

Chapitre 1

Synthèse sonore numérique par modulation de fréquence

1.1 Introduction

Nous sommes au début des années 1970. Il y a déjà presque une décennie que Max Mathews a publié « The Digital Computer as a Musical Instrument » [1]. Pierre Boulez vient d'être chargé par le président George Pompidou de créer un institut dédié à la création musicale (le futur IRCAM) associé au centre d'art et de culture Pompidou (l'ouverture au public aura lieu en 1977). Le champ de recherche est en plein essor, mu par le rêve de fabriquer et de contrôler des sons « naturels » à l'aide d'une machine informatique et d'incorporer ces nouveaux instruments électroniques dans une organologie étendue pour des compositeurs tels que Iannis Xenakis ou John Cage. Les chercheurs sont aussi souvent des musiciens et compositeurs accomplis comme Jean-Claude Risset, Max Mathews ou John Pierce. Parmi ceux-ci se trouve John Chowning, qui en 1973 publie une idée très intéressante pour la synthèse sonore [2] : utiliser un concept connu en transmission radio, la modulation de fréquence, pour réaliser des sons dont on peut contrôler l'évolution spectrale avec une économie de moyens étonnante !

Cette technique, qui a eu un succès important, y compris du point de vue industriel (série des circuits OPL de chez Yamaha) et du point de vue des synthétiseurs grand public (DX7 de Yamaha) peut être reproduite de manière simple à l'aide d'un logiciel de calcul numérique tel que MATLAB[®], Scilab ou Octave ou encore d'un logiciel dédié à la création musicale tel que Pure Data, écrit par Miller Puckette et disponible sous licence libre [3, 4]. Dans ce qui suit, les principes de base seront abordés puis quelques éléments de pour régler les algorithmes, appliqués ensuite sur trois exemples (vents, piano, gong). Les algorithmes seront donnés sous forme de pseudo-code, aisés à mettre en oeuvre par la suite à l'aide du logiciel choisi.

Notations. Les signaux à temps continu seront des fonctions de $t \in \mathbb{R}$, et les signaux à temps discret des séquences fonctions de $n \in \mathbb{Z}$. TF est l'abréviation de Transformée de Fourier, notée \hat{x} .

1.2 Principe général

En transmission radio, la modulation de fréquence utilise une porteuse f_p à haute fréquence (typiquement quelques dizaines de MHz), c'est à dire un signal sinusoïdal, dont la fréquence est modulée par un signal audio avec une bande n'excédant pas 20 kHz. Le résultat est un signal dont le spectre est centré sur la porteuse avec un étalement très faible comparativement à celle-ci. L'idée de Chowning est d'utiliser le même principe mais avec des fréquences porteuse et modulante *du même ordre de grandeur*.

1.2.1 Modulations

Le système élémentaire sur lequel repose la technique décrite par Chowning est un oscillateur sinusoïdal dont il est possible de moduler la fréquence f et l'amplitude m . Ces deux grandeurs sont alors des fonctions du temps : la fréquence instantanée –ou loi de modulation en fréquence– $f(t)$ et l'amplitude instantanée –ou loi de modulation en amplitude– $m(t)$.

Modulation d'amplitude. La modulation d'amplitude est obtenue à l'aide multiplicateur : le signal est écrit sous la forme

$$s(t) = m(t) \cos(2\pi f_0 t + \phi_0) \quad (1.1)$$

où la fréquence f_0 ne dépend pas du temps.

Remarque 1.1. Variations lentes. De manière souvent implicite, $m(t)$, l'amplitude instantanée, est une fonction qui « varie lentement » en référence à f_0 . En effet si, à titre de contre-exemple, on prend $m(t) = 1 + \cos(2\pi f_0 t + \phi_0)$, il vient $s(t) = \frac{1}{2} + \cos(2\pi f_0 t + \phi_0) + \frac{1}{2} \cos(4\pi f_0 t + 2\phi_0)$. Pour $f_0 = 440$ Hz, notre perception conclura à un son complexe, composés de deux composantes à f_0 et $2f_0$ mais dont l'amplitude est constante.

Remarque 1.2. Modulation d'amplitude et modulation en anneau. De nombreux auteurs distinguent le cas $m(t) \geq 0$ (modulation d'amplitude) et le cas de signaux de modulation bipolaires, c'est à dire présentant des alternances positives et négatives (modulation en anneau).

Modulation de fréquence : approche intuitive. Comment définir un signal sinusoïdal d'amplitude constante dont la fréquence varie? Ce signal sera noté

$$s(t) = m_0 \cos(\phi(t)) \quad (1.2)$$

où $\phi(t)$ désigne la *phase instantanée* du signal. Pour une fréquence constante f_0 on aura $\phi(t) = 2\pi f_0 t + \phi_0$, ϕ_0 étant la phase à l'origine, comprise entre $-\pi$ et π . Lorsque la fréquence dépend du temps, on note $f(t)$ cette *fréquence instantanée*, et l'expression de la phase instantanée découle d'un raisonnement infinitésimal. Pendant une durée très faible δt le signal est considéré comme sinusoïdal pur et l'incrément de phase correspondant vaut $\delta\phi = 2\pi f(t)\delta t$. Le lien entre fréquence et phase instantanées s'exprime par suite sous la forme

$$\phi(t) = 2\pi \int_0^t f(u) du + \phi_0 \quad (1.3)$$

Exemple 1.3. pour fabriquer un « chirp » linéaire, c'est à dire un sinus dont la fréquence varie linéairement de f_1 à $f_1 + \Delta f$ en Δt secondes, on aura $f(t) = f_1 + \frac{\Delta f}{\Delta t} t$ et $\phi(t) = 2\pi f_1 t + \frac{\Delta f}{2\Delta t} t^2 + \phi_0$. La phase instantanée est quadratique.

Remarque 1.4. Non unicité de la définition intuitive.

Le signal défini précédemment est de module borné par m_0 . Définissons un nouveau signal $s_1(t) = (2m_0)^{-1} s(t)$. On a $|s_1(t)| \leq 1/2$ et donc il existe une fonction $\phi_1(t)$ à valeurs dans $[-\pi/6, \pi/6]$ telle que $s_1(t) = \cos \phi_1(t)$ et donc telle que $s(t) = 2m_0 \cos \phi_1(t)$. Dès lors on serait tenté de dire que $s(t)$ est maintenant un signal d'amplitude $2m_0$ et de phase instantanée $\phi_1(t)$ (et par suite de fréquence instantanée $(2\pi)^{-1} \phi_1'(t)$ où le prime indique la dérivée par rapport à t). De manière plus générale, soit $x(t)$ un signal dont le module est borné par le réel positif M . Il est alors possible de définir un signal $x_1(t) = M^{-1} x(t) = \cos \phi_x(t)$ où $\phi_x(t)$ est à valeur dans $[-\pi/2, \pi/2]$. D'une part, on comprend que pour un signal $x(t)$ quelconque de $L^\infty(\mathbb{R})$, $(2\pi)^{-1} \phi_x'(t)$ ne représentera en général pas une fréquence instantanée au sens intuitif décrit précédemment mais également, on voit qu'il est possible de définir une infinité d'expressions du type $x(t) = M_1 \cos \phi_{x, M_1}(t)$ dès lors que $M_1 \geq M$.

Note : définition univoque des modulations de fréquence et d'amplitude.

La définition univoque de la fréquence instantanée est basée sur la notion de signal analytique et de paire canonique {modulation d'amplitude, modulation de phase}; elle est notamment étudiée et commentée dans [5]. On rappelle que le signal analytique associé à un signal réel $x(t)$ est le signal $z(t)$ tel que sa TF vaut $\hat{z}(\nu) = 2\hat{x}(\nu)$ pour $\nu > 0$ et $\hat{z}(\nu) = 0$ pour $\nu < 0$ (pour plus de détail, voir par exemple [6]). Ce signal est s'écrit à l'aide de la transformée de Hilbert de x , notée $\mathcal{H}[x](t)$ comme $z(t) = x(t) + i\mathcal{H}[x](t)$.

Définition 1.5. *Paire canonique.* Soit $x(t)$ un signal réel et $z(t)$ le signal analytique associé, écrit sous la forme $z(t) = m(t) \exp i\phi(t)$, avec $m(t) > 0$. Le couple $\{m(t), \phi(t)\}$ est appelé *paire canonique* associée à x . Elle définit de manière univoque l'écriture de $x(t) = m(t) \cos \phi(t)$ où $m(t)$ est l'*amplitude instantanée* du signal et $\phi(t)$ sa *phase instantanée*. La fréquence instantanée du signal s'obtient alors par dérivation en suivant $\nu_x(t) = (2\pi)^{-1} \phi'(t)$.

Application à un chirp linéaire. Quelle est la fréquence instantanée d'un signal « chirp linéaire » simple de la forme $x(t) = \cos \phi(t) = \cos(bt^2)$? Considérons le signal $z_\phi(t) = \exp i\phi(t)$. Ce signal est une gaussienne complexe, dont la TF est une autre gaussienne complexe, soit $\hat{z}_\phi(\nu) = e^{i\frac{\pi}{4}} \sqrt{\frac{\pi}{b}} e^{-i\pi^2 \nu^2 / b}$ (ici on emploie la TF des distributions tempérées). On en déduit que $\hat{z}_\phi(\nu)$ n'est pas nulle pour $\nu < 0$ et donc *a fortiori*, $z_\phi(t)$ n'est pas le signal analytique associé à $x(t)$. La fréquence instantanée de $x(t)$ n'est donc pas $\frac{bt}{\pi}$. Cela semble contredire le résultat de la figure 1.2, où le signal est constitué de chirps linéaires par morceaux et pour lequel le spectrogramme semble suivre la modulation proposée.

Lever cette contradiction n'est pas trivial. Pour cela B. Picinbono [5] propose de calculer l'erreur $\epsilon(t) = \mathcal{H}[x](t) - \sin \phi(t) = -i(z(t) - z_\phi(t))$, qui caractérise la différence entre le signal analytique $z(t)$ associé à x et $z_\phi(t)$. Il montre que $\mathcal{H}[x](t) = \left\{ \sqrt{\frac{2}{\pi}} \int_0^t \cos(bu^2) + \sin(bu^2) du \right\} \sin bt^2 + \sqrt{\frac{2}{\pi}} \left\{ \int_0^t \cos(bu^2) - \sin(bu^2) du \right\} \cos bt^2$. En utilisant les résultats classiques sur les intégrales de Fresnel, soit $\int_0^{+\infty} \cos u^2 du = \int_0^{+\infty} \sin u^2 du = \frac{1}{2} \sqrt{\frac{\pi}{2}}$ on en déduit que l'erreur $\epsilon(t)$ tend vers 0 quant $t \rightarrow +\infty$. D'autre part, cette expression permet de calculer le signal z .

Schéma général d'un oscillateur sinusoïdal modulé. Dans le cas général, un oscillateur modulé sera de la forme $s(t) = m(t) \cos \phi(t)$, en gardant à l'esprit que $\cos \phi(t)$ peut très bien ne pas ressembler à une sinusoïde, selon le choix de $\phi(t)$. Cependant cette forme se réfère à la réalisation pratique de tels oscillateurs analogiques, pour lesquels un VCO (*Voltage Controlled Oscillator*, oscillateur contrôlé en tension) est l'élément de base. Ce système électronique génère un signal dont la fréquence est contrôlée par la tension d'entrée. Cette brique élémentaire d'un système de modulation de fréquence est schématisé dans la figure 1.1.

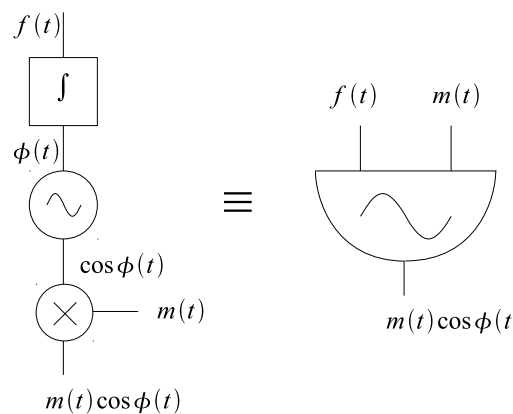


FIGURE 1.1 – Oscillateur modulé en fréquence et en amplitude

Modulation de fréquence pour les signaux à temps discret. Dans le cadre des signaux à temps discret (noté n) on utilise usuellement une approximation de l'intégrale par la méthode des rectangles, ce qui conduit à remplacer l'intégrale par une somme cumulée sur la fonction de modulation de fréquence échantillonnée $f(t_n)$, $t_n = nT_e$ avec $T_e = F_e^{-1}$ représente la période d'échantillonnage. Son inverse, F_e , est la fréquence d'échantillonnage. Pour une phase à l'origine ϕ_0 nulle, on obtient alors le pseudo-code suivant :

Pcode 1.1: Synthèse numérique d'un sinus modulé en fréquence

Fonction $x = \text{oscFM}(f, m, F_e)$

1. Calculer la somme cumulée $\phi(n) = 2\pi T_e \sum_0^n f(t_n)$
2. Calculer le signal $x(n) = m(t_n) \cos(\phi(n))$.

Ainsi en définissant la fonction $f(t)$ de la figure 1.2 on constate sur le spectrogramme¹ en vis-à-vis que le contenu fréquentiel du son obtenu suit bien cette fonction modulante.

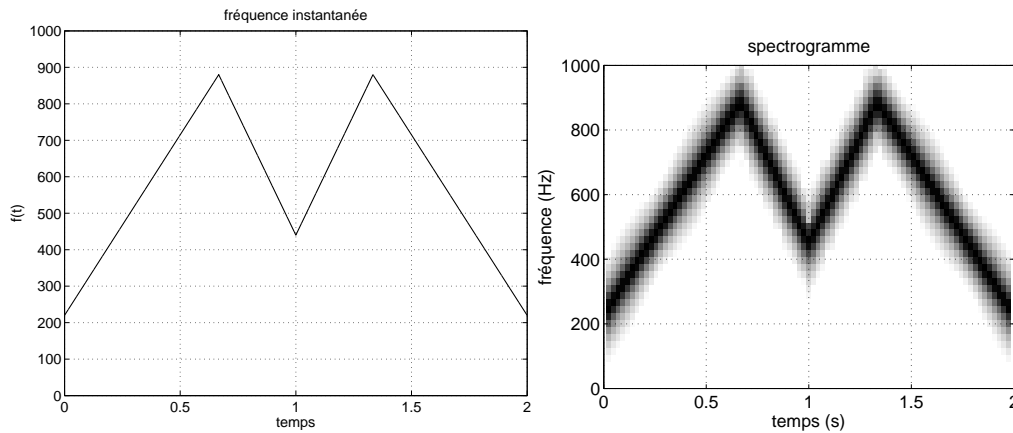


FIGURE 1.2 – Modulation de fréquence $f(t)$ et le spectrogramme du signal obtenu à l'aide de l'algorithme

1.2.2 Synthèse sonore « par modulation de fréquence »

Le principe de synthèse sonore décrit par Chowning [2] s'inspire des transmissions radio. La fréquence instantanée dans ce cas est définie comme une oscillation autour d'une fréquence porteuse f_p . Pour une oscillation sinusoïdale, cette fréquence instantanée vaut $f(t) = f_p + D \cos 2\pi f_m t$, en supposant f maximale à $t = 0$. D désigne l'excursion ou déviation en fréquence, c'est à dire l'amplitude des variations de fréquence autour de f_p . D'après 1.3 le signal correspondant s'écrit

$$s(t) = m_0 \cos(2\pi f_p t + I \sin(2\pi f_m t)) \quad (1.4)$$

en supposant la phase instantanée nulle à l'origine. $I = \frac{D}{f_m}$ est dénommé *index de modulation*. En suivant les conventions du paragraphe précédent, le signal $s(t)$ est obtenu à l'aide de la combinaison d'oscillateurs modulés, telle que présentée figure 1.3.

Spectre de la modulation de fréquence L'intérêt musical de la technique repose sur sa capacité à produire des spectres variés en réglant le paramètre de contrôle I . Les deux théorèmes suivant permettent de caractériser le spectre de $s(t)$.

1. Le spectrogramme est la représentation en niveau de couleur du module de TF à court terme, ou TF glissante. C'est à dire les spectres successifs de courts segments du signal qui se recouvrent. Le résultat est une représentation qui localise l'énergie dans un plan temps-fréquence et permet de juger des variations du spectre au cours du temps.

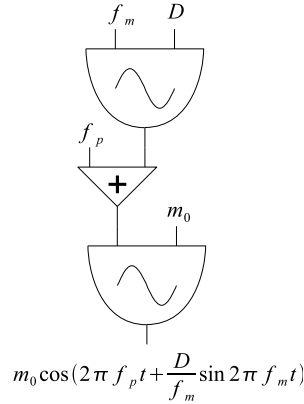


FIGURE 1.3 – Synthèse par modulation de fréquence par cascade d'oscillateurs modulés

Théorème 1.6. *Le spectre de $s(t)$ est un spectre de raies, dont les fréquences sont de la forme $f_p \pm kf_m$ (et $-f_p \pm kf_m$ par symétrie) où $k \in \mathbb{N}$.*

Démonstration. On pose $e(t) = m_0 \exp(2i\pi f_p t + iI \sin(2\pi f_m t)) = m_0 \exp(2i\pi f_p t) \exp(iI \sin(2\pi f_m t))$. Le facteur $\exp(iI \sin(2\pi f_m t))$ est une fonction périodique de période $T_m = f_m^{-1}$, de classe \mathcal{C}^∞ et de carré intégrable sur $[0, T_m]$. Elle est décomposable en série de Fourier sous la forme $\sum_{k \in \mathbb{Z}} \alpha_k e^{2i\pi k f_m t}$, on déduit que $e(t) = m_0 \sum_{k \in \mathbb{Z}} \alpha_k e^{2i\pi f_p t + 2i\pi k f_m t}$. Comme $s(t) = \text{Re}(e(t))$ cela clôt la démonstration. \square

Théorème 1.7. *Le signal $s(t)$ peut être développé sous la forme :*

$$s(t) = m_0 \sum_{k \in \mathbb{Z}} J_k(I) \cos(2\pi(f_p + kf_m)t) \quad (1.5)$$

où $J_k(I)$ désigne la fonction de Bessel de première espèce d'ordre k . On rappelle que $J_{-k}(I) = (-1)^k J_k(I)$.

Démonstration. Les coefficients du développement en série de Fourier de la fonction $e^{ix \sin \theta}$ s'écrivent $c_k = (2\pi)^{-1} \int_0^{2\pi} e^{i(x \sin \theta - n\theta)} d\theta$. On reconnaît là l'expression intégrale de la fonction de Bessel de première espèce d'ordre n , notée $J_n(x)$. Cela conduit à la formule de *Jacobi-Anger*² :

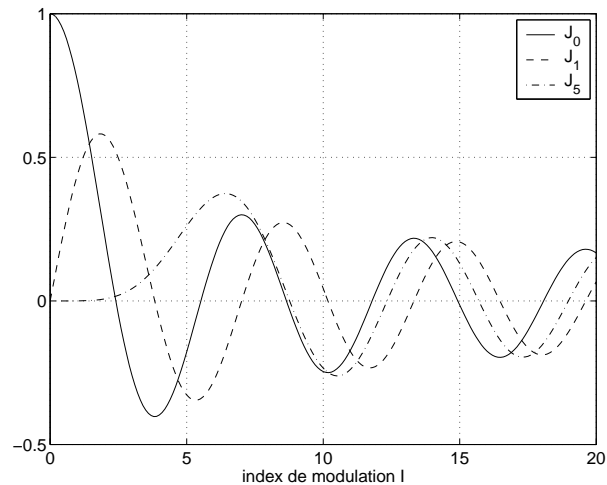
$$e^{ix \sin \theta} = \sum_{n=-\infty}^{+\infty} J_n(x) e^{in\theta}$$

le résultat s'en déduit immédiatement. \square

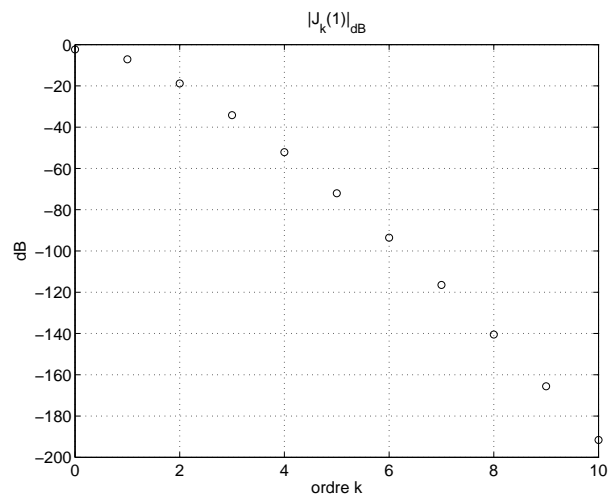
Influence de I sur les bandes latérales. L'équation 1.5 permet de mettre en évidence une propriété intéressante des spectres obtenus : *les bandes latérales du spectre sont d'autant plus larges que I est élevé*. Cette propriété est une conséquence de la forme des fonctions de Bessel, $|J_k(I)|$ étant l'amplitude de la composante de fréquence $f_p \pm kf_m$. Cette forme est décrite dans la figure 1.4. Cette figure appelle plusieurs remarques utiles :

1. les fonctions de Bessel ont une forme d'ondulation amortie, $J_k(I) \rightarrow 0$ lorsque $I \rightarrow \infty$;
2. $J_k(0) = 0, \forall k \neq 0$. On retrouve le fait que si $I = 0$ la seule composante d'amplitude non nulle est la composante centrale à la fréquence f_p . Le signal n'est alors pas modulé en fréquence ;
3. les fonctions d'ordre élevé (cas $k = 5$ sur la figure) présente « un retard à l'ondulation ». Ce retard est d'autant plus grand que k est grand.

2. deux mathématiciens du 19ème siècle.

FIGURE 1.4 – Allure de quelques fonctions de Bessel en fonction de I

En conséquence de la remarque 3 : quelque soit la valeur de I , il existe toujours une valeur k_m de k au delà de laquelle l'amplitude des composantes est aussi petite que l'on veut. Le spectre peut être considéré avec une bonne approximation comme à bande limitée. La figure 1.5 ci-dessous montre par exemple que pour $I = 1$, pour $k > 6$ toute les composantes seront d'amplitude inférieure à -80 dB. Pour $I = 2$ la même propriété s'appliquerait

FIGURE 1.5 – Fonction de Bessel pour $I = 1$ en fonction de l'ordre k

pour $k > 8$. On en conclut que le nombre de composantes significatives (*i.e.* dont l'amplitude est supérieure à un certain seuil, ici -80 dB) devient plus grand lorsque I augmente. Dans le cadre de la synthèse sonore, cela permet de contrôler l'enrichissement du spectre à l'aide du paramètre I . Cette propriété est intéressante car de nombreux sons naturels possèdent un spectre de largeur variable au cours du temps. De manière générale par exemple, les instruments à son libre tels que le piano, les cloches de carillon ou la guitare voient leurs composantes aiguës s'éteindre avant les composantes graves.

Représentation graphique du spectre. La représentation de la figure 1.6 montre l'étalement croissant du spectre lorsque I augmente. Cette représentation montre la valeur absolue du coefficient de chaque composante

en fonction de sa fréquence. Elle est obtenue à partir de l'équation 1.5 mise sous la forme

$$s(t) = \frac{m_0}{2} \left[\sum_{k \in \mathbb{Z}} J_k(I) \exp(2i\pi(f_p + kf_m)t) \right] + \frac{m_0}{2} \left[\sum_{k \in \mathbb{Z}} (-1)^k J_k(I) \exp(2i\pi(-f_p + kf_m)t) \right]. \quad (1.6)$$

Pour la figure on a choisi $f_p = 10f_m$ et $m_0 = 1$. On peut également y lire la symétrie en module autour de f_p

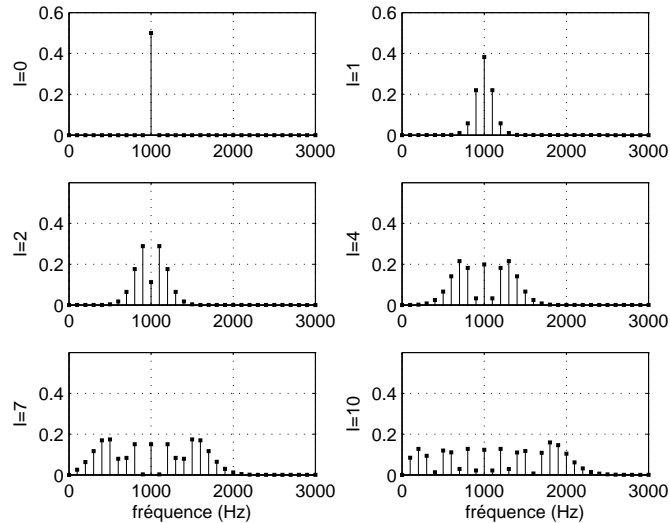


FIGURE 1.6 – Spectre de la modulation de fréquence, pour $f_p = 1000$ Hz et $f_m = 100$ Hz. Les amplitudes des composantes sont représentées en module. Seule la partie des fréquences positives est représentée.

lorsque $I = 1$ ou $I = 2$. A partir de $I = 7$, l'étalement du spectre devient supérieur à f_p , c'est à dire que les bandes latérales autour de $-f_p$ interfèrent avec celles autour de $+f_p$. Ce phénomène est mis en évidence dans l'équation 1.6 où il apparaît que l'amplitude de la composante de fréquence qf_m vaut $\frac{1}{2} [J_{q-10}(I) + (-1)^{q+10} J_{q+10}(I)]$. Il en résulte que le spectre perd sa symétrie autour de f_p .

Remarque 1.8. L'interférence entre bandes latérales telle que vue ci-dessus tient au choix de f_p comme multiple de f_m . Dans le cas où f_m et f_p ne sont pas dans un rapport entier, l'étalement du spectre produira des composantes « réfléchies » ainsi qu'il est montré dans la figure 1.7.

Utilisation musicale de la modulation de fréquence. Dans le cas des transmissions radio la fréquence porteuse ($f_p \sim 100$ MHz) est d'un ordre de grandeur bien supérieur à la fréquence modulante ($f_m \sim 10$ kHz), le spectre est alors faiblement étalé autour de f_p . C'est une situation du type du graphe pour $I = 1$ de la figure 1.6. L'idée de Chowning est de prendre des fréquences porteuse et modulante dans des rapports faibles, typiquement 1 ou 2. On est alors dans la situation où les bandes latérales autour de $-f_p$ et $+f_p$ interfèrent pour *des valeurs faibles de l'index de modulation*. Cela conduit à avoir des spectres très variables lorsqu'on modifie le paramètre de contrôle I . Cette forte variabilité du spectre est une raison du succès de la méthode : avec une architecture économe (il suffit de deux oscillateurs sinusoidaux pour fabriquer $s(t)$) on dispose d'un générateur de son dont le spectre est hautement variable et contrôlable à l'aide d'un seul paramètre.

Remarque 1.9. Note sur l'appellation « synthèse par modulation de fréquence ». D'après ce qui précède, le spectre de $e(t) = m_0 \exp(2i\pi f_p t + iI \sin(2\pi f_m t))$, lorsque f_p et f_m sont du même ordre de grandeur, peut contenir des composantes significatives dans les fréquences négatives pour des I faibles. Dans ce cas, on voit que $e(t)$ n'est pas un signal analytique et par suite la fréquence instantanée du signal *n'est pas* $f(t) = f_p + D \cos(2\pi f_m t)$. Le terme « modulation de fréquence » se réfère alors d'avantage à une méthode de mise en oeuvre pratique. Mathématiquement, il s'agirait plutôt d'une technique non-linéaire de production de spectre.

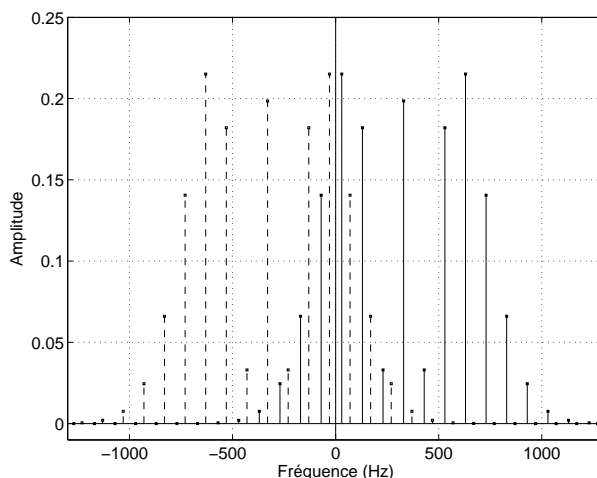


FIGURE 1.7 – Spectre de la modulation de fréquence, pour $f_p = 330$ Hz et $f_m = 100$ Hz. Les amplitudes des composantes sont représentées en module. La contribution du spectre autour de $-f_p$ est représentée en pointillés et les fréquences « réfléchies » correspondent à celles dont la fréquence est supérieure à 0.

Production de spectre « tonaux ». Pour reproduire le caractère harmonique du spectre d'une note de musique, tel qu'on peut l'observer pour de nombreux instruments, Chowning a proposé de prendre $f_m = Hf_p$, $H \in \mathbb{N}$ avec une valeur de H faible. Par exemple, pour $H = 1$ on produit des spectres harmonique à la fréquence fondamentale $f_p = f_m$. Pour $H = 2$, les spectres produits ne contiennent que des harmoniques impaires, ce qui peut permettre de réaliser des sons qui présentent un spectre dont les harmoniques paires sont fortement atténuées, tels que ceux de la clarinette.

1.2.3 Mise en oeuvre numérique

La mise en oeuvre sur ordinateur de l'instrument FM représenté sur la figure 1.3 fait appel à deux oscillateurs et un additionneur. Le pseudo-code correspondant est donné par :

Pcode 1.2: Synthèse numérique d'un instrument FM

Initialisation : on définit $m(t_n) = m_0$ et $d(t_n) = D$, $\forall n$ avec $t_n = nT_e$

Fonction $x = \text{instFM}(f_p, f_m, d, m, F_e)$

1. $x_1 = f_p + \text{oscFM}(f, d, F_e)$, avec $f(t_n) = f_m \forall n$
2. $x = \text{oscFM}(x_1, m)$.

Lorsque les fonctions m et d sont prises variant dans le temps, cela conduit respectivement à une modulation d'amplitude du signal et à un contrôle temporel de l'étendue (richesse) spectrale.

Note sur le repliement de spectre. La programmation sur ordinateur de la technique de synthèse FM manipule des séquences discrètes, obtenues par échantillonnage. Considérons le signal discret $s_d(n) = s(t_n)$. La formule 1.5 montre que $s_d(n)$ se décompose en somme de composantes sinusoïdales discrètes $\cos(2\pi\nu_k n)$ de fréquence réduite $\nu_k = \frac{(f_p + kf_m)}{F_e}$. Les composantes dont la fréquence réduite est supérieure à $\frac{1}{2}$ sont soumises au phénomène de repliement de spectre. En effet $\cos(2\pi\nu_k n) = \cos(2\pi(1 - \nu_k)n)$; et par exemple, pour $F_e = 16000$ Hz une composante à 9000 Hz sera en fait entendue à $F_e - 9000 = 7000$ Hz. La figure 1.8 montre le spectre de l'instrument FM à deux oscillateurs de la figure 1.3, tenant compte du repliement. On a choisi $f_p = 4f_m$ avec $f_m = 1047$ Hz (Do 5) et $I = 4$. Cette figure met en évidence les difficultés afférentes : les composantes repliées ne sont pas en rapport harmonique avec la fondamentale (f_m). Ces composantes sont perçues comme parasites et modifient la hauteur et le timbre perçus du son. Les précautions pour limiter ce phénomène sont

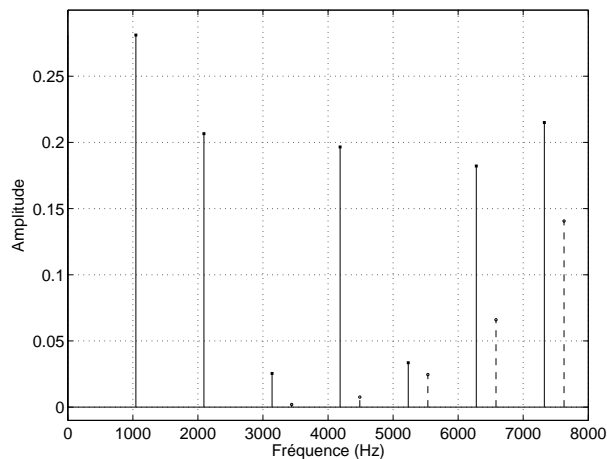


FIGURE 1.8 – Repliement de spectre occasionné par l'échantillonnage. Les composantes repliées sont en pointillés

par exemple :

- effectuer les calculs à des fréquences d'échantillonnage élevées (quitte à rééchantillonner avant d'envoyer la séquence vers le convertisseur analogique/numérique) ;
- limiter les index de modulation lorsque la note jouée est aiguë car le phénomène est d'autant plus sensible que la fondamentale est élevée.

1.3 Mise en oeuvre : comment régler les paramètres ?

La technique de synthèse FM est remarquablement économique : pour produire des spectres variant dans le temps, ou avec la nuance, le schéma simple à deux oscillateurs est efficace. La fonction $d(n)$ contrôle l'étendue spectrale ce qui permet de traduire des phénomènes acoustiques liés aux instruments de musique : le son plus « clair » d'une note de piano frappée *ff* (son plus riches en harmoniques), l'enrichissement progressif du spectre d'un cuivre lors de l'attaque ou encore l'atténuation plus rapide des aiguës que des graves dans un instrument à son libres.

Cette efficacité du modèle est atteinte en raison de la non linéarité du modèle de production. Le revers de cette médaille est la difficulté connue à inverser le modèle. Pour ajuster les paramètres du modèle, les réglages sont le plus souvent effectués empiriquement. Les paragraphes qui suivent ont pour objectif de donner des repères pour réaliser ces réglages.

1.3.1 Spectres harmoniques et inharmoniques

Le spectre des signaux obtenus par la technique de synthèse FM sont des spectres de raies aux fréquences $|f_p \pm kf_m|$, $k \in \mathbb{N}$. Une répartition quelconque des raies ne pourra donc pas toujours être obtenue par cette technique. Le rapport $r = f_p : f_m$ fixe la position des raies latérales et par suite, contrôle cette répartition.

Spectres harmoniques. Le spectre de nombreux instruments mélodiques est réparti, de manière approchée ou exacte, selon la séquence $f_0, 2f_0, \dots, Nf_0$; c'est à dire selon une série de partiels multiples du fondamental f_0 . Un tel spectre *harmonique* (ou encore *tonal*), peut-être obtenu par synthèse FM en choisissant $r = 1 : 1$. Dans ce cas la série des partiels est complète de la forme kf_p , $k \in \mathbb{N}$. Chowning propose ce rapport r pour synthétiser un instrument qui s'apparente à un cuivre. Ce type d'instrument présente en effet des partiels d'ordre pair et impair d'importance comparable. Une telle répartition harmonique peut également être obtenue par des rapports de

la forme $r = P : 1$, $P \in \mathbb{N}^*$. Le fondamental de la note est alors f_m tandis que f_p est le P -ième partiel. Cela conduit à des spectres présentant des renforcements autour de f_p .

Pour un rapport $r = 1 : 2$, le fondamental est toujours à f_p mais seule les harmoniques impaires à $(2k+1)f_p$ sont présentes dans la série. Cette propriété est utile pour approcher les spectres de certains instruments de la famille des bois tels que la clarinette. Là encore, il est possible de choisir $r = P : 2$ avec P un entier non nul impair. Chowning propose $r = 3 : 2$ pour une clarinette.

Spectre inharmoniques. Dans cette catégorie rentrent des répartitions très différentes des composantes spectrales : un piano présente une série faiblement inharmonique, où l'échelle des partiels est légèrement étirée (le k -ième partiel au dessus de kf_0) alors qu'un gong ou une cloche présente une série largement différente de la série harmonique. Une cloche mineure peut par exemple être accordée à la série $0,5 : 1 : 1,2 : 1,5 : 2 : 2,5 : 2,6 : \dots$. Pour réaliser des spectres inharmoniques pour les cloches ou les gongs, Moorer [7] utilise un rapport $r = 1 : \sqrt{2}$.

Pour essayer d'approcher le comportement légèrement inharmonique du piano, Schottstaedt [8] propose un instrument à double modulante avec des rapports du type $r = 1 : 1 + \epsilon$ et $r = 1 : 4 + \epsilon$. Le choix $r = 1 : 1 + \epsilon$ produit des spectres où la k -ième composante est à la fréquence $(k + (k-1)\epsilon)f_p$ et où les bandes latérales réfléchies donnent une composante à la fréquence voisine $(k + (k+1)\epsilon)f_p$. Les partiels seront des doublets.

Une analyse plus complète de la répartition des partiels est donnée dans [9] qui explore plus systématiquement les familles de spectres obtenues. Un compositeur peut ainsi régler $r = f_p : f_m$ de façon à générer des progressions timbrales.

1.3.2 Réglage de l'enveloppe

L'enveloppe générale du son est la fonction qui donne les variations de son amplitude en fonction du temps. Dans le schéma de synthèse FM élémentaire à deux oscillateurs, c'est la fonction $m(t)$ qui génère cette enveloppe par modulation d'amplitude.

Analyse de son réels. Pour déterminer cette enveloppe on peut par exemple analyser des sons enregistrés d'instruments en situation de jeu. Un algorithme simple consiste à redresser le signal puis à le filtrer passe-bas, suivant le pseudo-code donné par :

Pcode 1.3: Détection d'enveloppe

Initialisation :

1. $w(n)$, fenêtre de Hann de longueur $\frac{F_e}{100}$ (10 ms).
2. $h(n) = \frac{w(n)}{\sum_n w(n)}$, réponse impulsionnelle du filtre

Fonction $e = \text{env}(s, F_e)$

1. $E(n) = s(n)^2$
2. Filtrage zéro-phase par h : # permet de compenser le retard du filtre
 - (a) $x = h \star E$, $x \leftarrow x(-n)$
 - (b) $e = h \star x$, $e \leftarrow e(-n)$
3. Normalisation $e \leftarrow \sqrt{2}e$

La figure 1.9 montre le résultat obtenu lorsque l'algorithme est appliqué à un son de trompette.

Paramétrisation des fonctions d'enveloppe. Pour que nombre de coefficients à régler pour la synthèse reste faible, les fonctions de modulation d'amplitude sont paramétrées. Le plus souvent cette paramétrisation est donnée par morceaux sous forme de fonctions simples (exponentielles, droites). Pour être cohérent avec notre

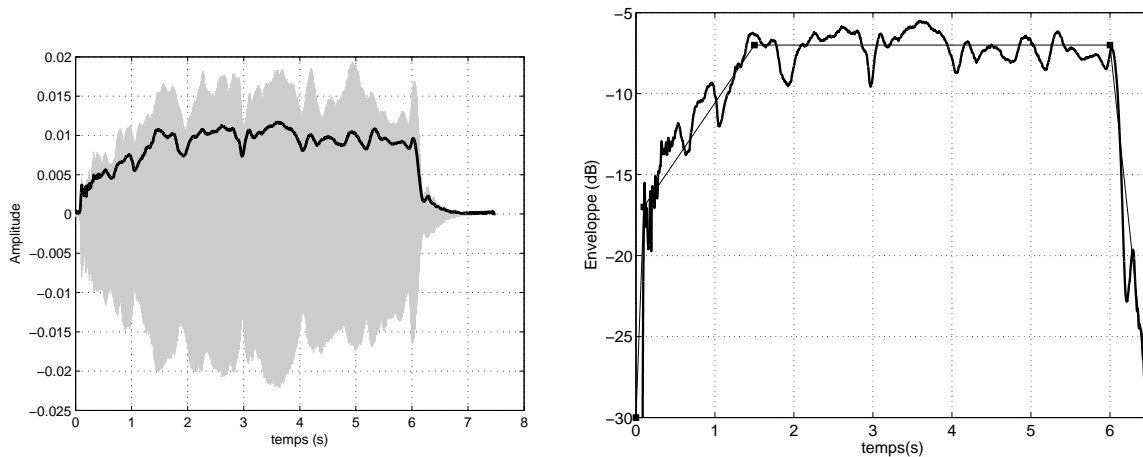


FIGURE 1.9 – Détection d’enveloppe pour un signal de trompette. (a) résultat obtenu, (b) enveloppe représentée en dB et sa paramétrisation en segments de droite.

perception, on pourra par exemple se baser sur la représentation obtenue plus haut et affichée en décibels. La figure 1.9 (b) montre une telle paramétrisation sous forme de segments de droite en dB.

Remarque 1.10. Enveloppe ADSR. Une paramétrisation courante est donné sous forme de 4 segments de droite décrivant 4 parties dans l’évolution du son : une partie d’attaque (A), une partie de décroissance (D) qui débouche sur une partie entretenue du son (le S pour *sustain*). Enfin le son s’éteint dans la dernière partie (R pour *release*), ainsi qu’illustré figure 1.10.

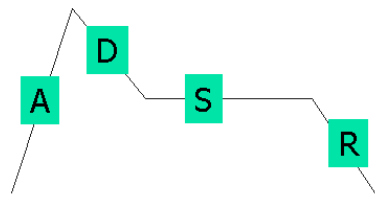


FIGURE 1.10 – Enveloppe ADSR

1.3.3 Réglage de l’indice de modulation

C’est l’indice de modulation I ou, de manière équivalente, la déviation $D = If_m$ qui règle l’enrichissement du spectre en harmoniques. Une règle simple dit que le nombre de raies latérales significatives est de l’ordre de $I + 1$. Cette enrichissement peut-être analyser à l’aide d’une représentation temps-fréquence tel que le spectrogramme de la figure 1.11, obtenu pour le son de trompette précédent.

En pointillé, on a représenté la largeur du spectre de manière approchée, ce qui permet de définir un gabarit pour la fonction $d(t_n)$ définie pour l’instrument FM (*cf.* pseudo-code 1.2). Un aspect intéressant est la similarité entre ce gabarit est la fonction d’enveloppe temporelle trouvée figure 1.9, qui trouve son explication dans les mécanismes physiques de production des sons dans les cuivres [10].

1.3.4 Compléments

Ce paragraphe donnent quelques pistes pour améliorer le rendu de la synthèse.

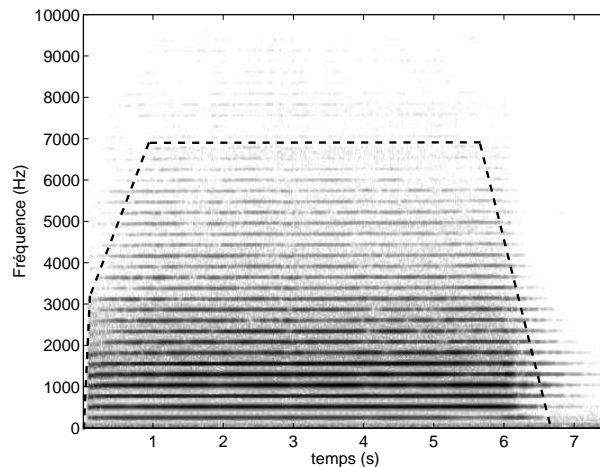


FIGURE 1.11 – Spectrogramme d'un son de trompette

Vibrato. Le vibrato des instruments de musique à cordes frottées ou de la voix humaine correspond à une modulation lente de la hauteur du son, typiquement à une fréquence comprise entre 3 et 7 Hz et n'excédant pas une excursion d'un 1/2 ton. Cette modulation de fréquence s'accompagne souvent d'une modulation d'amplitude synchronisée. Dans le cas de la synthèse FM, pour créer cet effet, on pourra moduler les fréquences porteuse et modulante soit :

$$\begin{aligned} - f_m &\leftarrow f_m(1 + a \sin(2\pi f_v t)) \\ - f_p &\leftarrow f_p(1 + a \sin(2\pi f_v t)) \end{aligned}$$

où f_v est la fréquence du vibrato (3 à 7 Hz) et a l'excursion (typiquement 0.03 pour une amplitude de ± 1 quart de ton). a pourra être modulé dans le temps pour que le vibrato intervienne après la phase d'attaque de la note. Pour obtenir une variation d'amplitude synchronisée, on peut appliquer une modulation d'amplitude similaire en sortie soit :

$$- m(t) \leftarrow m(t)(1 + a_m \sin(2\pi f_v t + \phi_m))$$

Variations aléatoires. Pour les instruments à vent, l'interaction entre le conduit vocal et la colonne d'air introduit un facteur humain dans la production du son et notamment de sa hauteur. Cela conduit à des variations aléatoires de faible amplitude de cette hauteur. De même l'amplitude sonore sera soumise à des aléas, qu'il est possible de constater pour un son de trompette sur la figure 1.9. Les générateurs de nombres pseudo-aléatoires permettent de fabriquer des signaux aléatoires. On pourra par filtrage passe-bas obtenir ainsi des fonctions aléatoires présentant des variations plus lentes.

Pcode 1.4: Génération d'un Jitter (variation de hauteur) aléatoire

1. Génération d'un bruit blanc Gaussien de variance 1 : $b(n) \sim \mathcal{N}(0, 1)$
2. Filtrage du bruit par un passe-bas de fonction de transfert de la forme $H(z) = \frac{\sqrt{(1-a^2)}}{1-az^{-1}}$ où a est un réel compris entre 0 et 1. On trouve $y(n) = \{h \star b\}(n)$ # y est un processus aléatoire de variance 1
3. Ajout du Jitter : # on pourra choisir σ de l'ordre de 0.5%
 - $f_p \leftarrow f_p(1 + \sigma y)$
 - $f_m \leftarrow f_m(1 + \sigma y)$

Effet de réverbération. Il existe de nombreuses techniques pour simuler l'effet d'une salle [11]. Les plus simples sont basées sur des mise en oeuvre récursives de ligne à retard et de filtres passe-tout pour rendre l'effet des échos successifs. Une mise en oeuvre directe et simple consiste cependant à convoluer directement le signal de

synthèse par une réponse impulsionnelle de salle. Cette opération est effectuée souvent au prix d'une complexité accrue (les réponses impulsionnelles de salle peuvent être longues). Soit h_s la réponse impulsionnelle de la salle utilisée, appliquer l'effet de réverbération consiste alors en :

$$- s \leftarrow h_s \star s$$

Cuivrage. Le phénomène de cuivrage, caractéristique du timbre de la trompette ou du trombone par exemple, correspond à une vitesse de propagation des ondes différentes pour les hautes et les basses pressions acoustiques. Une manière simplifiée de simuler cet effet sur un son préexistant est de d'appliquer un retard dépendant de l'amplitude instantanée. La figure 1.12 montre le résultat d'un tel effet sur une onde sinusoïdale.

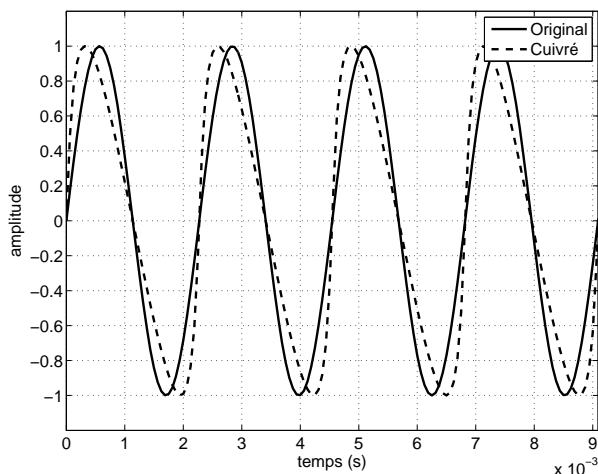


FIGURE 1.12 – Cuivrage d'une onde sinusoïdale

La littérature sur des effets non-linéaires similaires [12, 10, 13] propose plusieurs pistes de mise en oeuvre. Le lecteur est invité à s'y référer pour un abord plus complet mais le pseudo-code d'une réalisation simple est donné ci-dessous.

Pcode 1.5: Cuivrage simple d'un son

Fonction $x_c = \text{cuivrage}(x, \beta)$

β désigne le maximum du retard en nombre d'échantillons

1. Filtrage passe-bas du signal (optionnel) : $y = h \star x$ ($y = x$ si pas de filtrage)
2. Rééchantillonnage de y d'un facteur P ($P = 8$ par exemple) : $\rightarrow y_r$
et renormalisation $y_r \leftarrow \frac{y_r}{\max(|y_r|)}$
y_r est de longueur PN si x est de longueur N
3. Rééchantillonnage de x d'un facteur P : $\rightarrow x_r$
4. Calcul des temps $t_n = n - \beta y_r(n)$ où sont réaffectés les échantillons de x_r
le signal avec effet est maintenant la fonction (t_n, x_r) , mais les t_n ne sont plus régulièrement distribués
5. Interpolation de x aux instants $\tau_k = kP$: $\rightarrow x$
revient à l'échantillonnage initial (facteur P de sous-échantillonnage)
l'interpolation peut-être faite avec des méthodes variées (linéaire, splines, Lagrange...)

1.4 Mise en oeuvre : exemples

Dans cette partie, quelques systèmes sont décrits graphiquement pour réaliser des sons musicaux. La difficulté de s'approcher des sons réels est connue mais, à l'aide des remarques précédentes, il est possible de chercher

à générer des timbres que notre perception apparente à ceux d'instruments réels. Les schémas donnés sont considérés ici comme des bases de travail à améliorer et à ajuster pour obtenir les sons cibles.

1.4.1 Schéma de base pour instrument à vent

La figure 1.13 regroupe les diverses remarques des paragraphes précédents.

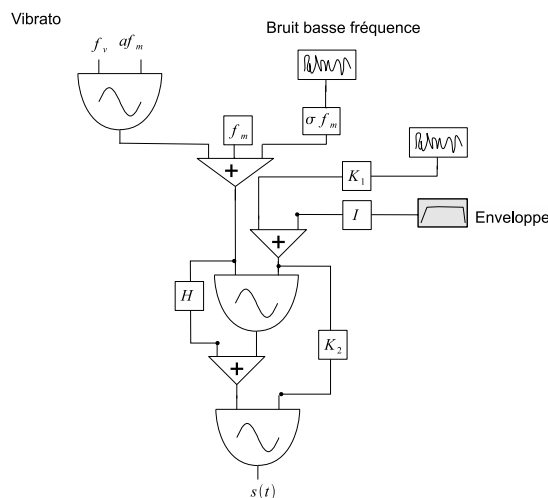


FIGURE 1.13 – Schéma de base pour instrument à vent

Trompette. Pour régler le schéma de base pour un cuivre tel que la trompette, il est possible de prendre $H = 1$ ou $H = 3$ ou 4 . On pourra alors combiner plusieurs schémas de manière à renforcer le spectre dans les zones situées autour du troisième ou du quatrième harmonique. D'autres pistes d'amélioration du schéma sont :

- ajout de bruit aléatoire au signal obtenu et mise en forme spectrale à l'aide d'un filtre en sortie
- un schéma à double oscillateur pour enrichir le vibrato (un vibrato sinusoïdal sonne en général artificiel)
- la partie destinée à générer un vibrato peut être reconduite pour moduler l'enveloppe générale du son et conduire ainsi à une modulation d'amplitude synchrone avec la modulation de fréquence liée au vibrato.

Clarinette. Pour une clarinette, le renforcement des harmoniques impaires peut être simulé en prenant $2f_p = f_m$ soit $H = \frac{1}{2}$ avec une fréquence fondamentale à $f_0 = f_p = \frac{f_m}{2}$.

1.4.2 Piano FM

Le piano FM est réalisé ici selon une idée proposée par Schottstaedt [8] d'une double modulante. Pour représenter le bruit mécanique lié à la frappe de la touche sur le bâti et aux résonances de la table d'harmonie lors de l'attaque, un générateur de bruit basse fréquence a été ajouté.

Remarque 1.11. Un tel schéma ne peut réellement approcher le timbre d'un piano acoustique. Les timbres obtenus ont leur propre spécificité, et ils ont acquis leur identité sonore par une utilisation importante. Il n'est pas rare de retrouver ces sons dans des banques de sons de synthèse sous une appellation telle que « piano FM ».

1.4.3 Cloches et gong

Le schéma présenté ici permet de réaliser un gong de synthèse en prenant un f_0 suffisamment grave (typiquement 70-200 Hz). Pour approcher un son de cloche mineure telle qu'utilisée dans les carillons, on pourra

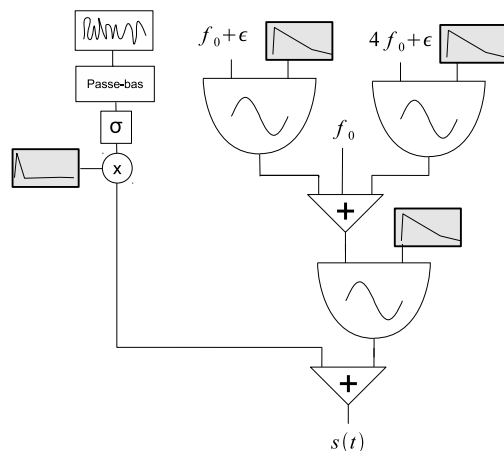


FIGURE 1.14 – Schéma de base pour un piano FM

utiliser l'addition de deux montages à deux oscillateurs avec $f_p : f_m = 1 : 1$ avec un avec un rapport 1.2 entre les modulantes du premier et du deuxième montage.

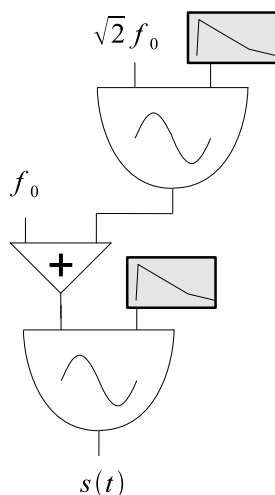


FIGURE 1.15 – Schéma de base pour un instrument inharmonique de type gong ou cloche

1.5 Conclusion

Cette partie explique les principes fondamentaux qui régissent la synthèse sonore par modulation de fréquence. Le rôle du paramètre d'index de modulation est notamment étudié et montre que l'enrichissement du spectre en harmoniques est contrôlé par ce paramètre. Quelques illustrations sur des schémas de synthèse sont données. Une extension naturelle de ce travail serait de réaliser des instruments de synthèse complet, c'est à dire d'ajuster le comportement des enveloppes, des index de modulation et des fréquences sur toute une tessiture. Une autre question intéressante est de spécifier le lien entre la nuance jouée et les variations des paramètres. Le spectre du piano par exemple est différent lorsque la note est jouée *forte* (il est plus riche) et lorsqu'elle est jouée *piano*. Enfin, les schémas proposés ici restent simples et il existe dans la littérature et dans le commerce

des schémas plus complexes utilisant des cascades de plus de 2 oscillateurs (typiquement jusqu'à 6) mis en série ou en parallèle et des schémas utilisant des oscillateurs bouclés.

Toutefois, la synthèse par modulation de fréquence est connue pour être peu appropriée pour s'approcher d'un son réel. Le système non-linéaire de production ne s'inverse pas aisément. En revanche, il propose un espace timbral à explorer et ce, en manipulant peu de paramètres de contrôle.

Bibliographie

- [1] M. MATHEWS, « The digital computer as a musical instrument », *Science*, vol. 142, no. 3592, p. 553, 1963. 1.1
- [2] J. CHOWNING, « The synthesis of complex audio spectra by means of frequency modulation », vol. 21, no. 7, p. 526–534, 1973. 1.1, 1.2.2
- [3] M. PUCKETTE, « Pure data : another integrated computer music environment », *Proceedings of the Second Intercollege Computer Music Concerts*, p. 37–41, 1996. 1.1
- [4] M. PUCKETTE, *The theory and technique of electronic music*. World Scientific Pub Co Inc, 2007. 1.1
- [5] B. PICINBONO, « On instantaneous amplitude and phase of signals », *IEEE Transactions on Signal Processing*, vol. 45, no. 3, p. 552–560, 1997. 1.2.1, 1.2.1
- [6] C. GASQUET et P. WITOMSKI, *Analyse de Fourier et applications : filtrage, calcul numérique, ondelettes (coll sciences sup)*. Paris : Dunod, 2000. 1.2.1
- [7] J. A. MOORER, « Signal processing aspect of computer music », vol. 65, Aug 1977. 1.3.1
- [8] B. SCHOTTSTAEDT, « The simulation of natural instrument tones using frequency modulation with a complex modulating wave », *Computer Music Journal*, vol. 1, no. 4, p. 46–50, 1977. 1.3.1, 1.4.2
- [9] B. TRUAX, *Foundations of Computer Music*, chap. Organizational Techniques for c :m Ratios in Frequency Modulation, p. 68–82. MIT Press, 1985. 1.3.1
- [10] J. GILBERT, L. MENGUY et M. CAMPBELL, « A simulation tool for brassiness studies », *The Journal of the Acoustical Society of America*, vol. 123, p. 1854, 2008. 1.3.3, 1.3.4
- [11] U. ZÖLZER et X. AMATRIAIN, *DAFX : digital audio effects*. John Wiley & Sons Inc, 2002. 1.3.4
- [12] C. M. COOPER, J. S. ABEL et N. ERDOL, « Analysis and digital enhancement of "brassiness". », *The Journal of the Acoustical Society of America*, vol. 128, no. 4, p. 2310–2310, 2010. 1.3.4
- [13] T. TOLONEN, V. VALIMAKI et M. KARJALAINEN, « Modeling of tension modulation nonlinearity in plucked strings », *Speech and Audio Processing, IEEE Transactions on*, vol. 8, no. 3, p. 300–310, 2000. 1.3.4