CHAPITRE 7

Application à la musique

Le modèle acteur-observateur s'applique bien au contrôle de la synthèse sonore. Les acteurs modélisent le dispositif d'entrée pour l'utilisateur et le dispositif de synthèse sonore; ces deux entités sont mises en relation par un tissu d'observateurs qui gèrent une ou plusieurs stratégies de contrôle du son. Nous avons choisi pour illustrer ce principe, d'appliquer les méthodes d'animation dynamique au contrôle d'un modèle de synthèse de la voix chantée : la synthèse par forme d'onde formantique.

7.1 Synthèse sonore par forme d'onde formantique

La synthèse de la voix chantée a été étudiée depuis le milieu des années 70 par Xavier Rodet [Rode77], pour aboutir dix ans plus tard à un système général de synthèse adapté à la voix chantée [RodePotaBarr85] : Chant. A la fin des années 80, le système Chant a été porté sur la station de travail spécialisée de l'IRCAM pour proposer aux compositeurs un contrôle temps-réel de la synthèse sonore par forme d'onde formantique [EckeIovi94]. Au milieu des années 90, Chant a été porté sur micro-ordinateur Macintosh afin de permettre aux compositeurs de l'utiliser dans un cadre de travail per-

sonnel, mais sans pouvoir bénéficier d'un contrôle temps-réel, ou même interactif, de la synthèse sonore.

C'est pourquoi, en raison de l'intérêt de la synthèse par forme d'onde formantique (FOF) pour une illustration convaincante de l'application du modèle acteur-observateur au contrôle de la synthèse, nous avons réalisé une nouvelle implantation de Chant sur micro-ordinateur, contrôlable en temps-réel et intégrée au gestionnaire multimédia Quicktime.

7.1.1 La synthèse de la voix chantée

On considère que la voix humaine est le résultat d'une modification continuelle par le conduit vocal du signal émis par trois types de sources sonores :

- **une source voisée** : elle correspond à la vibration des cordes vocales et se présente sous la forme d'un signal quasi-périodique.
- **Une source fricative** : elle correspond aux turbulences engendrées par les rétrécissements en certains points du conduit buccal (lèvres, langue-palais, glotte).
- **Une source plosive** : elle correspond au bruit d'explosion engendré par la fermeture puis l'ouverture brusque du conduit buccal avec les lèvres ou la langue.

Dans le cas de voyelles chantées, les sources fricatives et plosives peuvent être omises.

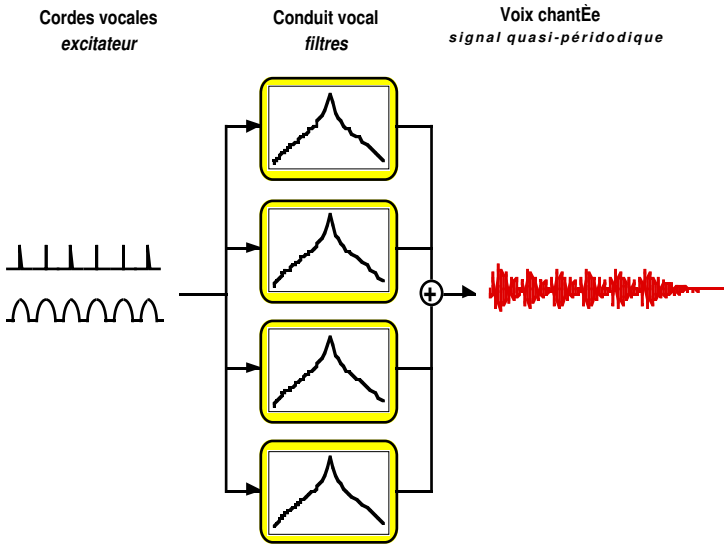
7.1.2 Modèle excitateur-résonateur

On modélise la production de voix chantée par un modèle de synthèse sonore du type excitateur-résonateur :

- **l'excitateur** correspond aux cordes vocales et définit le timbre de la voix; il est modélisé par une impulsion ou un arc.
- **le résonateur** correspond au conduit vocal et définit la voyelle chantée; il est modélisé par un jeu de filtres en parallèle. La fréquence centrale d'un de ces filtres est

appelée un formant. Une suite de formants permet de créer une voyelle. En effet, la perception sonore de l'homme « détecte » les pics d'énergie dans le spectre d'un son quasi-permanent pour le caractériser.

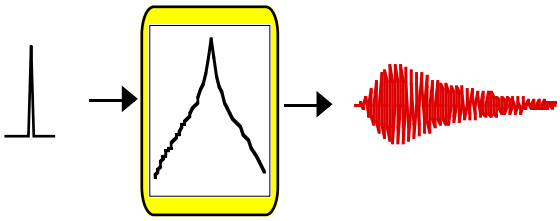
FIGURE 42 Modèle de synthèse de la voix chantée



Simplification du modèle

Le principe de la synthèse sonore par FOF est de considérer que les variations des paramètres des filtres sont suffisamment lentes pour être négligées lors du filtrage d'une excitation élémentaire. On peut calculer alors formellement l'expression correspondant à une impulsion élémentaire filtrée : c'est ce que l'on appelle une FOF ou forme d'onde formantique.

FIGURE 43 Une FOF est la réponse d'un filtre à une impulsion élémentaire



Ainsi, la réponse d'un filtre élémentaire à une impulsion est une sinusoïde amortie par une exponentielle (voir [RodePotaBarr85]).

$$y(t)=e^{-\alpha t}\sin\left(\omega t+\phi\right) \qquad \textbf{(EQ 19)}$$

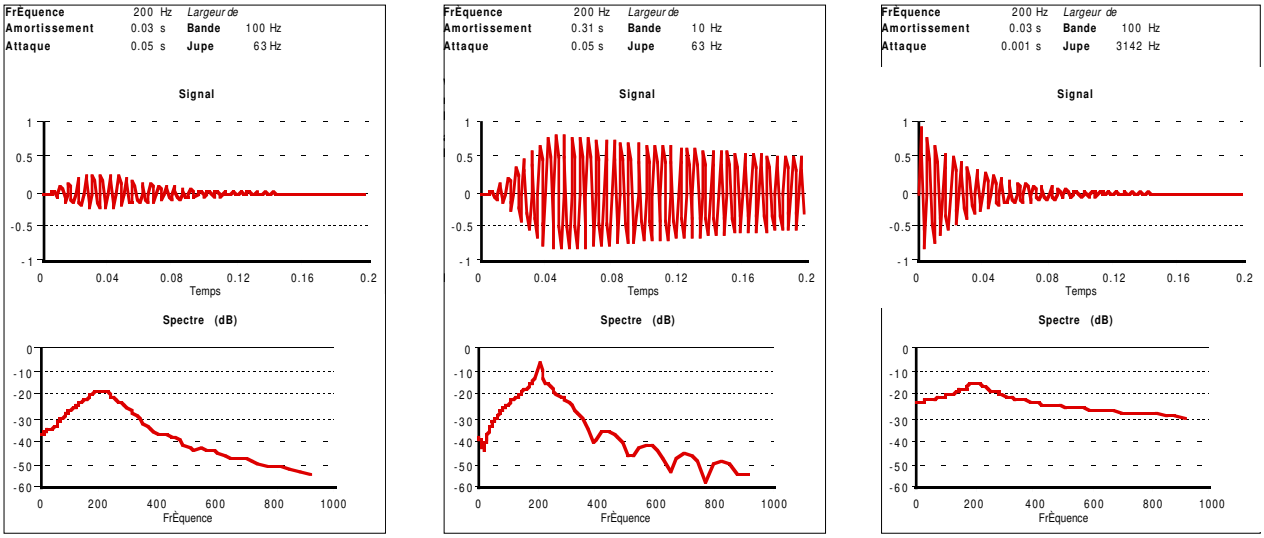
Afin de mieux modéliser les excitations engendrées par les cordes vocales, on va « lisser » l'attaque de la sinusoïde amortie par une portion de cosinus :

$$\begin{aligned} t<0 \;&\Rightarrow\; y(t)=0 \;(t<0)\\ 0<t<\frac{\pi}{\beta} \;&\Rightarrow\; y(t)=\frac{1}{2}\left(1-\cos\left(\beta t\right)\right)e^{-\alpha t}\sin\left(\omega t+\phi\right)\\ \frac{\pi}{\beta}<t \;&\Rightarrow\; y(t)=e^{-\alpha t}\sin\left(\omega t+\phi\right) \end{aligned} \qquad \textbf{(EQ 20)}$$

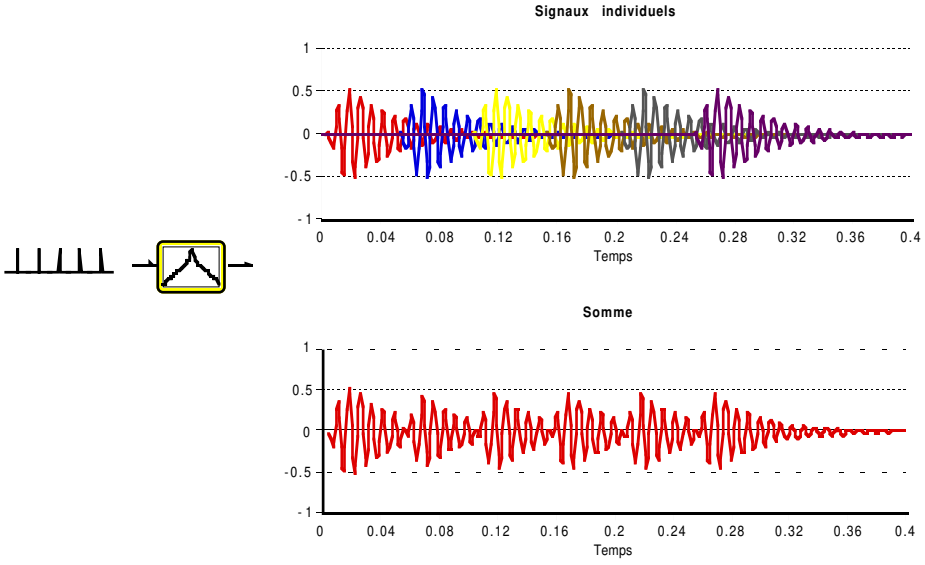
L'expression obtenue dépend de trois paramètres :

- ω : il correspond à la fréquence centrale du filtre;
- α : il correspond à la largeur de bande du filtre, c'est-à-dire à largeur du pic d'énergie à -3 dB du maximum;
- β : il correspond à la largeur de jupe du filtre c'est-à-dire à l'étalement du spectre dans les hautes fréquences.

FIGURE 44 Signal et spectre de trois FOF ayant la même fréquence d'oscillation mais différentes largeurs de bande ou de jupe



Enfin, pour calculer le signal résultant d'une excitation quasi-périodique, on va superposer un grand nombre de FOF décalées dans le temps de façon quasi-périodique.



Algorithme de calcul rapide

Afin de calculer rapidement une FOF, on utilise une méthode par itération : en résolvant par différences finies une équation différentielle linéaire du premier ou du deuxième ordre, les valeurs successives correspondant au sinus, au cosinus ou à l'exponentielle de la FOF peuvent être obtenues en un minimum d'additions et de multiplications. Cette méthode est beaucoup plus rapide qu'une tabulation dans la mesure où l'accès aux champs d'un tableau est pénalisé par la taille réduite de la mémoire cache des machines à processeur RISC. De plus, elle ne dépend pas d'une évaluation de cosinus ou d'exponentielle par l'unité arithmétique du processeur.

7.1.3 Un composant Quicktime de synthèse en temps réel

L'implantation de l'algorithme repose sur l'écriture d'un composant (au sens du *Component Manager* de MacOS) s'interfaçant avec le gestionnaire multimédia Quicktime. Cette implantation présente plusieurs avantages :

- **Elle est multiplateforme** : elle peut être portée sur tout système compatible avec Quicktime, comme Windows ou IRIX[1].
- **Elle permet un contrôle générique** : dans la mesure où les logiciels multimédias, MIDI ou compatibles OMS ont accès à Quicktime de façon transparente.
- **Elle est modulaire et réutilisable** : le code de synthèse est concentré sur une dizaine de lignes en C. En changeant ce code, on peut créer très rapidement de nouveaux modules de synthèse sonore.

Quicktime Music Architecture (QMA)

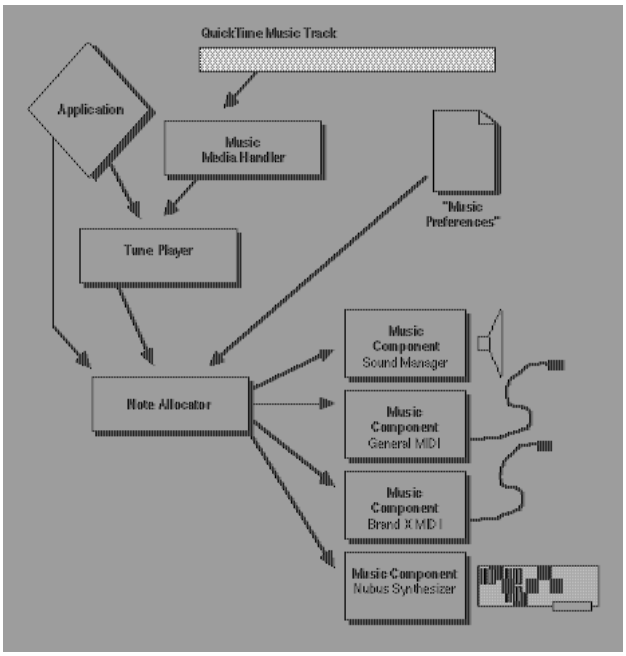
L'architecture musicale de Quicktime définit quatre nouveaux types de composants :

- **Music Data Handler** : ce composant gère un format de données musicales, comparable au format MIDI mais incluant un contrôle micro-tonal de la hauteur, un accès générique aux paramètres de synthèse et une description du timbre par nom et par mode de synthèse. Il permet de lire des données musicales dans une séquence multimédia.
- **Tune Player** : ce composant se charge de la mise en temps d'événement musicaux. Une application souhaitant jouer des notes de musique à des dates précises, pourra faire appel à une instance de ce composant.
- **Note Allocator** : ce composant gère la répartition de la polyphonie et des timbres à reproduire en fonction de l'environnement de synthèse disponible. Une application souhaitant jouer des notes de musique immédiatement, pourra faire appel à une instance de ce composant.
- **Music Synthesizer** : ce composant synthétise des notes en fonction de paramètres de hauteur, de sonie et de timbre. En créant de nouveaux composants synthétiseurs, on

1. IRIX est la version du système UNIX commercialisée par Silicon Graphics.

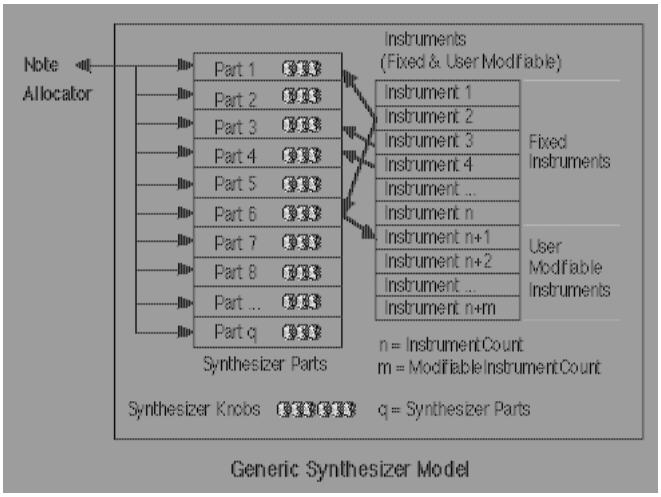
étend le nombre d'algorithmes de synthèse disponibles aux applications.

FIGURE 45 L'architecture musicale de Quicktime



L'implantation du synthétiseur de FOF passe donc par la conception d'un composant de synthèse. Dans QMA, ces composants comportent un certain nombre de voix (polyphoniques ou monophoniques) correspondant chacune à un timbre donné. L'ensemble de ces voix se partagent la polyphonie totale du composant. Elles peuvent chacune être contrôlées par un jeu de paramètres définis par le concepteur du composant.

FIGURE 46 Chaque composant de synthèse peut produire simultanément plusieurs timbres provenant d'une même banque de timbres



Paramètres de synthèse

Nous avons choisi de définir tout d'abord un jeu de 48 paramètres de contrôle de la synthèse. Ces 48 paramètres correspondent à un jeu de 6 formants par voix, chacun de ces formants étant défini par les 8 paramètres :

- **Fund** : définit la fréquence fondamentale de l'excitation associée au formant.
- **Freq** : définit la période du formant.
- **Bw** : définit la largeur de bande des FOF associées au formant.
- **Amp** : définit l'amplitude du formant.
- **Attack** : définit la largeur de jupe du formant.
- **Debatt** : limite le temps de relaxation d'une FOF.
- **Atten** : définit le temps d'atténuation d'une FOF.
- **Cross-over** : définit le recouvrement de FOF maximum du formant.

On définit aussi au niveau de chaque voix un jeu paramètres de contrôle :

- **Relax** : temps de relaxation du filtre du premier ordre appliqué aux données de contrôle pour éviter un effet « crémaillère ».
- **Amp** : amplitude de la voix.
- **Pitch** : hauteur de la voix.

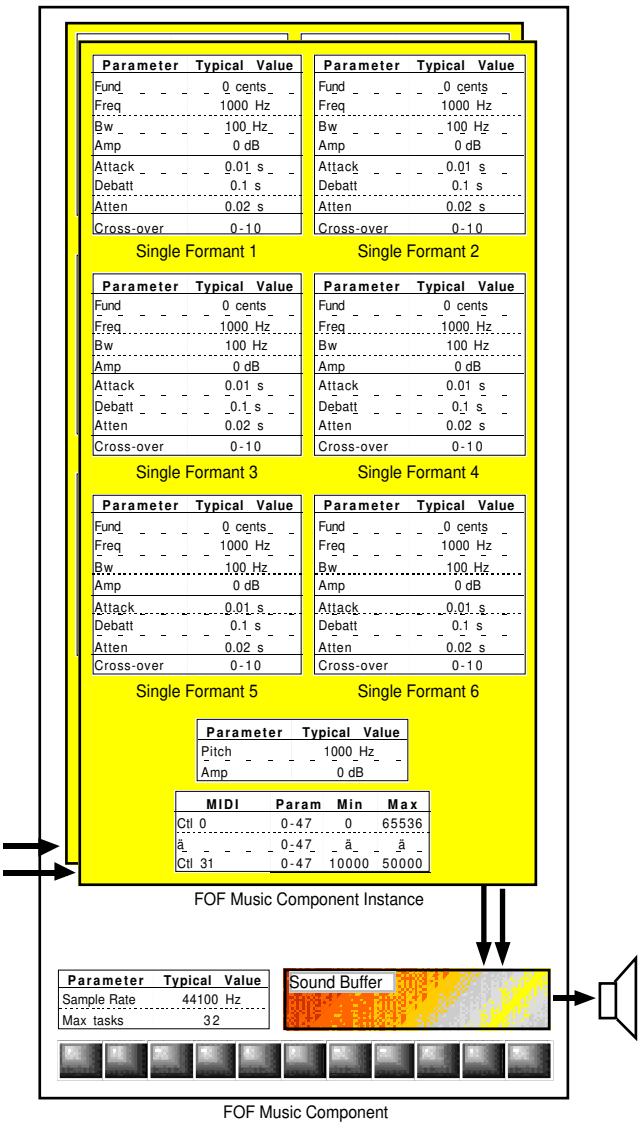
Enfin, au niveau du composant, on définit :

- **SampRate** : fréquence d'échantillonnage du signal
- **MaxTask** : polyphonie ou recouvrement maximal du composant.

Contrôle MIDI

128 autres paramètres permettent de piloter par contrôle MIDI les paramètres de synthèse en définissant pour chacun des 32 contrôles continus MIDI, un paramètre de synthèse cible, une valeur minimum et une valeur maximum.

FIGURE 47 Architecture du composant de synthèse de forme d'onde formantique pour Quicktime Music Architecture



Principe de fonctionnement

Lorsque le composant reçoit un message lui ordonnant de créer une note, il alloue pour chaque formant un certain nombre de tâches en fonction du recouvrement défini par les paramètres de synthèse. Dans le même temps, un système d'interruption est mis en place pour remplir régulièrement un

buffer où est stocké le signal sonore. A chaque interruption, on parcourt la liste des tâches en calculant pour chaque tâche une portion de FOF correspondant à la taille du buffer. Lorsqu'une FOF a été complètement calculée, la tâche est réinitialisée pour être exécutée ultérieurement en fonction de la fréquence fondamentale d'excitation.

7.2 Contrôle activisuel de la synthèse sonore

Le contrôle de la synthèse sonore a pour objectif de faciliter la manipulation de modèles de synthèse dépendant d'un grand nombre de paramètres en offrant au musicien des outils capables de gérer, de représenter et de manipuler simplement le modèle. De nombreux outils de contrôle de la synthèse sonore fondés sur des interfaces graphiques ont déjà été développés [CaroOrla91, Opco90]. A l'instar de générateurs d'interface utilisateur, ces outils permettent de placer sur l'écran des composants de contrôle et de spécifier leur relation avec le modèle de synthèse.

C'est ce principe que nous allons appliquer à la synthèse par forme d'onde formantique en proposant deux innovations :

- **Une interface en trois dimensions**. Les modèles graphiques que nous allons utiliser sont ceux introduits au chapitre précédent pour l'animation.
- **Un contrôle actif**. Le dispositif de contrôle est actif dans la mesure où il peut évoluer en dehors du contrôle de l'animateur.

La mise en interaction avec un modèle graphique et la mise en oeuvre d'un formalisme de contrôle sont obtenus grâce au modèle acteur-observateur.

7.2.1 Observateurs de voyelles

Une voyelle est obtenue comme un ensemble de quelques formants dont les paramètres de fréquence, de largeur de bande et d'amplitude vérifient une contrainte donnée. Pour

construire une phrase, on intercale entre deux voyelles une transition. Les transitions sont calculées au moyen de règles d'interpolation obtenues par expérimentation sur la voix humaine.

Création de voyelles

On représente les voyelles comme des observations entre FOF. Toute modification d'un paramètre d'une des FOF de la voyelle conduira à une résolution de contrainte permettant d'assurer la continuité de la voyelle dans une transition.

Métastructuration des voyelles

On définit les transitions de voyelles par méta-observation. En se plaçant dans le formalisme d'animation par points clés, une transition correpond à un interpolateur (voir paragraphe 6.2.3 page 114).

7.2.2 Contrôle mécanique des voyelles

Au lieu d'animer les voyelles par des observateurs de points clés, on va désormais les faire interagir avec un modèle d'animation dynamique.

Utilisation d'un modèle d'animation

On se donne un système d'animation physique composé d'objets élémentaires reliés par des contraintes mécaniques. On définit un observateur pour faire interagir les objets de la scène et les formants. C'est l'observateur dit de « mécavoyelle ».

Utilisation d'une métastructure voyelle-contrainte

Pour améliorer le résultat, on définit un observateur faisant interagir voyelle et contrainte mécanique. Ainsi, lorsque les contraintes de voyelle deviennent difficiles à satisfaire, elles se traduisent en augmentation des contraintes mécaniques entre objets élémentaires.

7.2.3 Introduction d'hystérésis

Le modèle de contrôle mécanique présenté précédemment permet de mieux appréhender la manipulation de voyelles grâce à une représentation graphique des contraintes de la synthèse sonore. Cependant, le caractère bijectif de l'interaction reste assez pauvre. C'est pourquoi, on va chercher à introduire de l'hystérésis. On anime l'observateur de mécavoyelle par un système d'états et l'on fait interagir l'observateur d'état avec celui de contrainte mécanique pour assurer les changements d'état. Il s'agit encore d'une application de la métastructuration.