Les nouveaux gestes de la musique, dir. H. Genevois et R de Vivo, Marseille : Parenthèses, 1999

Contrôle interactif d'une voix chantée de synthèse

Laurent Pottier

Résumé *: Nous décrivons un synthétiseur utilisé pour produire des sons de voix chantée dans le cadre d'une installation virtuelle interactive conçue par Catherine Ikam et Jean-Louis Flery. Ce synthétiseur a été mis au point sur la Station d'Informatique Musicale de l'Ircam. L'installation mettait en scène un visage et une voix de synthèse, créés et modifiés en temps réel par les déplacements d'un émetteur à infrarouge.*

1 Présentation du projet

1.1 Principe du dispositif

Le dispositif de synthèse de voix chantée que nous allons décrire est un système « temps réel » à contrôle interactif. Il a été utilisé au cours de deux installations virtuelles réalisées sous la direction de Catherine Ikam et Jean-Louis Flery. La première installation, « Le Messager », a fonctionné en continu de mars à août 1995 lors de l'exposition spectacle Cité-Ciné 2 à la Défense. La seconde, « Alex », a été présentée à l'Ircam pendant le mois de juin 1996 à l'occasion de l'inauguration de la médiathèque.

Les installations étaient présentées dans une pièce sombre dans laquelle un visage virtuel était projeté sur un grand écran. L'image était accompagnée d'une voix chantée de synthèse diffusée par plusieurs haut-parleurs.

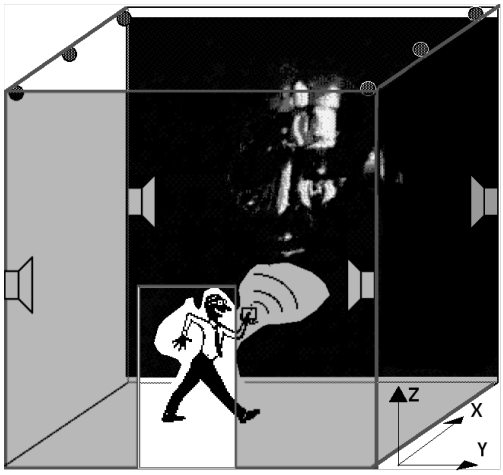

figure 1 : schéma du dispositif

Le visiteur pouvait modifier les expressions du visage et les transformations du son grâce à un émetteur à infrarouge qu'il tenait à la main lors de ses déplacements dans la pièce. Les positions de l'émetteur étaient transmises en permanence à des ordinateurs par l'intermédiaire de capteurs accrochés au plafond.

1.2 L'équipe ayant réalisé le projet

Les images de synthèse et leurs logiciels d'animation ont été produits par les établissements MacGuff Line.
La partie sonore a été dirigée Jean-Baptiste Barrière, compositeur et directeur de la pédagogie de l'Ircam. Nous avons programmé le synthétiseur sur la Station d'Informatique Musicale (SIM) de l'Ircam. Le programme d'interaction a été construit avec l'aide de Tom Mays en utilisant le programme Max sur ordinateur

Macintosh. L'installation et le mixage ont été réalisés avec la participation de Laurent Ghys, Daniel Raguin et David Meysonnier.

1.3 Le matériel utilisé

Le dispositif a requis l'utilisation de quatre ordinateurs.
En début de chaîne, un ordinateur IBM P.C. servait à analyser les informations en provenance des capteurs afin de délivrer un flot continu de triplets de position (x, y, z).
Un ordinateur SGI « Reality Engine » a été employé pour produire l'image de synthèse. Il transmettait à un ordinateur Macintosh les informations de position provenant de l'émetteur et y ajoutait des informations relatives à l'image.
Sur l'ordinateur Macintosh, le programme Max effectuait le traitement des informations fournies par la SGI afin de déclencher des sons sur un échantillonneur Akai « S1000 » et d'envoyer un flot continu de paramètres vers la Station d'Informatique Musicale installée sur un ordinateur NeXT. Ces paramètres étaient destinés à un contrôle en temps réel du synthétiseur Chant implémenté sur la SIM.

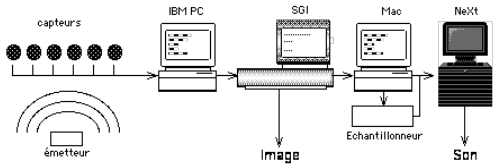

figure 2 : dispositif informatique utilisé

2 La Station Musicale de l'Ircam

2.1 Matériel

La Station d'Informatique Musicale est un système de synthèse et de traitement du son sur ordinateur fonctionnant en temps réel. Il est formé d'un ordinateur NeXT comportant une à trois cartes DSP munies de deux processeurs Intel i860 et de quatre entrées et sorties indépendantes.

2.2 Logiciel

La SIM est contrôlée par les programmes FTS et Max. Le programme FTS (Faster Than Sound) gère les calculs pour la synthèse et le traitement du son. L'environnement Max permet une programmation graphique du contrôle des dispositifs de synthèse. [Puckette-1991]

2.3 Librairies

Max et FTS sont des programmes « ouverts » qui permettent la programmation de librairies spécialisées dans différents types d'opérations.
Pour ce projet, nous avons utilisé la librairie Chant destinée à la synthèse de la voix chantée et la librairie SIMLIB contenant de nombreuses fonctions de contrôle de la synthèse.

3 Le synthétiseur Chant

Le synthétiseur Chant, conçu à l'Ircam en 1978/79, est un outil de synthèse par règles construit à partir d'études effectuées sur la voix chantée. Son principe repose sur la modélisation des processus de transformation des sons émis par la glotte par le conduit vocal. [Rodet, Bennett-1980]
Chant a été initialement écrit en SAIL sur Dec-PDP10 par Xavier Rodet et Yves Potard en 1979. Réécrit en C sous Unix en 1989, il est devenu polytimbral et polyphonique et a bénéficié du contrôle par le programme Formes. En 1992, il a été porté sur Macintosh pour être piloté par le programme PatchWork. [Iovino, Laurson, Pottier-1994]
Une version a été implémentée sur la SIM en 1994 par Francisco Iovino et Gerhardt Eckel.

Pour être modélisé, l'appareil vocal est dissocié en deux parties : d'un côté les cordes vocales qui produisent un son harmonique au spectre riche, de l'autre le conduit vocal qui filtre le son à la manière d'un résonateur.

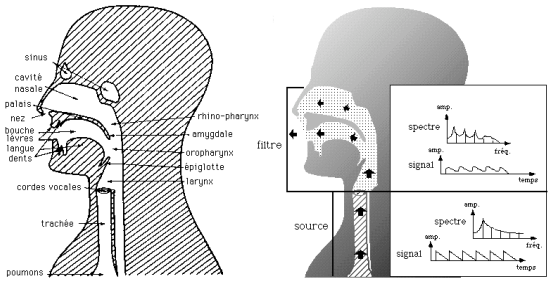

figure 3 : filtrage du son émis par la glotte

La simulation de l'appareil vocal s'effectue par un modèle de type excitation-résonance, en employant pour chaque résonance une Fonction d'Onde Formantique ou Fof.

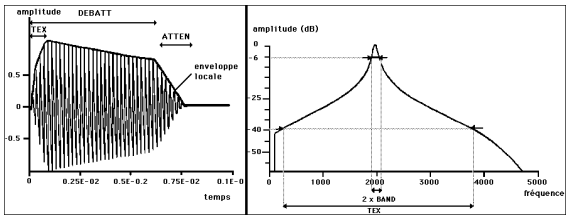

figure 4 : représentations temporelle (à gauche) et spectrale (à droite) d'une Fof

En additionnant plusieurs Fofs on peut reproduire les spectres caractéristiques des différentes voyelles de la voix chantée.

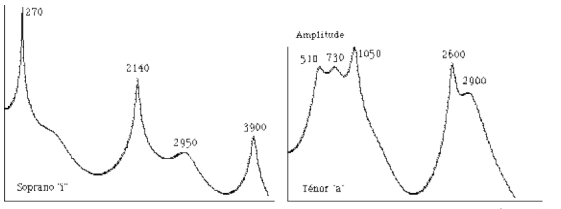

figure 5 : spectres de deux voyelles chantées : à gauche « i » par une voix de soprano, à droite, « a » par une voix de basse

Le synthétiseur Chant dispose d'un ensemble de règles permettant de contrôler de façon automatique les variations des formants en fonction des variations d'intensité et de hauteur de la voix. [Sunberg-1985]

4 Les librairies d'Adrien Lefevre

Nous avons utilisé environ quarante paramètres pour contrôler le synthétiseur Chant. Ces paramètres étaient générés par le programme Max sur Macintosh en fonction des données de position en provenance de l'émetteur. Deux librairies de Max, VTboule et MIDIcrypt, ont permis de réaliser ce contrôle de façon efficace.

4.1 La librairie VTboule

La librairie VTboule sert à la production de données numériques par l'interpolation de listes de valeurs stockées dans des boules, objets graphiques disposés sur une fenêtre à l'écran.
La position d'un pointeur entre les boules permet de calculer des interpolations entre les valeurs contenues dans les différentes boules. L'interpolation est donnée par la moyenne des valeurs pondérées par la taille des boules et par l'inverse de la distance entre le pointeur et chaque boule.

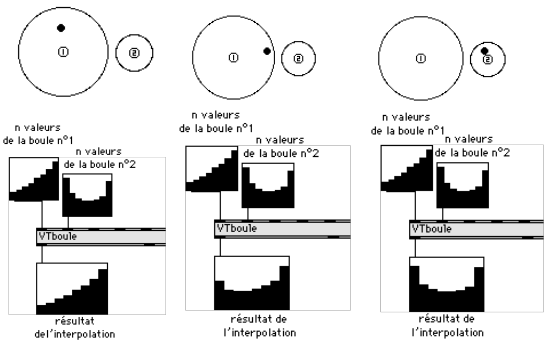

figure 6 : relation entre le déplacement du pointeur et les valeurs produites par l'objet VTboule

Nous avons utilisé simultanément plusieurs groupes de boules pour le contrôle global d'une vingtaine de

paramètres de la synthèse. Le pointeur se déplaçant entre les boules représentait la position de l'émetteur à infrarouge dans la salle de projection.

4.2 La librairie MIDIcrypt

Pour permettre la transmission d'une grande quantité de données en MIDI entre les ordinateurs Macintosh et NeXT, les données étaient envoyées sous forme de messages MIDI de type « Système Exclusif ». Pour cela, nous avons utilisé les objets *midicrypt* mis au point par Adrien Lefevre.
L'objet *midicrypt* permet de coder une information dans un message de type « Système Exclusif » sur une machine pour l'envoyer en MIDI vers une autre machine où l'objet *mididecrypt* sert à la décoder. Il est possible de choisir la résolution de chaque information (8 bits, 12 bits…) et d'envoyer plusieurs informations dans un même message.

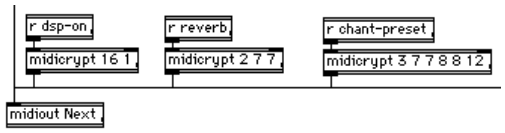

figure 7 : exemple de transmission des informations à l'aide d'un module midicrypt.

5 Description du synthétiseur

Le dispositif reposait sur la synthèse de la voix réalisée à l'aide du synthétiseur Chant. Différentes fonctions de traitement du son étaient également utilisées : un harmoniseur, des filtres et des fonctions de spatialisation.

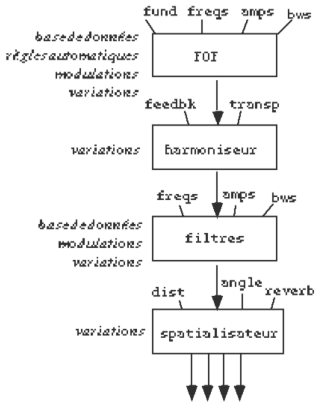

figure 8 : schéma de production du son sur le dispositif du Messager et principaux paramètres de contrôle

5.1 Les paramètres de la synthèse de la voix

Chant comprend à la fois des fonctions qui génèrent le signal et des règles qui en contrôlent les paramètres.
Les fonctions qui produisent le son sont les Fofs. Les trains de Fofs génèrent un signal périodique. Ils présentent chacun un formant. Pour la synthèse de la voix, nous avons utilisé cinq trains de Fofs dont les différents formants permettaient la reproduction de plusieurs phonèmes.

5.2 Les formants

Huit phonèmes ont servi de modèles. En fonction des mouvements de l'émetteur, les fréquences des formants étaient en permanence interpolées, de façon à ce que la voix effectue des transitions entre les voyelles.

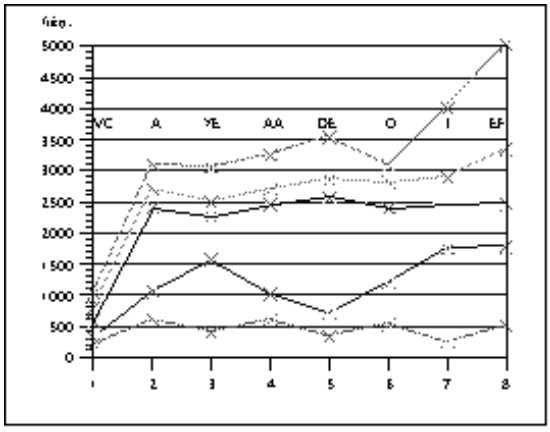

figure 9 : fréquence des cinq formants utilisés pour les huit phonèmes de référence

Les formants étaient contrôlés par un ensemble de règles importantes pour la production de sons réalistes. Certaines de ces règles sont entièrement automatiques alors que d'autres comportent des accès servant à modifier leurs paramètres en fonction du temps.

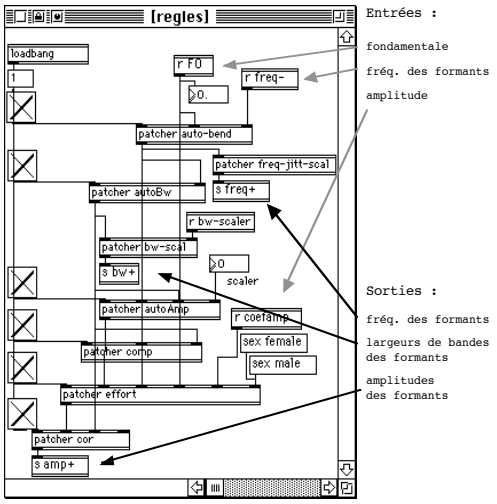

figure 10 : patch de connexion des différentes règles de contrôle des paramètres formantiques

La figure 10 présente les règles ayant été utilisées. Ces règles contrôlent automatiquement les fréquences (*freq+*), les amplitudes (*amp+*) et les largeurs de bande (*bw+*) des formants en fonction des paramètres des notes jouées. Les paramètres des notes sont leur hauteur (*F0*), leur intensité (*coefamp*) et le type de voix utilisé.

Tant que l'utilisateur du synthétiseur Chant garde pour ces paramètres des valeurs compatibles avec celles de la voix chantée, le comportement de tous les paramètres formantiques reste cohérent et produit un son de type vocal crédible.

En plus de ces règles, des fonctions de contrôle agissant sur les paramètres des formants ou sur la fréquence fondamentale des Fofs ont été programmées. Elles ont permis de s'éloigner du modèle vocal pour produire de temps en temps des sons plus artificiels. Des modulations aléatoires ou « jitters » ont été employées pour créer des variations des fréquences des formants.

D'autres paramètres permettaient le contrôle global des amplitudes et des fréquences des deux formants graves ou des trois formants aigus. Une large palette sonore était ainsi disponible.

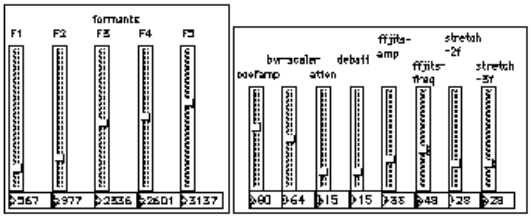

figure 11 : paramètres de contrôle des formants

5.3 La fréquence fondamentale

La voix chantée ne suivait pas de mélodie mais délivrait un son continu dont la hauteur était modifiée par plusieurs fonctions : deux vibratos dont l'un servait aux modulations à basse fréquence ; un triple « jitter » (*f0-jitt*) ; un « portamento » (*bend*) ; un mélisme (*chœur1tab*).

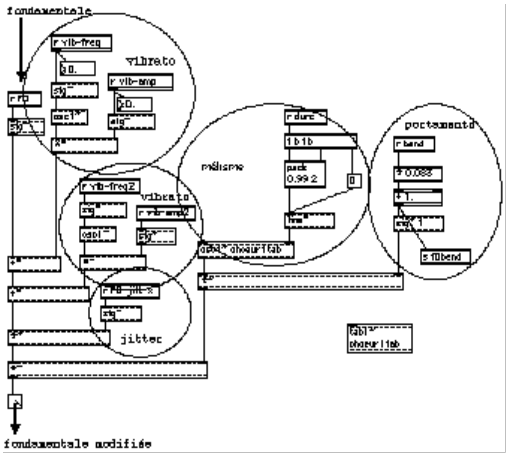

figure 12 : dispositif de modification de la fréquence fondamentale comportant deux vibratos, un « jitter », un mélisme et un « portamento »

Les paramètres permettant le contrôle de la fondamentale sont indiqués dans la figure 13.

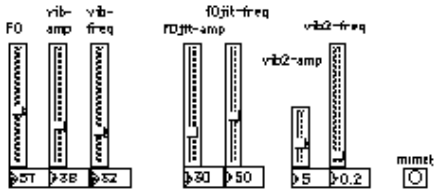

figure 13 : paramètres de contrôle de la fondamentale

5.4 Les traitements

Différents traitements étaient appliqués à la voix de synthèse.

Un harmoniseur était employé pour des transpositions de l'ordre du quart de ton afin de créer un léger effet de chœur sur la voix.

Des filtres ont été utilisés pour faire résonner la voix sur les fréquences de ses formants. Lorsqu'ils étaient en action, le son de la voix devenait très cristallin.

Un dispositif de spatialisation du son a permis de créer des déplacements de la voix dans l'espace, en jouant sur le taux de réverbération, l'équalisation et le volume du son dans les différents haut-parleurs.

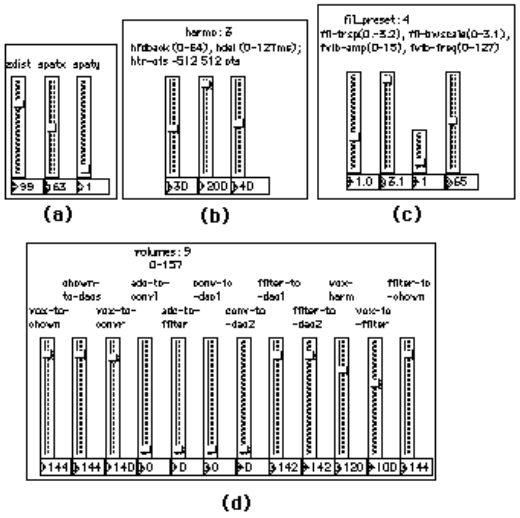

figure 14 : paramètres pour les fonctions de traitement du son : (a) spatialisateur, (b) harmoniseur, (c) filtres, (d) mixage

6 Les patchs pour le contrôle

La principale fonction utilisée dans Max sur le Macintosh pour le contrôle de la synthèse a été l'objet *VTboule*. Il a permis de traduire directement les informations de position horizontale de l'émetteur en vecteur d'interpolation des boules. Trois ensembles de boules ont servi au contrôle des différents groupes de paramètres.

6.1 Contrôle des fréquences des formants

Le contrôle des fréquences des formants était réalisé par un premier groupe de boules. Un phonème était attribué à chaque boule sous la forme d'une liste de cinq fréquences. Le déplacement de l'émetteur dans la pièce, représenté par le point noir sur la figure 15, permettait de parcourir les différents phonèmes par des transitions progressives.

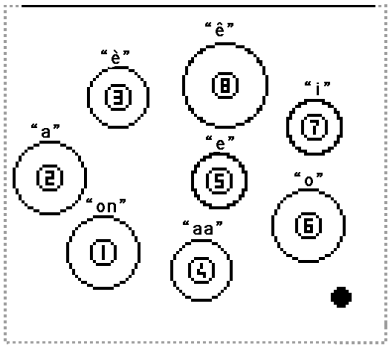

figure 15 : disposition des boules contrôlant les fréquences des formants

6.2 Contrôle des paramètres du timbre

Un ensemble de boules contrôlait trois groupes de paramètres agissant sur le timbre : un vibrato sur la fondamentale, des microvariations aléatoires (« jitter ») sur les formants et des fonctions de transpositions (« stretch ») des fréquences des formants.

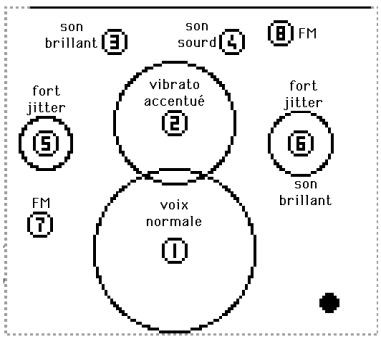

figure 16 : disposition des boules contrôlant le timbre

De légères variations des paramètres du vibrato permettent de changer notablement le caractère de la voix. Par contre, les valeurs extrêmes de la fréquence ou de l'amplitude de cette modulation, peuvent produire des transformations du timbre similaires à celles obtenues en modulation de fréquence. Pour permettre un jeu sur les deux registres, les variations des paramètres du vibrato suivaient une courbe exponentielle.

Le « jitter » appliqué sur les fréquences des formants anime le son. Lorsque ces variations prennent de l'ampleur, le timbre devient perturbé et chaotique. Comme pour le vibrato, les paramètres du « jitter » suivaient des courbes exponentielles.

Le paramètre « stretch » servait à amplifier ou à diminuer l'ensemble des fréquences des formants. Ce paramètre agissait à la façon d'un équaliseur en déplaçant les formants dans le spectre. Il permettait de rendre les sons très sourds ou très brillants.

6.3 Contrôle de la spatialisation

Le contrôle de la spatialisation a été réalisé par un troisième ensemble de boules. La boule centrale de la figure 17 assurait une diffusion équilibrée du son dans tous les haut-parleurs. Les boules périphériques, numérotées de 1 à 4, plaçaient le son uniquement dans le haut-parleur le plus proche. A l'inverse, les boules 6 et 7 rejetaient le son vers le côté opposé de la pièce. La boule 8 faisait office de « trou noir » faisant disparaître le son lorsqu'on s'en approchait.

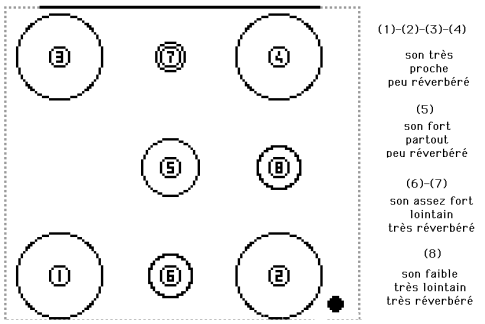

figure 17 : disposition des boules contrôlant la spatialisation

6.4 Autres contrôles

Certains des paramètres de contrôle de la synthèse ne dépendaient pas des boules d'interpolation. Ils étaient réglés à l'initialisation du patch ou dépendaient de facteurs dynamiques.
Par exemple, la vitesse du déplacement de l'émetteur dans la pièce modifiait le volume global du son tandis que ses déplacements verticaux modifiait la hauteur du son. Par ailleurs, l'intensité des filtres augmentait lorsque l'émetteur s'écartait du centre de la pièce.

7Conclusion

La technique d'interpolation par boules et quelques algorithmes supplémentaires nous ont permis de contrôler de nombreux paramètres de la synthèse par le déplacement de l'émetteur. La superposition des zones d'influence des ensembles de boules a créé une sorte de labyrinthe à l'intérieur duquel le spectateur se déplaçait en essayant d'en percevoir les méandres.
L'intérêt de ce dispositif d'interaction réside à la fois dans la grande sobriété du produit présenté - un visage en mouvement, un son de voix chantée - et dans sa grande complexité - une quarantaine de paramètres de contrôle pour l'image et pour le son, variant en continu lors des déplacements du capteur dans la salle.
Ce dispositif combinait une grande lisibilité à une grande diversité. Il permettait au visiteur de comprendre très vite le principe de l'interaction et de découvrir toutes ses subtilités en jouant avec cet instrument insolite produisant à la fois image et son.

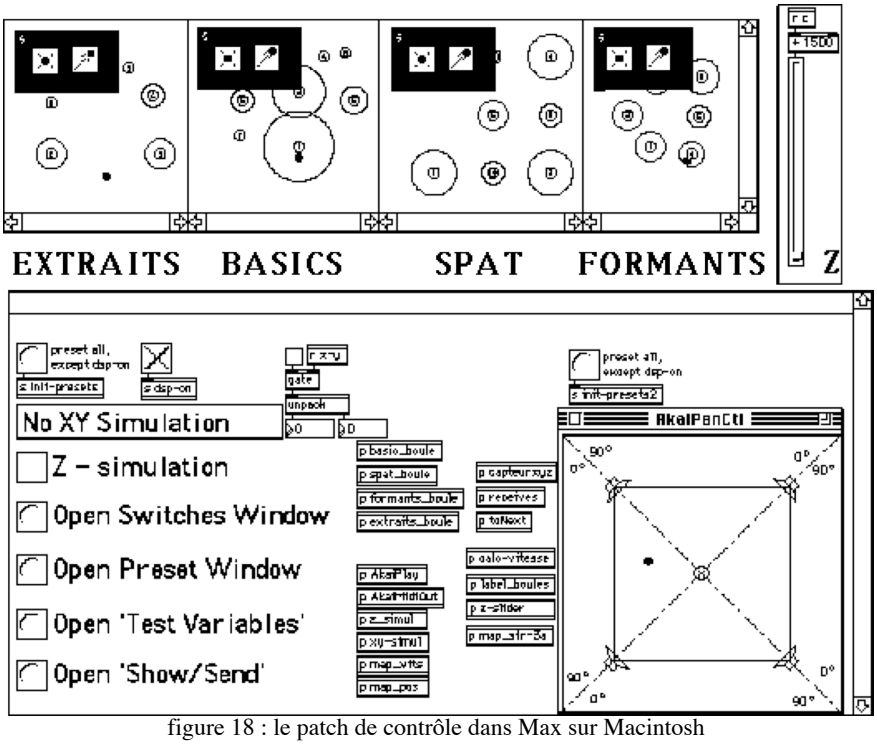

figure 18 : le patch de contrôle dans Max sur Macintosh

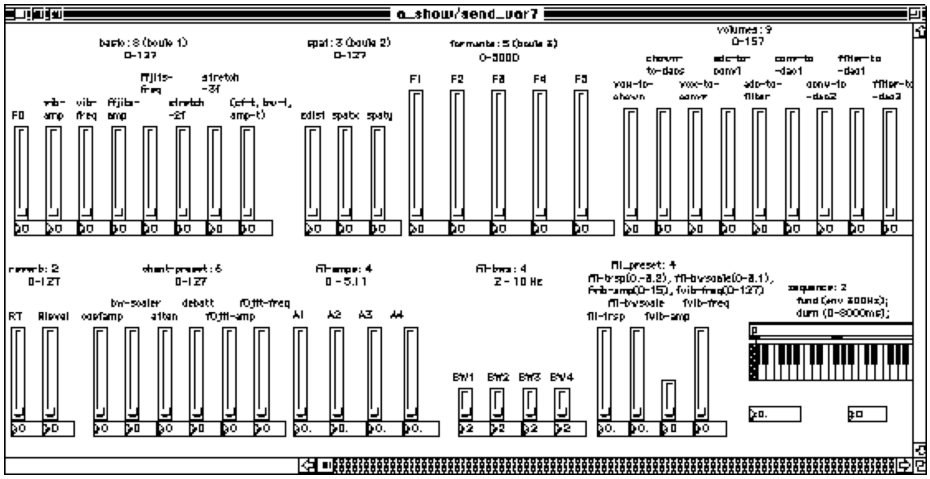

figure 19 : l'ensemble des paramètres de contrôle de la synthèse adressable du Macintosh vers la NeXT

Références

Iovino, Laurson, Pottier-1994 : Iovino (Francisco), Laurson (Mikael), Pottier (Laurent), « PW - Chant - Reference », *Document Ircam*, Paris, Juil. 1994.

Puckette-1991 : Puckette (Miller), « Combining event and signal processing in the MAX graphical programming environment », *Computer Music Journal*, vol. 15, n°3, 1991, p. 68-77.

Rodet, Bennet-1980 : Rodet (Xavier), Bennett (Gerald), « Synthèse de la voix chantée par ordinateur », *Conférences des Journées d'Etude*, Festival International du Son, Paris, 1980, p. 73-91.

Sunberg-1987 : Sundberg (Johan), *The Science of the Singing Voice*, Northern Illinois Univ. Press, Paris, 1987.