

Support de formation : Synthèse sonore

1 Introduction

La synthèse sonore regroupe un ensemble de procédés et d'algorithmes destinés à produire des sons musicaux. Dans les applications musicales, ces algorithmes sont commandés par un ensemble de paramètres de contrôle.

L'ambition de ce document est de proposer un rapide survol des méthodes et quelques clefs pour en sélectionner une en fonction d'une application visée. Des références intéressantes pour prolonger sont données dans (Roads 98; Kahrs 98).

Les différentes méthodes de synthèse sont ici présentées en suivant une taxonomie qui reprend certaines des catégories définies par Smith (Smith 91). Ces méthodes sont ici regroupées selon qu'elles reposent sur :

1. un *modèle de signal* : on construit des *formes d'onde* qui sont ensuite utilisées pour fabriquer le son. Exemples :
 - synthèse additive : le son est reconstruit en additionnant des composantes élémentaires comme les partiels d'un son de cloche ;
 - synthèse granulaire : c'est un cas particulier de synthèse additive mais cette fois les formes élémentaires sont localisées en temps et en fréquence ;
 - synthèse à table d'onde (dite également "par échantillonnage") : le son est reconstruit par contraction ou dilatation en temps et/ou fréquence de formes d'onde élémentaires contenues en mémoire ;
2. un *modèle de production* : on choisit les paramètres et la structure d'un système qui produit le son. Exemples :
 - synthèse par modulation de fréquence (FM) : on connecte des opérateurs élémentaires tels que oscillateurs, additionneurs, gains et tables de lecture pour produire le son. La cascade de deux oscillateurs sinusoïdaux produit une configuration élémentaire de synthèse FM ;
 - synthèse source/filtre ;
3. un modèle physique de fonctionnement d'un instrument. Cette synthèse, dite *par modèle physique*, vise à reproduire le comportement réel de l'instrument via une description mathématique en terme de source, d'équations d'évolution (ou de propagation) et de conditions aux limites.
 - synthèse par discrétisation des équations de propagation (différences finies ou éléments finis), description physique de l'interaction avec la source et simulation du rayonnement acoustique ;
 - synthèse par lignes de transmission : adaptées aux propagations mono-dimensionnelles comme les cordes (méthode de Karplus-Strong) ou les instruments à vent (tubes acoustiques).

Certains algorithmes de synthèse évolués font appel à plusieurs techniques mises en connexion, comme par exemple le logiciel CORDIS-ANIMA (ACROE, Grenoble) qui permet d'assembler empiriquement divers éléments physiques élémentaires ou encore MODALYS (Ircam, Paris) qui permet de travailler avec des modèles physiques de structures mécaniques (cordes, tubes, plaques) décrits par analyse modale.

Compromis entre précision et versatilité. Comment juger de la qualité d'une synthèse ? Des facteurs très différents rentrent en jeu : esthétique sonore, facilité d'utilisation et d'exploration de l'espace sonore, capacité à générer un discours musical riche... Dans la littérature, il est souvent fait mention de deux critères principaux : la capacité à se rapprocher perceptivement d'un son cible réel ou imaginé, ce que nous appellerons ici *précision* et d'autre part, la capacité de la synthèse à produire des sons diversifiés lorsqu'on fait varier les paramètres de contrôle, comme par exemple la vitesse (force d'enfoncement). Ce dernier critère, que nous dénommerons *versatilité* est relié à la variabilité spectro-temporelle des sons produits. La plupart des synthèses ne parviennent pas à souscrire simultanément à ces deux critères. La synthèse additive par exemple, peut atteindre une excellente en précision mais elle reste en général peu versatile (le contenu spectral ne varie que peu avec la force d'excitation). En revanche la synthèse FM est réputée pour produire des spectres rapidement variables avec l'amplitude et la possibilité de fabriquer des sons inouïs dans le monde réel mais elle reste inopérante pour produire un son de piano acoustique réaliste. La synthèse par modèle physique contourne ce compromis expressivité/précision mais cette double performance est obtenue au prix d'une complexité accrue. Cependant cette complexité est de moins en moins hors de portée de la puissance des micro-ordinateurs actuels. C'est pourquoi les "softsynths" actuels s'intéressent à ce type de technique qui est un des fer de lance de la recherche scientifique sur le sujet.

Historiquement, la synthèse FM et la synthèse par table d'onde sont les plus anciens et le plus développés commercialement. La première parce qu'elle produit des sons riches et expressifs pour une faible complexité de calcul, la deuxième en raison de l'abaissement des coûts de la mémoire. Une tendance des "softsynths" actuels est de simuler les anciens synthétiseurs analogiques (Moog, Roland Jupiter, Prophet 5) en proposant des modules élémentaires interconnectables et une variété d'effets (filtres, boucles, ...). A côté de ces free (WaveCraft) ou sharewares (VAZ, Virtual Waves, Reaktor, Rebirth), on trouve d'autres logiciels qui intègrent une partie "modèles physiques" (Tassman, Attack).

2 Contrôle de la synthèse

Parmi les questions qui se pose au manipulateur de son et au créateur, celle du contrôle du système tient une place importante. De nombreuses propositions existent, depuis celles liées à la création contemporaine impliquant un contrôle gestuel via la captation du mouvement d'un danseur par exemple au très généralisé, langage MIDI (Musical Instrument Digital Interface). Un rapide survol de ses caractéristiques est fait ci-dessous.

Langage midi Le protocole MIDI est sans doute le plus répandu des langages de commande d'unités de synthèse musicale électronique. Il est largement utilisé depuis sa conception au début des années 80, à la fois par les musiciens et les compositeurs, mais aussi par les applications multimédia et les jeux vidéo, pour représenter, coder, l'interprétation d'une oeuvre musicale.

Un pas décisif fut accompli par la publication du système GM (Général Midi) qui permit de lutter contre la non-standardisation des choix effectué par les constructeurs de synthétiseurs.

Ce qu'est le MIDI, ce qu'il n'est pas. Un des arguments les plus souvent employés en faveur du MIDI est la légèreté des fichiers obtenus, de l'ordre de 1000 fois moins encombrants que leur équivalent sous forme d'onde échantillonnée (en PCM à 44.1kHz par exemple). Faire cette comparaison n'a cependant, en général, pas de sens. Le MIDI *ne contient pas de représentation numérique des sons.*

C'est un langage de commande, qui instruit le synthétiseur sur quel instrument doit jouer, quelles notes il doit générer, quand et comment il doit les générer. C'est une forme de partition généralisée. C'est le synthétiseur qui produit le flux sonore correspondant.

Technologie et terminologie MIDI. Le MIDI est un flux de données asynchrone, fonctionnant par paquets de 10 bits à 31.25 Kbit/s. La connexion entre unités (de commande et/ou de production) est réalisée par *une interface MIDI* (connecteur din 5 broches) qui comprend généralement 3 connecteurs : IN, OUT et THRU. L'unité de commande peut par exemple être un clavier, capable de produire en temps réel, la traduction des commandes physiques reçues en flux de donnée midi. On parle alors de *contrôleur MIDI*. Via la prise OUT de ce contrôleur, le flux est acheminé vers le connecteur IN d'une unité de production des sons, capable d'interpréter ce flux de *messages MIDI* : un *séquenceur* (cf. figure 1).

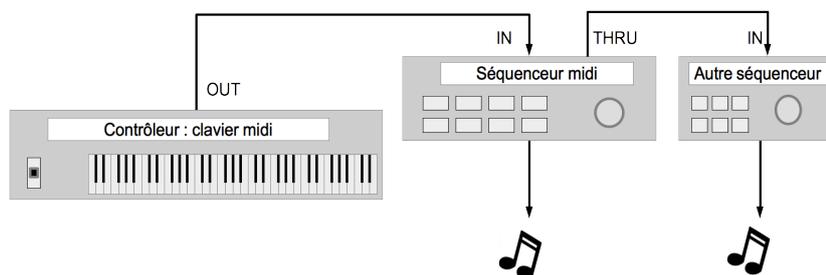


FIGURE 1 – Système MIDI générique

Le canal MIDI physique est divisé en 16 canaux logiques, permettant de commander simultanément plusieurs séquenceurs via le connecteur de réplication MIDI THRU. La plupart des séquenceurs sont cependant multitimbraux, c'est à dire capable de recevoir et jouer plusieurs canaux midi simultanément.

Principaux messages MIDI. Un message MIDI est constitué de deux ou trois paquets de 8 bits, le premier codant le type de l'instruction et les autres les paramètres afférents. Ces messages sont de deux types : les messages système et les message de canal. Ces derniers sont spécifiquement adressé à un canal MIDI spécifique dont le numéro est inclus dans le message.

Parmi ces derniers, 2 seront cités ici (la description exhaustive du langage dépasse le cadre de ce cours), qui définisse de manière minimale une pièce musicale : le "Note On" et le "Note off". Ils spécifient respectivement le point de départ et d'extinction de la note. Le "Note On" précise sa hauteur (numéro midi, ou "key number" avec $La4=69$), la force d'enfoncement de la note ("velocity") et le canal midi affecté.

3 Méthodes de synthèse reposant sur un modèle de signal

3.1 Synthèse à table d'onde

Avec la synthèse FM, ce type de synthèse est le plus utilisé commercialement, présent sur la plupart des cartes son d'entrée de gamme du commerce. Son principe est celui d'un dictionnaire de formes

d'onde élémentaires, comportant une ou plusieurs périodes et stockées en mémoire morte. L'utilisation courante de ces tables d'onde consiste alors à les relire à l'aide d'un index (souvent appelé phase) fractionnaire cyclique pour réaliser la transposition de hauteur.

3.1.1 Enregistrement des formes d'onde et repliement

La constitution d'une table d'onde de $N = 256$ échantillons pour obtenir un signal carré doit se prémunir contre le repliement spectral. La solution simple

$$x(n) = \mathbf{1}_{[0 \ N/2]}(n) - \mathbf{1}_{[N/2+1 \ N]}(n)$$

en comporte puisque la série de Fourier de cette forme d'onde présente des raies aux harmoniques n impaires pour $n \in \mathbb{N}^+$. Les harmoniques de rang $n \geq 125$ sont repliées. Ce repliement peut paraître faible (il correspond à des amplitudes d'harmonique inférieures d'au moins 40 dB à celle du fondamental) mais il peut devenir fortement audible lorsque hauteur de la note jouée augmente.

Plusieurs solutions coexistent pour pallier ce problème :

- les formes d'ondes sont enregistrées en limitant le nombre d'harmoniques (souvent on limite leur fréquence réduite à la valeur maximale de $1/4$),
- on reconstruit la sortie en interpolant les échantillons de manière à approximer la formule de reconstruction (Shannon) ou, au moins à réaliser un filtrage qui limite le repliement lors du rééchantillonnage.

3.1.2 Transposition de hauteur et interpolation

Procédé asynchrone. Une solution simple quoique que coûteuse en matériel consiste à attribuer un CNA par voies (par note). Ce procédé est dit asynchrone car les horloges de chaque voies sont alors indépendantes. Si cette solution a été appliquée dans le passé (jusque vers le milieu des années 80) elle est abandonnée maintenant. En effet, elle souffre de plusieurs désavantages : on ne peut pas faire de post-traitement en numérique sur la sortie puisque le mixage des voies est réalisé en analogique et d'autre part, toutes les caractéristiques temporelles sont rendues proportionnelles à la période de la note jouée. Par exemple les attaques vont dépendre de la note ce qui est peu réaliste dans de nombreux cas.

Interpolation Ce procédé consiste à rééchantillonner la table d'onde avant de l'envoyer vers le CNA. De cette manière le mixage et les effets peuvent être réalisés en numérique et un seul CNA est nécessaire par canal analogique.

Les procédés classiques en traitement du signal numérique de sous/sur-échantillonnage (L -interpolateur + filtre + M -décimateur) ne sont pas envisageables en général en raison de la complexité requise : les valeurs de L et M ainsi que la synthèse du filtre doivent être modifiées à chacune des notes jouée.

Les procédés les plus couramment utilisés pour réaliser cette fonction sont les interpolations de la table d'onde. Aucun filtrage n'est réalisé préalablement et par conséquent, la décimation effectuée provoque du repliement.

Phase incrémentale fractionnaire L'interpolation la plus simple consiste à parcourir la table d'onde, à l'aide d'un index cyclique (modulo la longueur de la table), incrémenté par une valeur fixe fractionnaire, correspondant au rapport souhaité des périodes. La numérisation impose d'utiliser la partie entière de cet index. L'opération réalisée est alors un blocage d'ordre 0. L'effet spectral de blocage est connu : modulation du spectre par la fonction $T_e \text{sinc}(\pi f T_e)$. Le signal reconstitué en sortie s'écrira alors :

$$\tilde{x}(t) = \sum_{n=-\infty}^{+\infty} x(n)\delta(t-n) \star \mathbf{1}_{[0T_e]}(t) \quad (1)$$

Interpolation linéaire l'interpolation linéaire se traduit par un rééchantillonnage du signal qui n'est pas à bande limitée! En effet, interpoler linéairement entre les échantillons du signal revient à convoluer le signal échantillonné par une fonction triangulaire $w(t)$ de hauteur 1 et de largeur deux périodes d'échantillonnage. Là encore on a :

$$\tilde{x}(t) = \sum_{n=-\infty}^{+\infty} x(n)\delta(t-n) \star w(t) \quad (2)$$

Cette opération donne en sortie un signal à temps continu $\tilde{x}(t)$ dont le spectre est le spectre du signal original périodisé (à cause de l'échantillonnage), et multiplié par la transformée de Fourier de la fonction triangulaire, un sinus cardinal au carré. Le signal à temps continu $\tilde{x}(t)$ est ensuite ré-échantillonné avec une période d'échantillonnage $(1-\alpha)\Delta T$ et cela sans filtrage passe-bas préalable. Ce ré-échantillonnage qui n'est pas à bande limitée produit donc un repliement spectral mais moindre que dans le cas de l'interpolation par phase fractionnaire en raison des lobes secondaires plus faible du sinc^2 . Pour les signaux audio dont le spectre est principalement localisé en basse-fréquence, ce repliement spectral ne pose pas de problème majeur (Rossum 93), toutefois on prend souvent la précaution de fabriquer des formes d'onde dont la bande de fréquences d'énergie significative est situé en deçà de la fréquence réduite 0.25.

3.2 Synthèses par sommation

Ce type de synthèse consiste à sommer des composantes élémentaire. Le qualificatif d'"additive" est parfois d'avantage employé lorsque les motifs élémentaires sont sinusoïdaux. La synthèse granulaire et la synthèse à forme d'onde formantiques (FOF) sont d'autres exemples de synthèse par sommation.

3.2.1 Synthèse additive

C'est l'une des plus anciennes méthodes de synthèse dans laquelle le signal est obtenu en additionnant des sinusoïdes contrôlées dynamiquement en fréquence, en amplitude et en phase. C'est une méthode simple, mais coûteuse en temps de calcul : un son de piano peut contenir plus de 150 harmoniques (notes basses) qu'il est nécessaire de synthétiser une à une. Cependant, la transformée de Fourier inverse permet d'accélérer le processus de synthèse : à partir des valeurs des fréquences et amplitudes à synthétiser, on génère le spectre de Fourier, puis par transformée inverse le signal synthétique. D'autre part, le nombre de paramètres de contrôle (amplitude et fréquence pour chaque sinusoïde) devient très élevé pour un son complexe, rendant sa manipulation difficile.

L'analyse correspondant à la synthèse additive peut être effectuée par transformée de Fourier, ou à

l'aide de méthodes d'estimation spectrale paramétrique (ARMA, Prony, Matrix Pencil etc... (Laroche 93))

3.2.2 Synthèse granulaire et à FOF

La synthèse granulaire consiste à utiliser des motifs élémentaires localisés en temps et en fréquence. Ces motifs prennent la forme d'ondelettes. Les compositions qui utilisent ce type de synthèse prennent alors l'aspect de nuages de points du plan temps/fréquences, représentés sous forme de grains.

La synthèse par Formes d'Onde Formantiques s'appuie sur les résultats obtenus en traitement de la parole pour réaliser des synthèses vocales et notamment de voix chantée très performantes. Le conduit vocal est assimilé à un filtre, dont les résonances (ou formants) modulent le spectre de raies des productions vocales voisées. Le principe de la synthèse par FOF est de reproduire *dans le domaine temporel* l'effet de ce filtre en construisant sa réponse impulsionnelle par superposition des réponses élémentaires de chacun des filtres formantiques.

4 Méthodes de synthèse reposant sur un modèle de production

4.1 Synthèse FM

Cette méthode de synthèse, décrite initialement dans (Chowning 73), est basée sur le principe de la *modulation de fréquence*. Elle est très répandue notamment via les composants Yamaha qui lui sont dédiés (OPL2, OPL3, OPL4) implantés sur de nombreuses cartes son grand public.

4.1.1 Principe

À l'origine la modulation de fréquence fut utilisée dans les systèmes de communication radio. La porteuse est alors une onde électromagnétique dont la fréquence (≈ 100 MHz) est bien au-delà de la limite supérieure des fréquences audibles. L'idée de Chowning est d'étendre cette technique pour une porteuse et une modulante dont les fréquences sont du même ordre de grandeur et plus particulièrement, sont dans un rapport simple (1 :1, 1 :2, 1 :4, etc...).

Son écriture mathématique est celle d'une modulation de phase :

$$\phi(t) = 2\pi f_p t + I \sin(2\pi f_m t)$$

Il est à noter que cette modulation de phase correspond à la fréquence instantanée

$$f_i(t) = f_p + I f_m \cos(2\pi f_m t) = \frac{1}{2\pi} \frac{\partial \phi}{\partial t}$$

Le signal temporel obtenu est donné ci-dessous, ainsi que son développement en série trigonométrique :

$$\begin{aligned}
x(t) &= A \sin(\phi(t)) \\
&= AJ_0(I) \sin(2\pi f_p t) \\
&\quad + A \sum_{n=1}^{+\infty} J_n(I) \sin(2\pi(f_p + n f_m)t) \\
&\quad + A \sum_{n=1}^{+\infty} J_n(I) (-1)^n \sin(2\pi(f_p - n f_m)t)
\end{aligned} \tag{3}$$

L'expression 3 permet de comprendre l'intérêt du procédé : génération de spectres riches et rapidement variables (il suffit de moduler I) avec seulement deux oscillateurs et la possibilité de créer des spectres harmoniques donc pouvant être assimilé à des sons musicaux.

On note que l'amplitude de la n-ième raie latérale est réglée par la fonction de Bessel de première espèce d'ordre n. Ces fonctions sont représentées figure 2 pour n=0,1 et 5.

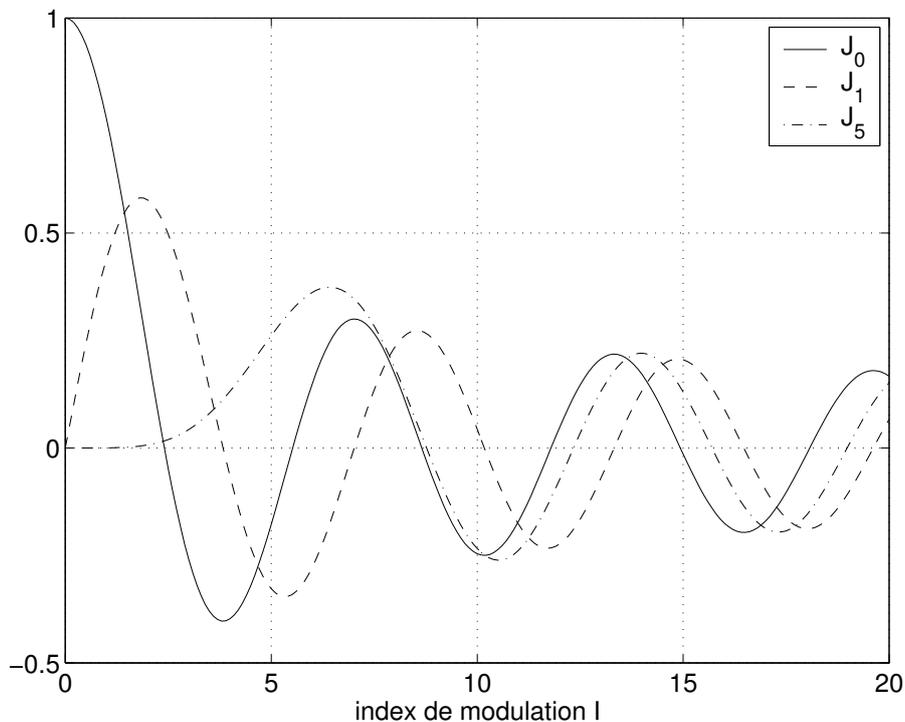


FIGURE 2 – Fonction de Bessel de première espèce pour n=0,1 et 5

De l'expression 3 et de la représentation 2 découlent plusieurs remarques :

- Les fonctions $J_n(I)$ ressemblent à des ondulations amorties et retardées d'autant plus que n est grand. Les bandes fréquentielles latérales seront donc d'autant plus larges que l'index I est grand et l'amplitude des raies "en bout de bande" tend vers zero.
- lorsque les bandes latérales sont suffisamment larges, les composantes sont reflétées (repliées) par rapport à la fréquence nulle. Si elle sont en opposition de phase avec les composantes non

reflétées, elle peuvent les atténuer ou les annuler. Dans le cas d'une implantation numérique, ce phénomène est également présent autour de la fréquence de Nyquist.

On peut toutefois remarquer que lors de l'implémentation numérique le repliement autour de la fréquence de Nyquist est en général gênant : il produit des spectres dont on ne contrôle plus l'harmonicité.

Le schéma de la figure 3 montre la variation du spectre avec l'index de modulation.

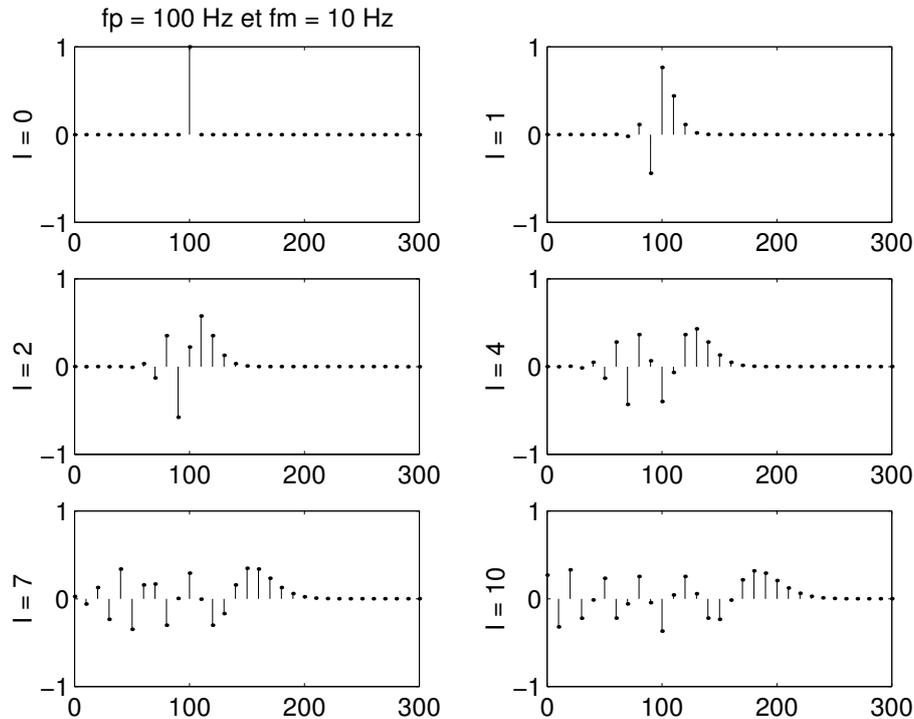


FIGURE 3 – Spectre de la modulation fm simple (sinus-sinus) en fonction de l'index de modulation et tenant compte de la phase des composantes.

4.1.2 Construction d'un système de synthèse fm

Une manière classique de représenter les systèmes est de les décomposer en *opérateurs* interconnectés. Un opérateur comprend en général un élément *oscillateur* principal alimenté par deux entrées — enveloppe du son et fréquence instantanée —. Cet oscillateur est paramétré par sa forme d'onde (sinus, carré, dent de scie, etc...) et sa fréquence centrale (porteuse). Sa représentation schématique est donnée figure 4. Le chip OPL3 de Yamaha rend possible l'interconnection de 4 opérateurs selon 4 schémas possibles : 1-2-3-4, 1 + 2-3 + 4, 1-2 + 3-4, 1 + 2-3-4.

L'enveloppe du son est décrite habituellement par quatre segments : attaque (A), décroissance (D), sustain (S) et extinction (R pour "release"). Ce qui est traditionnellement représenté sous la forme donnée figure 5. Si cette représentation est commode et très utilisée, il faut garder à l'esprit que les segments sont en fait des branches exponentielles de la forme a^n , $a < 1$ pour les parties décroissantes et $1 - a^n$ pour l'attaque.

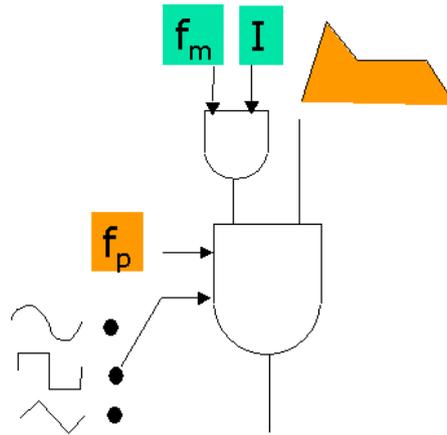


FIGURE 4 – Représentation schématique d'un oscillateur

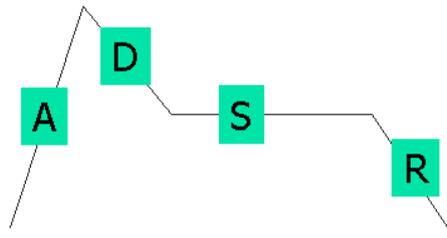


FIGURE 5 – Représentation d'une enveloppe ADSR

Application musicale. La synthèse FM est célèbre pour sa capacité à produire des sons cuivrés. On utilise pour cela un rapport $f_p : f_m$ égal à 1 :1 et un index compris entre 0 et 7. Le rapport 1 :2 produit des harmoniques impaires qui sont caractéristique des sons boisés tel que la clarinette ou le saxophone.

4.1.3 Possibilités étendues

2 compléments étendent les possibilités de la synthèse FM telle que décrite ci-dessus.

- la FM multiporteuses. L'utilisation de plusieurs opérateur dont les porteuses respectives ne sont pas aux mêmes fréquences permet de mieux contrôler l'enveloppe spectrale, chacun des oscillateurs produisant une partie plus spécifique de ce spectre ;
- le bouclage des oscillateurs. Ce bouclage s'opère d'un opérateur sur lui-même ou selon des schémas cascades plus complexes. Son effet est une "régularisation" timbre.

4.2 Synthèse par distortion non-linéaire ou *waveshaping*

L'idée principale sur laquelle repose cette méthode de synthèse consiste à transformer un forme d'onde simple, telle qu'une sinusoïde par exemple à l'aide d'une fonction non-linéaire. Expérimentée à l'origine par Jean-Claude Risset en 1969 (Kahrs 98), les bases mathématiques de cette technique ont été posée par Arfib (Arfib 79) et Le Brun (Le Brun 79).

Le principe de base, dans sa forme la plus concise, consiste à appliquer une distortion non-linéaire w , définie pour chacune des valeurs possibles de la forme d'onde d'entrée, tels que les exemples donnés figure 6. L'intérêt est de produire des spectres enrichis puis que la non-linéarité produit des harmoniques

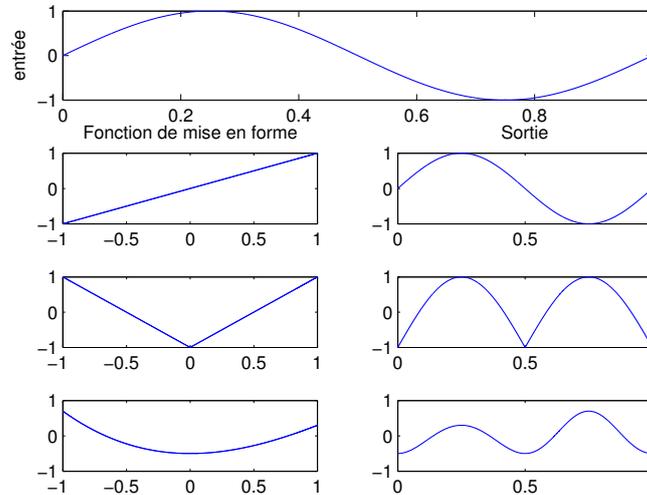


FIGURE 6 – Principe de la distortion non-linéaire, les 3 distortions sont respectivement l'identité, $|x|$ et $x^2 - 0.2x^3$.

multiples de la fréquence de la sinusoïde d'entrée (ou des harmoniques d'intermodulation si plusieurs composantes alimentent le système) avec une grande économie de moyens. En effet cette mise en forme peut être facilement mise en oeuvre en lisant des tables enregistrées préalablement en mémoire.

Les deux auteurs cités précédemment ont observé de plus que le degré d'enrichissement pouvait être contrôlé en utilisant des polynômes de Chebyshev comme fonctions de distortion. Ceux-ci présentent en effet l'intéressante propriété suivante : le polynôme d'ordre n appliqué à une sinusoïde de fréquence f_0 produit une sinusoïde pure de fréquence nf_0 . En définissant ainsi la distortion à l'aide d'une combinaison linéaire de polynômes de Chebyshev, l'importance relative de chaque harmonique peut-être contrôlée.

Cette méthode est souvent combinée à un post-traitement de modulation d'amplitude. Ce qui permet, soit d'enrichir le spectre autour des harmoniques, soit, lorsque la fréquence de modulation est suffisamment grande, de produire du repliement et des sons inharmoniques (Arfib 79).

4.3 Synthèse soustractive ou source-filtre

En synthèse source-filtre, l'onde de sortie est obtenue en appliquant un filtre variant dans le temps à une excitation prédéfinie. Elle est inspirée du codage et de la modélisation de la parole (Moorer 79). Un modèle simplifié de production de la parole consiste en effet à supposer que le débit produit au niveau de la glotte constitue un signal source filtré par la suite par le conduit vocal. Ce dernier est donc considéré comme une fonction de transfert dont les paramètres sont contrôlés par l'articulation de l'appareil phonatoire. Le signal d'entrée, dans sa version la plus élémentaire, traduit deux types de productions : les sons voisés (avec vibration des cordes vocales) et les sons non voisés (sans vibration, le son est produit par des turbulences de l'air au niveau de la glotte). Cela amène au schéma bloc suivant :

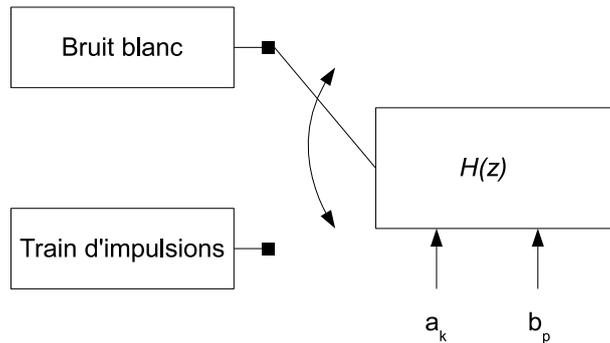


FIGURE 7 – Synthèse source-filtre.

Les paramètres de contrôle de cette synthèse sont les coefficients $a_k(t)$ et $b_p(t)$ du filtre, variant lentement avec le temps discret t . La fonction de transfert s'écrit alors

$$H(z) = \frac{\sum_{p=0}^P b_p z^{-p}}{1 - \sum_{k=1}^K a_k z^{-k}}.$$

Cette synthèse est aussi qualifiée de soustractive pour ce qu'elle met en forme le spectre de la source qui est plat à l'origine. Ainsi, les instruments à son entretenu, dont le fonctionnement peut-être rapproché de ce modèle de production vocale, sont bien adaptés pour ce type de méthode.

Les coefficients du filtre peuvent-être obtenus en analysant un signal de voix enregistré, par exemple en employant la prédiction linéaire. Celle-ci s'appuie sur un modèle plus restreint de filtre autorégressif, pour lequel, au numérateur de la fonction de transfert, seul le coefficient b_0 est non nul. Cette technique est optimale dans le sens des moindres carrés : l'échantillon modélisé à l'instant t est la meilleure approximation du signal original au sens de la norme quadratique, approximation obtenue comme combinaison linéaire des K échantillons précédents. D'autres technique d'estimation de l'enveloppe spectrale existe comme le cepstre ou l'interpolation directe des pics spectraux.

5 Synthèse par modèle physique et apparentés

Ce type de synthèse, basé sur l'étude physique du fonctionnement des instruments est devenue un axe de recherche et de développement majeur du domaine. Son intérêt principal réside dans sa capacité à reproduire la sensation d'interaction avec un instrument réel.

La synthèse dite de Karplus-Strong, du nom de ses initiateurs, fut sans doute un des précurseurs de ce type de synthèse en s'appuyant sur des lignes à retard pour traduire la propagation des ondes. Les méthodes actuelles, quant à elles, tendent à reproduire numériquement les équations d'évolution du système physique. Ces dernières ne seront pas développées ici ; leur description étant spécifique de l'instrument ou du type d'instrument (on peut citer par exemple les sons de flûte (Verge 95)).

5.1 Synthèse de Karplus-Strong

Cette méthode de synthèse a été étudiée par deux étudiants de Stanford, Karplus et Strong, en 1983. C'est une méthode bien adaptée pour les sons harmoniques, et en particulier les sons de cordes (guitare, piano etc...) Sa simplicité permet une implémentation en temps réel aisée. Avec quelques améliorations, on peut obtenir des synthèses de bonne qualité (Jaffe 83).

Le premier modèle est le suivant :

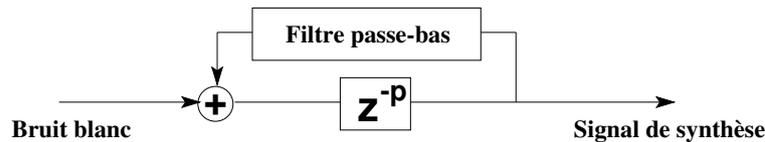


FIGURE 8 – Modèle simple de résonateur numérique

Il s'agit d'un filtre alimenté par une bouffée de bruit blanc. Le filtre dans sa version la plus simple est constitué d'une boucle récursive dans laquelle est introduit une cellule passe-bas. Une cellule passe-bas simple couramment utilisée est du type :

$$y_n = \frac{1}{2}(x_n + x_{n-1})$$

On peut montrer que dans ces conditions, la réponse impulsionnelle du filtre fait apparaître des sinusoïdes décroissantes en rapport harmonique, la (pseudo) période du signal étant de $pT + T/2$ secondes (où T représente le pas d'échantillonnage). On voit donc que le paramètre p peut être utilisé pour contrôler la fréquence du signal généré. De plus, la présence du filtre passe-bas dans la boucle provoque une atténuation plus rapide des hautes fréquences. On peut également évaluer l'amortissement de chaque partiel par :

$$\alpha_i = (\cos(\pi f_i T))^{1/(p+1/2)}$$

C'est le terme par lequel l'amplitude du partiel est multipliée à chaque échantillon. On voit que plus f_i est proche de $F_e/2$, plus le cosinus est petit, et plus le partiel est amorti. Ainsi

- La fréquence d'un partiel est fonction de la longueur du retard total (délai z^{-p} plus délai du passe-bas).
- L'amortissement d'un partiel de fréquence f_i est fonction du gain de la boucle de feedback à la fréquence considérée.

Le filtre passe-bas permet donc de reconstituer l'amortissement plus rapide des partiels de haute fréquence que l'on observe dans les cordes en régime libre.

En général, ce filtre est excité par une impulsion de bruit blanc de courte durée, ce qui donne un aspect plus naturel à l'attaque.

Un défaut de cette implémentation est la quantification des fondamentaux possibles. Les fréquences fondamentales réalisables s'écrivent :

$$f = \frac{F_e}{N + \frac{1}{2}} \quad \text{où } F_e \text{ représente la fréquence d'échantillonnage}$$

Puisque p est entier, l'ensemble des fréquences fondamentales est quantifié d'une façon malheureusement incompatible avec les applications musicales. Une façon de remédier à ce problème consiste à introduire dans la boucle de feedback une cellule passe-tout :

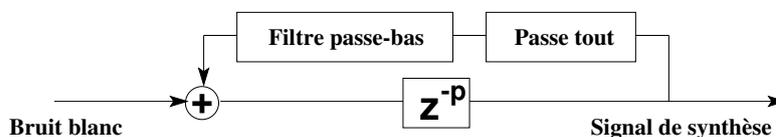


FIGURE 9 – Modèle amélioré pour réaliser des périodes non entières

On choisit une cellule du premier ordre pour sa simplicité :

$$H(z) = \frac{c + z^{-1}}{1 + cz^{-1}} \quad \text{avec} \quad |c| < 1$$

Le paramètre c contrôle le comportement en phase du passe-tout. Puisque ce dernier ne modifie pas le gain de la boucle de feedback, les amortissements des partiels sont inchangés. En revanche, le passe-tout introduit un déphasage qui correspond à un retard de phase. Ainsi la longueur de la boucle de feedback n'est plus de $pT + T/2$ mais de $pT + T/2 + \text{Arg}(H(e^{j\omega}))/\omega$. La fréquence fondamentale de la réponse impulsionnelle peut ainsi être ajustée avec une grande précision grâce au paramètre c . Ci-dessous est représenté le délai introduit par le passe-tout pour des valeurs de c comprises entre -0.9 et 0.9 .

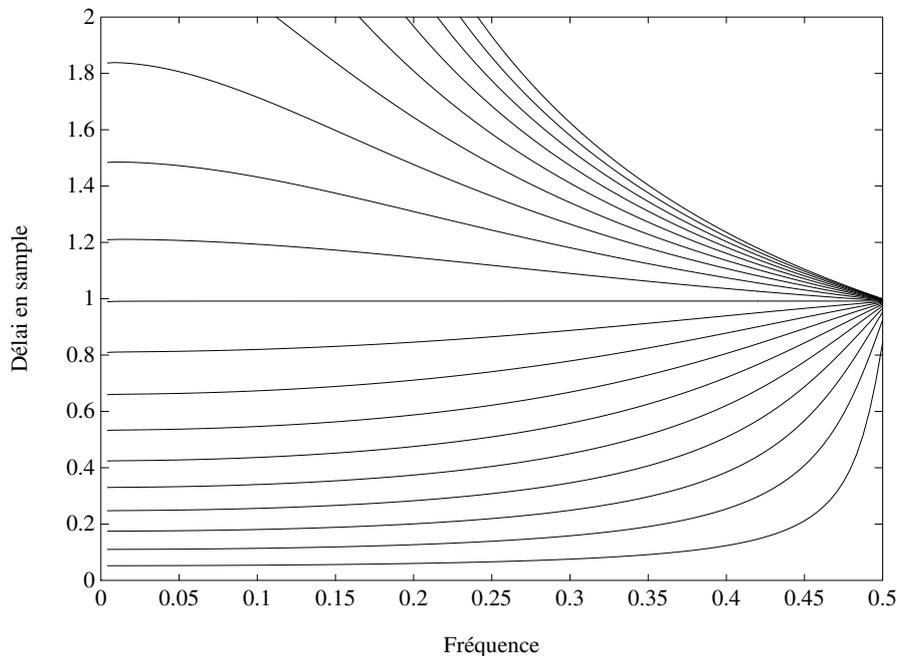


FIGURE 10 – Retard introduit par le passe-tout

Enfin, l'amortissement des hautes fréquences peut être mieux contrôlé par l'utilisation d'un filtre passe-

bas légèrement plus élaboré.

Ce type de synthèse se révèle très efficace pour la simulation de cordes de guitares, etc... Voir (Jaffe 83) pour plus de détail sur cette méthode de synthèse.

5.2 Solution numérique d'équations d'onde

Les techniques les plus avancées actuelles, sont basées sur des principes de discrétisation des équations d'onde. Elles s'appliquent de manière générale à tous les objets vibrants en utilisant des paradigmes tels que la plaque mince, la coque mince, les membranes, les cordes, les barres, etc... En étudiant les structures en un nombre fini de point de l'espace temps, on obtient des équations aux différences, c'est à dire des récurrences qui traduisent la propagation de l'onde dans le milieu élastique. Cette technique a donné lieu au développement de la synthèse à guide d'onde, dont un développement détaillé est disponible sur le web (Smith 09). Cette technique s'étend à des objets en 2 et 3 dimension, au prix bien sur, d'une complexité quadratique ou cubique respectivement.

Enfin, dans une autre approche, les équations sont résolues en découpant les objets en petits éléments et en propageant les solutions à l'aide de conditions de continuité aux frontières. Cette technique, appelée par éléments finis, est très répandue dans le champ industriel de la mécanique vibratoire, et a donnée lieu à des applications musicales comme par exemple celles développées par G. Dervaux (Dervaux 03).

Références

- D. Arfib. Digital synthesis of complex spectra by means of multiplication of nonlinear distorted sine waves. *Journal of the Audio Engineering Society*, vol. 14 (3), pp. 38–45, 1979.
- J. Chowning. The synthesis of complex audio spectra by means of frequency modulation. vol. 21 (7), pp. 526–534, 1973.
- G. Dervaux, A. Chaigne, E. Bécache, and P. Joly. Time-domain simulation of a guitar. model and method. *Journal of the Acoustical Society of America*, vol. 114 (6), pp. 3368–3383, 2003.
- D. A. Jaffe and J. O. Smith. Extensions of the karplus-strong plucked-string algorithm. *Computer Music J.*, vol. 7 (2), pp. 56–69, Summer 1983.
- M. Kahrs and K. Brandenburg. *Applications of Digital Signal Processing to Audio and Acoustics*. Kluwer Academic Press, Dordrecht, Netherland, 1998.
- J. Laroche. The use of the matrix pencil method for the spectrum analysis of musical signals. *J. Acoust. Soc. Am.*, vol. 94 (4), pp. 1958–1965, Oct 1993.
- M. Le Brun. Digital waveshaping synthesis. *Journal of the Audio Engineering Society*, vol. 27 (4), pp. 250–266, 1979.
- J. A. Moorer. The use of linear prediction of speech in computer music applications. *J. Audio Eng. Soc.*, vol. 27 (3), pp. 134–140, 1979.
- C. Roads. *L'audionumérique*. DUNOD, Paris, 1998.
- D. Rossum. Constraint based audio interpolators. *IEEE ASSP Workshop on app. of sig. proc. to audio and acous.*, 1993.
- J. O. Smith. Viewpoints on the history of digital synthesis. In *Proceedings of the International Computer Music Conference*, 1991.
- J. O. Smith. Site web de J. O. Smith sur la synthèse à guide d'ondes. <https://ccrma.stanford.edu/jos/wg.html>, 2009.

M.P. Verge. *Aeroacoustics of Confined Jets*. PhD thesis, TU Eindhoven, 1995.