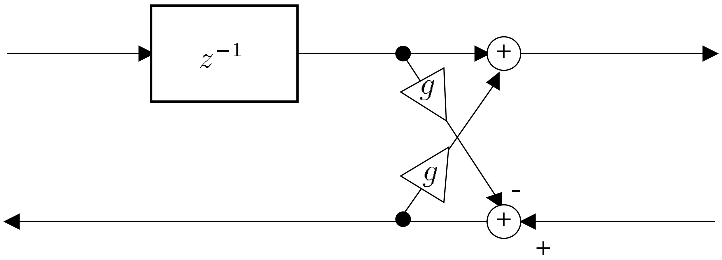


FIG. 21 – *Treilli élémentaire*

deux modèles à été développée dans les premières études sur le traitement de la parole (1970). Elle permet entre autre d'interpréter le résultat de l'analyse auto-régressive d'un signal de parole comme une modélisation du conduit vocal par un guide d'onde cylindrique de rayon variable.

2.4.8 Pertes viscothermiques

Nous avons également vu dans le chapitre précédant que l'introduction de filtres passe-bas permettait de prendre en compte le caractère résistif de certains éléments acoustiques. Il est également possible de prendre en compte l'aspect dispersif de la propagation des ondes dans le guide d'onde en répartissant des filtres passe-bas dans la propagation du signal, c'est à dire dans la ligne à retard. Pratiquement les pertes à prendre en comptes dans un instrument à vent sont les pertes visco-thermiques. Elles interviennent essentiellement dans des longs tubes[19] du type tuyau d'orgue. Dans ce type de tubes dit *long*, il n'est plus possible de faire une approximation faible pertes. L'intensité des pertes visco-thermiques est proportionnelle à $f^{\frac{1}{2}}$, où f est la fréquence. En conséquence, il faut introduire un filtre numérique dont le gain s'atténue de 3dB par octave (la moitié du classique 6dB par octave), ce qui n'est pas standart. Donc pour introduire les pertes viscothermiques dans les guides d'onde, il faut répartir des filtres optimisés dans un certaine intervalle de fréquence, afin que le résultat de la synthèse soit acceptable dans le même intervalle de fréquence.

2.4.9 Synthèse de cordes de piano

Le professeur J.O. Smith a présenté en 1995 (ISMA95) un modèle très convaincant de piano. Chaque corde est représentée par un guide d'onde. En fonction de l'octave, chaque note est simulée par une, deux ou trois cordes couplées, excitées ensemble par la frappe impulsive d'un marteau. L'ensemble des cordes est couplée à la table d'harmonie. La synthèse se réalise en temps-réel sur une station de travail NeXT pilotée par un clavier MIDI.

2.4.10 Modèle d'excitateurs

La modélisation complète d'un instrument de musique doit prendre en compte à la fois le système de propagation de l'onde acoustique, et aussi le phénomène

19. Un tube est dit *long* si le rapport de sa longueur par sa largeur est grand.

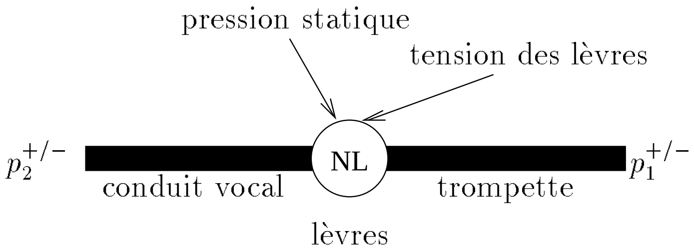


FIG. 22 – *Modèle non linéaire de trompette*

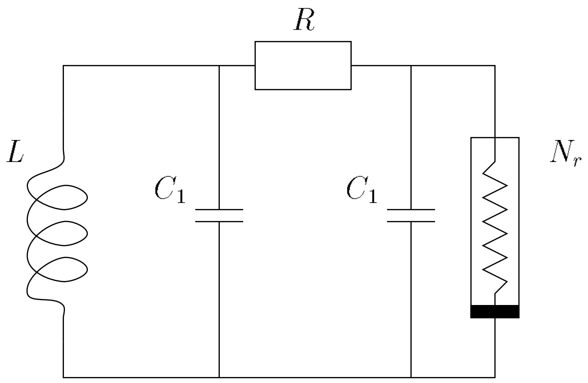


FIG. 23 – *Circuit de Chua*

qui est à l'origine de la production du son, c'est-à-dire, soit l'impulsion initiale (corde frappée, corde pincée, membrane frappée), soit le mécanisme auto-oscillant (corde frottée, anche simple, anche double), qui excite l'ensemble du système. Le système complet se décompose dans la grande majorité des cas suivant un ou plusieurs guides d'onde représentant les systèmes propagatifs et résonnants de l'instrument, couplés entre eux et avec un système non lin;eaire régissant les phénomènes transitoires et auto-oscillants. Tous les mécanismes entretenus, tels que le mouvement de la corde frottée par un archet de violon, ou bien le mouvement des cordes vocales, apparaissent alors comme des cycles limites, des attarcteurs étranges de systèmes non lineaires complexes.

L'étude des phénomènes auto-oscillants est loin d'être simple, et les phénomènes sont également loin d'être bien compris. Il est très facile d'obtenir des situations chaotiques, c'est-à-dire dans des situations, en général peu recherchées qui correspondent aux "canards" de l'instrumentiste. Le contrôle des paramètres est une tâche ambitieuse et très complexe. La théorie relève plus de l'automatisme (control system) et des mathématiques appliquées que du traitement du signal.

Les modèles d'excitateurs aujourd'hui répertoriés, étudiés et analysés sont aujourd'hui les anches simples, les anches lipales, l'interaction corde-archer, intéraction marteau-membrane ou marteau-corde. Les systèmes d'anche doubles continuent à poser beaucoups de problème de modélisation.

Le système (non-musical) le plus simple qui permet d'engendrer des auto-oscillation est le circuit de Chua (fig. 23 et fig. 24)

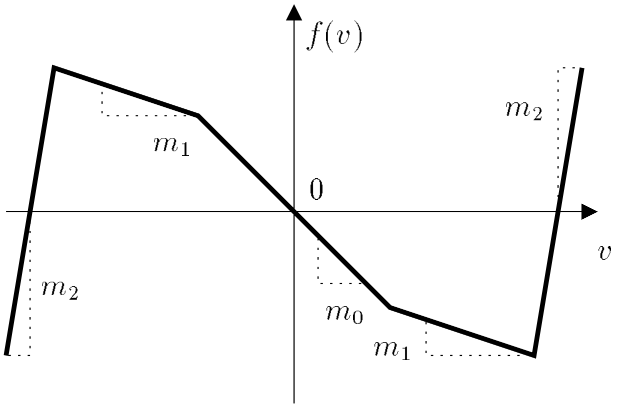


FIG. 24 – *Caractéristique de la diode de Chua*

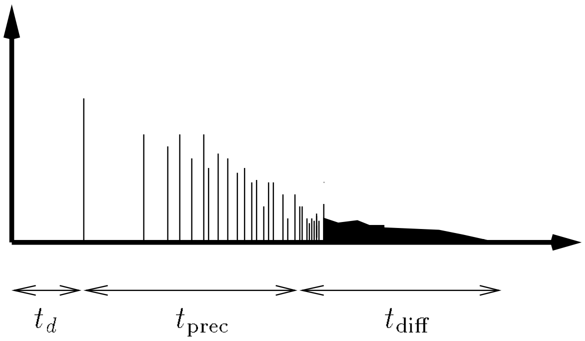


FIG. 25 – *Réponse impulsionnelle type d'une salle*

2.4.11 Restrictions d'utilisation

- Précision spatiale.
- Stabilité.
- Méthode d'analyse.

2.5 Table de réverbération numérique

Historiquement, la première application des guides d'onde concernait les tables de reverbérations numériques. L'idée générale consiste à réaliser un dispositif numérique qui reproduit artificiellement une sorte d'effet de salle. D'un point de vue pragmatique, il convient d'examiner la réponse impulsionelle provenant de l'effet de salle (typiquement la réponse mesurée en un point à un coup de feu tiré dans une salle). On distingue traditionnelement 3 zones (fig. 25), à savoir le son direct, les premières réflexions, et finalement la réverbération tardive, appelée également la diffusion.

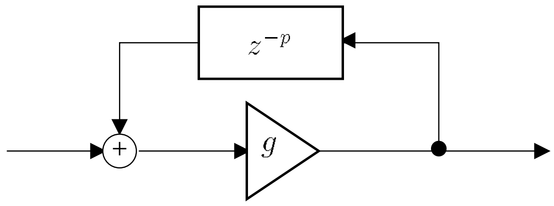


FIG. 26 – *Filtre en peigne*

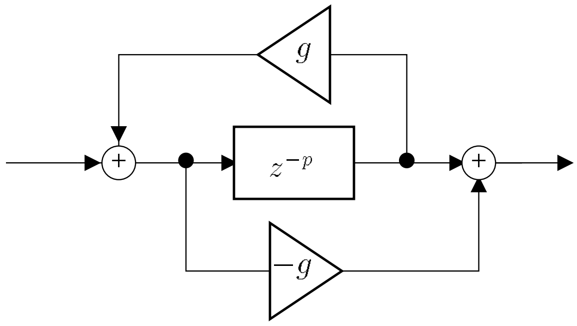


FIG. 27 – *Filtre passe-tout*

2.5.1 Réverbération numérique à l'aide de lignes à retard

La première idée ingénieuse qui vient à l'esprit pour réaliser un dispositif dont la réponse impulsionnelle se comporte de la même façon qu'une réponse impulsionnelle provenant de l'effet de salle, consiste à mettre en parallèle plusieurs filtres en peignes utilisant des retards qui sont à peu près premiers entre eux, comme indiqué sur la figure 26.

On constate que chaque filtre en peigne génère une impulsion tous les p échantillons. En conséquence, si chaque cellule génère un train d'impulsions dont les périodes sont premières entre elles, le signal de sortie n'est plus périodique. La réponse impulsionnelle d'un tel système est donc constitué dans un premier temps d'une série d'échos distincts successifs. Après un certain temps, la densité des échos est telle qu'il ne reste plus qu'un signal diffus : la réverbération tardive. Au bout du compte, la structure temporelle de l'effet de salle est reproduite. L'analyse fréquencielle d'un tel système est à rapprocher de l'analyse modale d'un salle parallélépipédique. La fonction de transfert de chaque cellule vaut $\frac{g}{1-gz^{-p}}$, ce qui fait apparaître p pôles uniformément répartis sur le cercle de centre 0 et de rayon $\sqrt[p]{g}$. Les pôles (i.e. les résonances) ne sont stables que si $g < 1$. Or l'analyse modale d'une salle parallépipédique aboutit précisement à 3 séries de modes dont les fréquences sont proportionelles $f_n = \frac{nc}{l}$, où l est la distance qui sépare deux murs qui se font face. En conséquence, le filtre en peigne reproduit la structure modale d'une paire de murs distants de $\frac{p}{2}L$, où L représente la distance que parcourt un front d'onde pendant le laps de temps qui sépare deux échantillons : $L = \frac{c}{F_e}$. Les trois séries de modes sont obtenus en utilisant en parallèle trois filtres en peigne dont les caractéristiques dépendent de la géométrie de la salle parallépipédique.

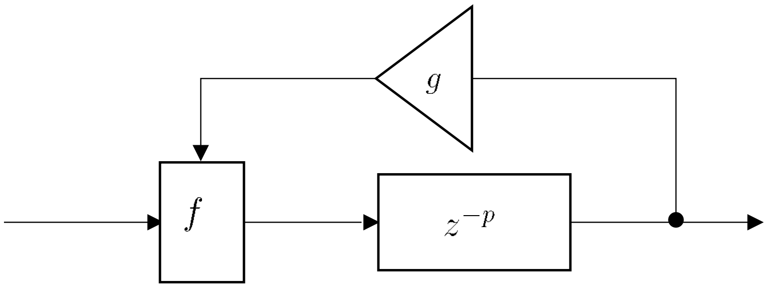


FIG. 28 – *Schéma idéal prenant en compte les erreurs de calculs*

Les inconvéniants de cette méthode sont divers. Tout d'abord cette méthode colore les signaux (certaines fréquences sont plus amplifiées que d'autres). Dans certaines applications de sonorisation, il est préférable de garder un spectre plat, auquel cas l'utilisation de cellules passe-tout du type de celle proposée figure 27 est préconisée à la place de cellule de filtres en peigne. Lexicon[20] utilise une méthode denotée filtre passe-tout en gigogne (*nested allpass filters*) pour réaliser leurs cellules réverbératrices. Elle consiste à remplacer dans la figure 27 le coefficient g par un autre filtre passe-tout ... et ainsi de suite. Pour des raisons d'efficacité (et au détriment de la mémoire utilisée) il est possible d'intervertir le coefficient g et la ligne à retard z^{-p} (c'est évident parce que la fonction de transfert est symétrique en g et en z^{-p} : $H(z) = \frac{g+z^{-p}}{1+gz^{-p}}$).

Il existe d'autres inconvéniants beaucoup plus structurels. Tout d'abord la constante de temps[21] des cellules passe-tout et en peigne étant de l'ordre de $\frac{p \ln 2}{1-g}$, nous sommes ammenés à utiliser la cellule avec des valeurs de g proche de 1, ce qui nous place en limite de stabilité. Rien ne nous garantit que les erreurs de calculs ne vont pas faire sortir un des p pôles à l'extérieur du cercle unité et faire diverger le système. De plus, la constante de temps est extrèmement sensible aux variations du paramètre g. Entre une cellule pour laquelle g vaut 0.99 et une autre pour laquelle g vaut 0.999, les constantes de temps sont dans un rapport 10.

D'autre part, toujours à cause des erreurs de calcul, il est possible que la position de repos ne soit pas le seul état stable du système. Les erreurs numériques se traduisent par l'introduction d'une fonction non linéaire dans la boucle idéale (fig. 28). Ce type de schéma (que nous retrouvons dans les modèles physiques d'intruments, voir chapitre 2.4.10) est susceptible d'auto-osciller. Les états stables auto-oscillatants non nuls sont appelés cycles limites et si ce sont des états recherchés dans les modèles physiques d'instruments, ce sont des états à fuire à tout prix dans la synthèse de table de réverbération numérique. Il est parfois possible d'entendre cette auto-oscillation qui refuse de disparaître sur certaines tables de réverbération numérique bon marchée, même en dehors de tout signal d'entrée.

20. Lexicon est une marque connue produisant des racks multi-effets et en particulier de nombreux effets de réverberations numériques.

21. La constante de temps correspond à la période de demi-vie de la réponse impulsionnelle.

2.5.2 Réverbération numérique à l'aide de guide d'onde

Il est troublant de constater à quel point la structure du modèle de la figure 26 est proche de celle du modèle compact de Karplus-Strong de la figure 16. Il est intéressant, et même utile de voir jusqu'à quel point le rapprochement de ces deux structures peut se faire. Les deux structures sont en fait identiques à un retard près.

Les différences entre les deux modèles interviennent dans l'interprétation que l'on en fait, et dans la manière d'assembler ensemble des cellules. Nous avons vu qu'une cellule représentait une paire de mur. L'interprétation en guide d'onde confirme cette représentation, la longueur du guide d'onde étant l'exact reflet de la distance qui sépare un deux mur. Là où nous nous contentions de sommer la sortie de chacune des cellules, il nous faut à présent, coupler les guides d'onde avec les jonctions adéquates afin de garder une interprétation physiques des signaux qui sont échangés.

L'interprétation physique du système numérique permet de traiter une information énergétique consistante dans le système. Le modèle physique assure que toutes les opérations effectuées dans le système ne rajoutent pas d'énergie et sont au pire dissipatives (par exemple en ce qui concerne les opération arithmétiques en arrondissant en valeur absolue les signaux à la valeure inférieure) ce qui nous assure de ne pas tomber dans des systèmes divergents. De plus il est très facile de rajouter des pertes dans le systèmes (avec un filtre passe-bas par exemple) afin d'éviter de rester coincé dans des états de cycles limites. En outre, les pertes physiques que les signaux acoustiques subissent lors de leur propagation ou de leur réflexion sur les parois s'introduisent très facilement avec les modèle de guides d'onde.

Les guides d'onde facilitent également la construction de réseaux complexes de guides d'onde. Quelque soit l'allure du réseau, le système est garanti de ne pas diverger numériquement.

Il est également possible de rendre dépendant du temps tous les paramètres du réseau (sauf peut-être la structure du réseau) sans risque de divergence ou d'artefact audible. Tout se passe comme si les murs de la salle se déplaçaient en même temps que les signaux sonores se propagaient à l'intérieur de la salle.

Dans le cas d'applications temps-réel (processeurs d'effets sonores par exemple), les modèles de filtres en peignes, filtres passe-tout, et de guides d'onde sont les technologies appropriées et utilisées. Dans le cas de simulations numériques d'effet de salle, les guides d'onde donnent une bonne approximation de la solution par une méthode analogue à la méthode des rayons. Cette méthodologie, quoique rapide et efficace, est évidemment un substitut approximatif par rapport à une simulation totale et complète, par exemple avec un code d'éléments finis, de l'acoustique d'une salle. Elle s'interprète pourtant relativement bien avec une analyse modale de la salle, ou une analyse de l'acoustique en termes de rayons (ray-tracing).

Références

[AR77] J. B. Allen and L. R. Rabiner. A unified approach to short-time Fourier analysis and synthesis. *Proc. IEEE*, 65(11):1558–1564, November 1977.

[Bel83] M. Bellanger. *Traitement numérique du signal.* Masson, CNET-ENST, 1983.

[Cho82] C. Chouzenoux. *Analyse Spectrale à Résolution Variable, Application au Signal Parole.* PhD thesis, École Nationale Supérieure des Télécommunications, 1982.

[Del91] J.-P. Delams. *Éléments de théorie du signal : les signaux déterministes.* Ellipses, 1991.

[EC97] G. Evangelista and S. Cavaliere. The discrete-time frequency warped wavelet transforms. In *Proc. IEEE ICASSP'97*, Munich, April 1997.

[GL84] D. W. Griffin and J. S. Lim. Signal estimation from modified short-time Fourier transform. *IEEE Trans. Acoust., Speech, and Signal Process.*, ASSP-32(2):236–242, April 1984.

[GW95] C. Gasquet and P. Witomski. *Analyse de Fourier et application, filtrage, calcul numérique et ondelette.* Masson, 2ème edition, 1995.

[KHL97] M. Karjalainen, A. Härmä, and U. K. Laine. Realizable warped IIR filters and their properties. In *Proc. IEEE ICASSP'97*, pages 2205–2208, Munich, April 1997.

[Por80] M. R. Portnoff. Time-frequency representation of digital signals and systems based on short-time Fourier analysis. *IEEE Trans. Acoust., Speech, and Signal Process.*, ASSP-28(1):55–69, February 1980.

[Sas96] T. Saso. On short-time fourier transform with single-sided exponential window. *Signal Processing*, 55:141–148, 1996.

Index

échantillonnage, 25
énergie, 31, 37

aliasing, 27
analyse harmonique, 23
analyse modale, 36, 40, 49, 51
Armstrong E.H., 25
attaque, 28
auto-oscillation, 23, 46, 50

chaos, 33
Chebyshev, 29
chevalet, 40
Chowning, 26
conditions initiales, 34, 40
convertisseur numérique-analogique, 1
Csound, 24

decay, 28
dispersion numérique, 33
divergence, 33, 35, 50, 51

EDO, 30
EDP, 33
excitateur, 46

filtres
- de Lagrange, 42
- en peigne, 40, 49
- en treillis, 45
- interpolateurs, 42
- passe-tout, 43

fréquence fondamentale, 27

harmonicité, 26
Heisenberg, 33

inharmonicité, 27, 41
interpolation, 24, 42

jonction, 44, 51

Lagrange, 42
ligne à retard, 37

MIDI, 28
modulante, 26
modularité, 24
Music V, 24

non linéarité, 46
non-linéarité, 29
Note-Off, 28
Note-On, 28

partiels, 26
patch, 24
pertes, 39, 42, 46
porteuse, 26
précalcul, 23

quadripole, 45

réflexion, 37, 40, 45, 48
ray-tracing, 51
relation d'incertitude, 33, 42
release, 28
repliement spectral, 27

Schaeffer P., 28
sinusoïdes, 23
stabilité, 33, 35, 51
sur-échantillonnage, 24
sustain, 28
système excitateur, 46

table d'harmonie, 40
table d'onde, 23
transmittance, 37, 40, 45

Table des matières

1 Synthèse par modèle de signaux **1**
1.1 Rappel sur les transformées 2
1.1.1 Notations et convolutions 2
1.1.2 Transformée de Fourier à Temps Continu (TFTC) 2
1.1.3 Transformée de Fourier à Temps Discret (TFTD) 3
1.1.4 Lemmes 4
1.1.5 Transformée de Fourier Discrète (TFD) 6
1.2 Transformée de Fourier à Court Terme (TFCT) 8
1.2.1 Interprétations 8
1.2.2 Simplifications 10
1.3 Reconstruction à partir de TFCT 12
1.3.1 Filter Bank Summation (FBS) 12
1.3.2 Overlap and add (OLA) 14
1.3.3 Équivalence OLA - FBS : 15
1.3.4 Exemple 16
1.4 Reconstruction à partir d'une TFCT modifiée 17
1.4.1 Critère quadratique 17
1.4.2 Comparaison avec la reconstruction classique 18
1.5 Filtrage par Transformée de Fourier à Court Terme 19
1.5.1 Gain variable 19
1.6 Vocodeur de phase 19
1.6.1 Analyse additive 20
1.6.2 Synthèse additive 20
1.7 Oscillateur numérique 22
1.8 Réseau d'oscillateurs, notion de patch 23
1.9 Synthèse par modulation de fréquence 24
1.9.1 Origines 24
1.9.2 Modulation de fréquence 25
1.10 Contrôle MIDI 26
1.11 Synthèse par distorsion d'amplitude 27

2 Synthèse par modèle physique **28**
2.1 Simulations numériques 28
2.1.1 Motivations 28
2.1.2 Réalisation 29
2.1.3 Les éléments finis 29
2.1.4 Exemple d'application des éléments finis 30
2.1.5 Les différences finies 31
2.1.6 Exemple d'application des différences finies 32
2.1.7 Limitations des méthodes numériques 32
2.1.8 Exemple de restriction d'utilisation 33
2.1.9 Simulations de la vibration d'une corde avec raideur . . . 34
2.2 Synthèse modale 34
2.3 Modèle virtuel 35
2.4 Modèles de guide d'onde 35
2.4.1 Guide d'onde acoustique sans perte 35
2.4.2 Systèmes avec pertes 36
2.4.3 Modèle de Karplus-Strong 37

2.4.4 Limitations et amélioration du modèle 39
2.4.5 Retards fractionnaires 40
2.4.6 Jonctions de N guides d'onde 42
2.4.7 Liens avec les filtres en treillis 43
2.4.8 Pertes viscothermiques 44
2.4.9 Synthèse de cordes de piano 44
2.4.10 Modèle d'excitateurs 44
2.4.11 Restrictions d'utilisation 46
2.5 Table de réverbération numérique 46
2.5.1 Réverbération numérique à l'aide de lignes à retard 47
2.5.2 Réverbération numérique à l'aide de guide d'onde 49