



HAL
open science

Concepts d'interfaces graphiques pour la création musicale et sonore

Hugues Vinet

► **To cite this version:**

Hugues Vinet. Concepts d'interfaces graphiques pour la création musicale et sonore. Hermes Science;
Hugues Vinet. Interfaces homme-machine et création musicale, , pp.97-121, 1999. hal-01250801

HAL Id: hal-01250801

<https://hal.science/hal-01250801v1>

Submitted on 5 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chapitre 5

Concepts d'interfaces graphiques pour la production musicale et sonore

HUGUES VINET

5.1. Introduction

L'essor de la micro-informatique, depuis le début des années quatre-vingt, a contribué à la généralisation de l'outil informatique dans les domaines de la création musicale et de la production sonore. D'un usage jusqu'alors réservé aux centres de recherche, notamment en rapport avec la création contemporaine, l'ordinateur a progressivement investi studios amateurs et professionnels, proposant de nouvelles modalités opératoires qui ont tendu à se substituer aux fonctions de production existantes : commande et programmation de synthétiseurs à travers le protocole MIDI, échantillonnage audionumérique, édition et montage non destructifs de sons numérisés sur disque dur, traitement en temps réel, formalisation et manipulation des structures musicales, synthèse sonore, simulation acoustique, etc.

Ces nouvelles applications se sont développées conjointement avec la démocratisation des interfaces graphiques, dont elles ont tiré parti pour proposer de nouvelles formes de représentation et de manipulation des contenus musicaux et sonores. Les concepts sur lesquels reposent les interfaces des différents logiciels existants sont suffisamment stables et convergents pour que leur analyse puisse être envisagée globalement, sur la base de critères communs. La présente étude vise à établir une telle synthèse, en définissant une typologie des interfaces graphiques rencontrées dans les différentes applications de production musicale et sonore, qu'il

s'agisse de produits commerciaux ou de développements plus expérimentaux issus de recherches en informatique musicale.

Avant d'aborder la présentation de la typologie, il convient d'abord de préciser le cadre de notre étude dans le contexte des interfaces graphiques et de mentionner certaines caractéristiques des modalités cognitives intervenant dans la représentation visuelle de processus sonores.

5.1.1. Interfaces graphiques

La notion d'interface graphique en micro-informatique fait référence à des modalités de communication homme-ordinateur s'inscrivant dans un cadre bien défini, associant un écran graphique, une souris et un clavier alphanumérique. Le lien entre la commande effectuée par l'intermédiaire de la souris et l'affichage graphique est assuré par la matérialisation d'un curseur de forme programmable, dont la position à l'écran est calculée à partir des déplacements relatifs de la souris. Le clavier est utilisé comme accès complémentaire pour la saisie de données alphanumériques, le déclenchement de fonctions particulières ou le déplacement du curseur, au moyen des touches de navigation. Les applications informatiques dans les domaines musical et sonore font souvent intervenir des modalités d'interface complémentaires. Ainsi, l'écoute du son produit est réalisée soit directement à partir de la sortie sonore de l'ordinateur, soit par la commande de synthétiseurs externes, notamment à travers le protocole MIDI. Celui-ci permet aussi l'utilisation d'interfaces gestuelles spécifiques, reprenant en particulier celles des différentes familles d'instruments, pour l'acquisition de données musicales. Certains systèmes informatiques de production sonore sont également dotés, parallèlement à l'usage de la souris, de surfaces de contrôle comportant un ensemble de boutons et potentiomètres ou reproduisant les réglages d'une console de mixage. L'étude que nous envisageons des principales interfaces graphiques ne fera pas d'autre hypothèse concernant les modalités de communication homme-ordinateur complémentaires au système clavier-écran-souris que celle de la possibilité d'une écoute du résultat sonore.

Les premiers principes des interfaces graphiques, comme l'utilisation de souris ou de fenêtres graphiques, sont attribués à Doug Engelbart pour la conception, vers le milieu des années 60, du système NSL, et ont été développés dans les années 1970 aux laboratoires de Xerox Parc pour la conception du système Xerox Star [SHN 98] [SMI 82]. Ils répondent à l'acronyme WIMP pour *Windows, Icons, Menus and Pointing* (fenêtres, icônes, menus et pointage) et sont décrits par Michel Beaudouin-Lafon dans le chapitre 6. Nous ne développons ici que les concepts d'interfaces graphiques spécifiques aux applications musicales et sonores et ne traitons pas de leurs principes généraux.

5.1.2. Représentation graphique des processus sonores

Les interfaces graphiques utilisées en production musicale et sonore constituent un champ d'étude singulier, dans la mesure où elles fournissent un support visuel à un processus de production dont le résultat est essentiellement de nature sonore. La conception de telles interfaces réside donc avant tout dans les formes possibles de correspondance ou, à l'inverse, d'indépendance entre modalités de cognition visuelle et auditive, dont les caractéristiques suivantes, dans la perspective qui nous intéresse, peuvent être mentionnées :

- *Instantanéité/séquentialité*. La prise de connaissance d'une scène visuelle, dans toute sa complexité, est quasi instantanée, alors qu'une information sonore se déploie dans le temps. Cette différence essentielle détermine à la fois l'intérêt et la forme particulière que revêtent ce que nous désignons dans la suite sous le terme de représentations temporelles.

- *Staticité/dynamicité*. Corrélativement à la remarque précédente, c'est le plus souvent sous forme statique que l'image est exploitée dans les interfaces graphiques pour la production sonore, les principaux changements à l'affichage résultant directement d'une manipulation de l'utilisateur, sauf dans des cas particuliers comme le déplacement d'un curseur sur une image fixe pendant l'écoute.

- *Localisation dans l'espace*. Le système auditif est doté d'une capacité autonome, plus imprécise que celle de la vision, de localisation de sources sonores dans l'espace, fournissant sous certaines conditions des informations d'incidence par rapport au repère de la tête [BLA 97]. La commande de dispositifs de spatialisation, effectuant notamment la simulation de la position de sources sonores, pourra donc s'appuyer naturellement sur une représentation géométrique de l'espace simulé.

- *Figuration/abstraction du support visuel*. Le traitement des informations visuelles par notre cerveau s'adapte aussi bien à des reproductions figuratives, éventuellement simplifiées, de notre environnement réel, qu'à des diagrammes abstraits décrivant des réseaux de relations complexes ou à des représentations symboliques, iconiques ou simplement textuelles. Cette capacité de gestion de différents niveaux d'abstraction visuelle est utilisée par les différentes fonctions de représentation et de manipulation des informations musicales et sonores des interfaces graphiques.

- *Différences de quantité d'information*. A fréquence et dynamique d'échantillonnage constantes, la quantité d'information, ou nombre de bits utilisés pour représenter les contenus sonores, augmente linéairement avec leur durée. L'utilisation d'un support visuel statique pour représenter les phénomènes sonores sur un certain intervalle de temps a donc en général pour contrepartie la nécessité d'une réduction de cette quantité d'information, correspondant à une description se situant à un niveau d'abstraction plus élevé. Ceci est d'autant plus vrai lorsque l'interface est utilisée non seulement pour visualiser le processus sonore, mais aussi pour le contrôler. Dans ce cas, illustré par la figure 5.1, la génération de

l'information manquante entre la représentation graphique et le signal sonore est assurée par un modèle de synthèse, fournissant une représentation paramétrique de signaux sonores, la quantité d'information transmise par unité de temps étant figurée par l'épaisseur du trait :

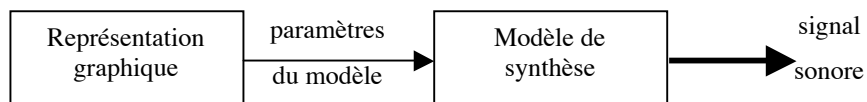


Figure 5.1. Utilisation d'un modèle de synthèse pour la production d'un signal sonore à partir d'une représentation graphique

Il est possible d'établir une mise en parallèle de ce schéma avec une situation d'exécution instrumentale traditionnelle, dans laquelle la partition contient l'ensemble des prescriptions du compositeur sous une forme symbolique (représentation graphique), le modèle de synthèse correspondant à la combinaison des particularités implicites, dans la notation, de fonctionnement de l'instrument (vibration acoustique, modes de jeu, etc.) et du processus d'interprétation. C'est par ce dernier qu'est principalement assurée, à travers le geste de l'instrumentiste, la génération d'information manquante, en spécifiant par des grandeurs physiques (force, vitesse, temps d'occurrence, etc.) les indications données sous forme discrète par la partition.

5.1.3. Principes et résumé de la typologie

Les interfaces graphiques constituent, du point de vue du concepteur informaticien, le support de médiation entre l'utilisateur et les fonctions de l'application informatique. Nous pourrions donc structurer notre typologie selon un principe fonctionnel, mais l'étude des logiciels montre qu'il peut exister plusieurs types d'interfaces graphiques pour une même fonction, ou inversement et plus rarement, qu'une même métaphore graphique soit appliquée à des types de fonctions différents. Il paraît ici plus approprié d'adopter un critère de classification des interfaces graphiques qui soit directement lié à l'expérience cognitive de l'utilisateur, comme le degré d'abstraction des représentations proposées en rapport avec les phénomènes sonores qu'elles illustrent. La typologie est donc organisée selon les catégories suivantes, par ordre croissant d'abstraction :

- *objets de commande virtuels*, reposant sur des représentations figuratives à l'écran d'objets physiques sur lesquels l'action de l'utilisateur se traduit par une commande en temps réel : potentiomètres, boutons, déplacement de sources sonores, etc.

- *représentations temporelles*, faisant apparaître une dimension spatiale liée au temps, constituant un degré d'abstraction supplémentaire,
- concepts de *programmation visuelle* utilisés notamment pour représenter les procédures intervenant dans le calcul de fonctions complexes par assemblage de modules élémentaires,
- représentations utilisées pour la *gestion de données multiples*, liées à la complexification des environnements informatiques en production musicale et sonore.

5.2. Objets de commande virtuels

Cette catégorie regroupe les interfaces graphiques associées à des métaphores d'objets, dont la manipulation, par l'intermédiaire de la souris, repose sur une interaction gestuelle simulant une commande réelle : déplacement d'un curseur ou d'une source sonore, clic correspondant à la pression d'un bouton. Le terme de *virtuel*, dont l'emploi généralisé donne lieu actuellement à des significations très diverses, est ici employé dans une acception traditionnelle : en optique géométrique, l'image d'un objet dans un miroir est source virtuelle des rayons lumineux au sens où tout se passe *comme si* ils en étaient effectivement issus. De manière analogue, les objets graphiques représentés à l'écran sont ici qualifiés de virtuels dans la mesure où leur apparence sensible, leurs modalités de manipulation et les effets qui en résultent tendent à simuler ceux d'objets de commande réels. Contrairement à de nombreuses applications interactives, comme les jeux vidéo, faisant appel à des interfaces homme-ordinateur proposant une navigation dans des mondes virtuels tridimensionnels, les interfaces graphiques destinées à la production sonore sont essentiellement conçues autour d'espaces bidimensionnels.

Une des principales caractéristiques des objets de commande virtuels est le fonctionnement en temps réel qui leur correspond, qu'il s'agisse du déclenchement d'une action particulière, ou de la variation continue d'un paramètre en fonction d'un déplacement. La notion de temps réel est ici envisagée du point de vue de l'utilisateur, c'est-à-dire relativement à un temps de latence imperceptible ou peu critique, typiquement compris entre 1 et 50 millisecondes selon les cas, entre commande et stimuli générés en réponse par le système. Dans le cas d'applications faisant intervenir une synthèse ou un traitement sonore en temps réel, la commande gestuelle s'appuie sur une double rétroaction, visuelle et sonore (figure 5.2).

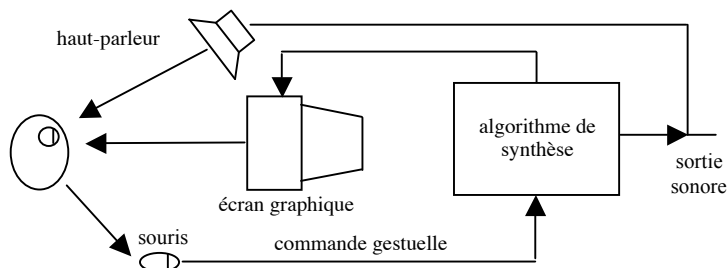


Figure 5.2. *Système d'interaction associé à un algorithme de traitement en temps réel, faisant intervenir une commande gestuelle à la souris et une double rétroaction visuelle et sonore.*

Ainsi, le logiciel GRM Tools de l'Ina-GRM [VIN 91], conçu comme outil de manipulation interactive de matériaux sonores, est constitué d'algorithmes de traitement, dotés chacun d'un panneau de contrôle comportant un ensemble de potentiomètres associés à chaque paramètre (figure 5.3). L'expérience d'utilisation montre qu'un tel système d'interaction est propice à une intériorisation rapide du lien existant entre geste de commande et effet sonore résultant, quels que soient l'algorithme et les paramètres utilisés.

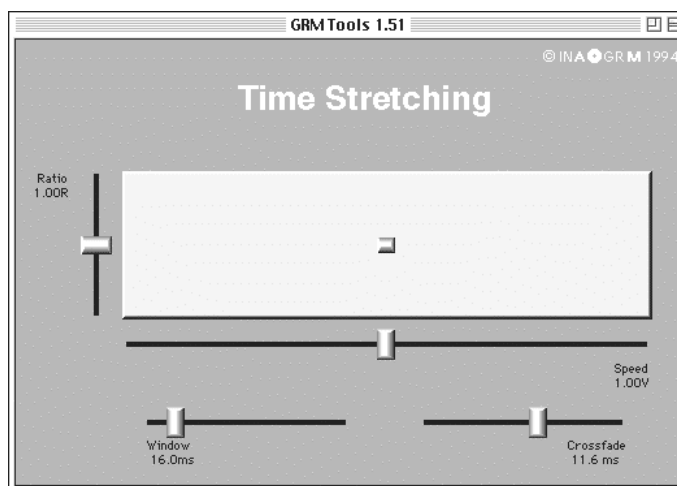


Figure 5.3. *Panneau de commande associé à l'algorithme « Time Stretching » du logiciel GRM Tools, (© Ina-GRM)*

5.2.1. Potentiomètres

L'ergonomie du studio et en particulier des consoles de mixage est profondément ancrée auprès des professionnels de production sonore. Les interfaces des systèmes informatiques développés dans ce domaine tendent donc généralement à reproduire celles des équipements analogiques traditionnels, même si leur architecture interne en est très différente. Les métaphores de potentiomètres sont ainsi couramment utilisées pour toute commande de paramètres devant prendre des valeurs continues dans un intervalle donné. La cohérence psychophysique nécessite le choix de bornes et d'une loi de variation conformes à celles de notre perception auditive du paramètre commandé, notamment logarithmique pour des hauteurs ou des intensités sonores. La géométrie de variation est déterminée par la sémantique de la commande, rotative pour un contrôle panoramique désignant une direction d'incidence, linéaire verticale pour un gain en amplitude, linéaire horizontale pour une position temporelle ou une vitesse de lecture, etc. Certains logiciels comme GRM Tools ont systématisé l'utilisation de plages bidimensionnelles (figure 5.3), équivalents graphiques de *joysticks*, pour la commande simultanée et indépendante de deux paramètres continus.

5.2.2. Variation simultanée de paramètres multiples

Plusieurs systèmes ont été conçus pour dépasser la limitation de commande simultanée de deux paramètres inhérente à la métaphore de potentiomètres graphiques.

5.2.2.1. Interpolation par boules

Le système le plus élaboré est incontestablement celui dit d'interpolation par boules, conçu par Jean-François Allouis pour la station audionumérique Syter [TER 98]. Etant donné un ensemble de N paramètres indépendants, l'utilisateur a la possibilité de créer des boules B_i de rayon R_i dans un espace bidimensionnel, associées à des configurations particulières $(a_{i1}, a_{i2}, \dots, a_{iN})$ des paramètres (figure 5.4). L'intersection entre deux boules distinctes doit être vide, contrainte imposée au moment de leur création. Si M boules ont ainsi été créées, la valeur de chaque paramètre, en mode de jeu, sera calculée en fonction de la distance d_i de la position du curseur de la souris à chaque boule B_i , selon la loi suivante :

– si le curseur de la souris est à l'intérieur de la boule i , les valeurs prises sont $(a_{i1}, a_{i2}, \dots, a_{iN})$.

$$- \text{sinon, } a_j = \frac{1}{k} \cdot \sum_{i=1, \dots, M} \frac{R_i}{d_i} \cdot a_{ij} \quad 1 \leq j \leq N, \text{ avec } k = \sum_{i=1, \dots, M} \frac{R_i}{d_i} \quad [5.1]$$

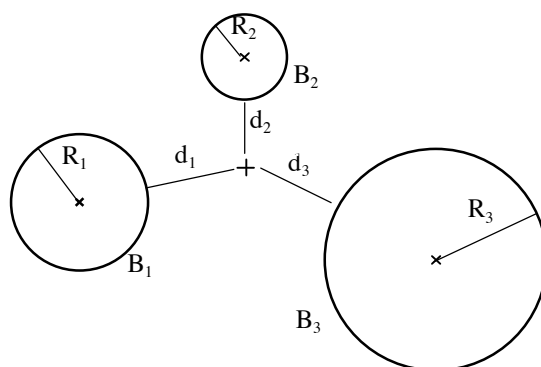


Figure 5.4. *Interpolation par boules*

Cela signifie que lorsque le curseur de la souris se trouve à l'intérieur d'une boule donnée, les paramètres prennent les valeurs correspondantes ; lorsque le curseur se trouve dans l'espace qui les entoure, les paramètres sont calculés comme interpolation des valeurs correspondant aux différentes boules, s'approchant d'autant plus des valeurs d'une boule donnée que son diamètre est important et que sa distance au pointeur de la souris est faible. Plus précisément, les boules ont une influence proportionnelle à leur rayon et inversement proportionnelle à leur distance de la souris. Certains utilisateurs attribuent ainsi la spécificité du « son » du système Syter à la loi de variation dynamique particulière induite par cette pondération en $1/d$. La formule [5.1] peut en fait être généralisée en remplaçant les d_i par $f(d_i)$ où f est une fonction croissante, nulle et continue en zéro, strictement positive ailleurs. Toute puissance d'exposant positif convient donc en particulier.

5.2.2.2. *Autres procédés d'interpolation*

Parmi les autres procédés de commande de paramètres multiples, on peut citer :

- le « super slider » conçu à l'Ina-GRM par Serge de Laubier dans le cadre du logiciel MIDI Formers de l'Ina-GRM [LAU 91], potentiomètre linéaire divisé en tronçons associés chacun à la variation entre deux réglages prédéfinis ;
- le logiciel Inter-Actors (Mil Productions), figurant à l'écran un ensemble d'icônes associées à des « acteurs » indépendants produisant chacun un certain type de données MIDI en fonction de leur distance à une origine. Le déplacement du point d'origine a donc pour effet la variation simultanée des données de l'ensemble des acteurs.

5.2.3. Spatialisation

La commande de dispositifs de spatialisation, destinés notamment à la simulation du positionnement de sources sonores, fait naturellement appel à des interfaces graphiques reposant sur des représentations géométriques. Les sources peuvent ainsi être disposées et déplacées dans un espace virtuel dans lequel la position de l'auditeur est également figurée. Le caractère bidimensionnel des interfaces graphiques a en général pour effet la limitation de l'espace de variation à un plan horizontal, la commande de positionnement en élévation, correspondant à un paramètre perceptif de plus faible sensibilité, pouvant éventuellement être réalisée séparément. La figure 5.5 montre un écran de commande du logiciel Spatialisateur de l'Ircam, combinant la simulation de la position de sources et d'un effet de salle, l'espace étant divisé en quatre zones associées à des qualités acoustiques différentes. De plus, le logiciel est doté d'un modèle de directivité des sources, dont la commande en rotation modifie l'orientation par rapport à l'auditeur et donc le rapport entre les champs acoustiques direct et réverbéré.

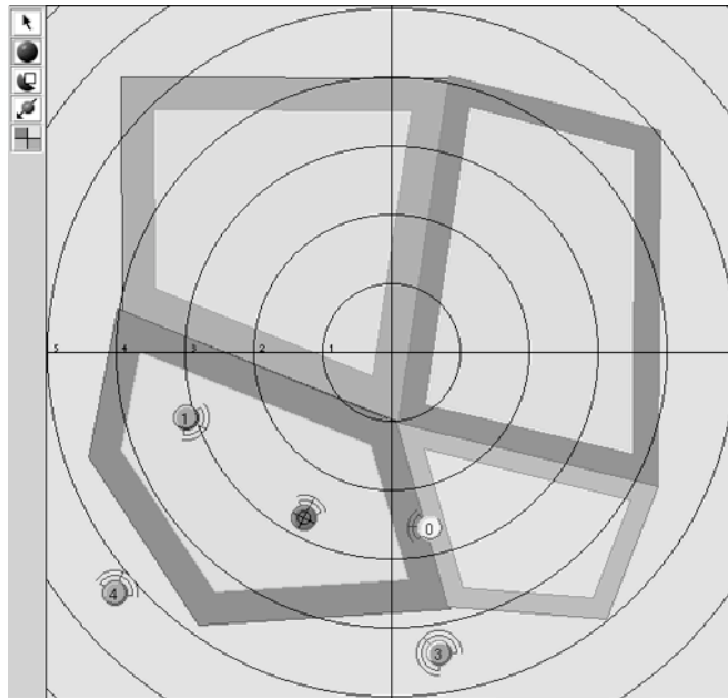


Figure 5.5. Logiciel Spatialisateur (©Ircam-Espaces Nouveaux)

La figure 5.6 montre l'écran de commande du logiciel MusicSpace de Sony CSL [PAC 98], dans lequel sont matérialisées, en plus des avatars de sources instrumentales et d'auditeur, des contraintes programmables sur la position des différents objets : déplacement conjoint de plusieurs sources, ambitus d'éloignement ou de rotation par rapport à l'auditeur, etc. A titre d'exemple, les deux cercles, centrés sur l'auditeur, auxquels est relié le piano, donnent les limites de déplacement de celui-ci. De même, les demi-droites à angle droit auxquelles sont reliés les instruments de la section rythmique (guitare, contrebasse, batterie) définissent les limites de déplacement angulaire de celle-ci.

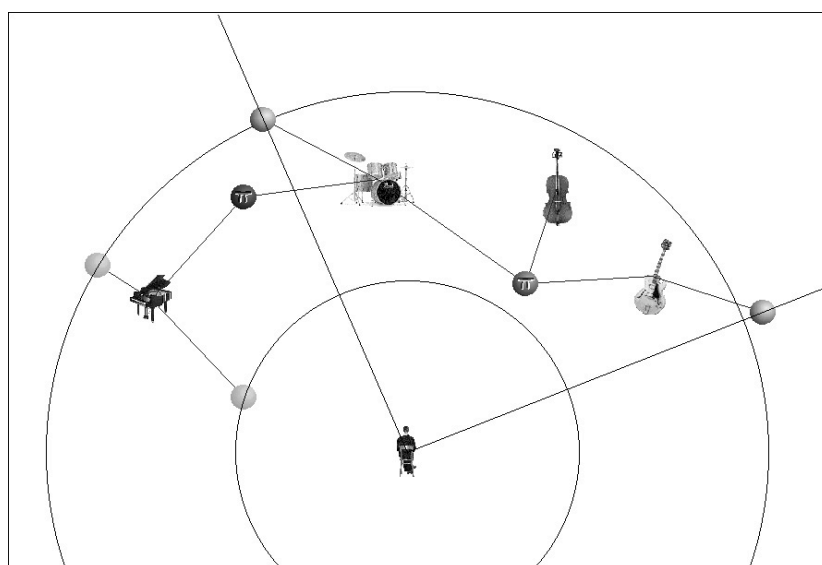


Figure 5.6. Logiciel MusicSpace (©Sony CSL)

5.2.4. Panneaux de commande

Les panneaux de commande sont des zones regroupant plusieurs objets graphiques autour d'un ensemble cohérent de fonctions. Un exemple caractéristique en est la généralisation des blocs de boutons de lecture, rembobinage, avance rapide, affichage numérique du temps, etc., inspirée de l'ergonomie des magnétophones et magnétoscopes, pour la commande de positionnement de tous types de supports de médias temporels, y compris ceux qui reposent sur un accès aléatoire (disque durs et optiques).

Dans le domaine de l'informatique musicale, l'implantation de synthétiseurs virtuels, simulant les fonctions d'appareils analogiques existants ou anciens,

s'accompagne de la reproduction fidèle de leurs faces avant (figure 5.7), avec comme argument commercial la possibilité de remplacement de tout cet appareillage, au prix du logiciel, par les fonctions de l'ordinateur.

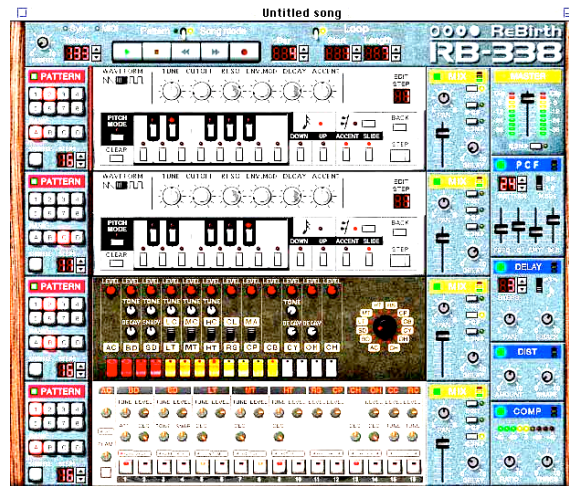


Figure 5.7. Face avant du logiciel Rebirth (© Steinberg)

Les environnements les plus avancés comme les logiciels Max (Ircam-Opcode) ou Scope (Creamware) sont dotés de fonctions de programmation portant non seulement, comme on le verra plus loin, sur les algorithmes des traitements, mais aussi sur la constitution par l'utilisateur de panneaux de commande personnalisés, qu'il organise à sa guise.

5.3. Représentations temporelles

5.3.1. Introduction

La dénomination de représentation temporelle se réfère ici à toute représentation graphique faisant apparaître une figuration spatiale explicite du temps. Quelles qu'en soient les origines, on ne peut que constater la convergence dans la culture occidentale des représentations du temps associées à un déplacement de gauche à droite sur un axe horizontal, qui sont également la règle générale dans les interfaces graphiques. Le lien avec l'écriture est manifeste. Ceci est d'autant plus vrai, dans le domaine qui nous intéresse, si l'on considère que l'écriture phonétique constitue déjà une forme de correspondance entre un phénomène sonore et sa matérialisation spatiale. En musique, le terme de composition est construit sur cette métaphore. De

plus, l'usage de la notation a donné lieu à l'emploi courant des termes de verticalité et d'horizontalité pour rendre compte respectivement de la simultanéité et de la succession d'événements.

L'un des principaux intérêts des représentations temporelles est qu'elles fournissent un point de vue global sur un intervalle de temps donné, avec une visualisation des relations de simultanéité et de succession des différents événements sonores qu'il contient. L'utilisation de fonctions de zoom donne accès aux différents niveaux d'organisation temporelle des sons, de la microstructure à l'échelle de la milliseconde contenue dans un événement donné à une forme globale de l'ordre de la dizaine de minutes. Cette possibilité d'anticipation visuelle, combinée au déplacement d'un curseur pendant l'écoute donnant la synchronisation entre son et image, n'est obtenue qu'au prix d'un enregistrement préalable du contenu sonore, sauf lorsque celui-ci est purement synthétique. L'emploi de supports de stockage donnant un accès quasi instantané à n'importe quelle portion du son permet l'agencement structuré d'extraits sonores sans limitation de complexité et leur manipulation non destructive par l'intermédiaire de pointeurs (montage virtuel).

5.3.2. Notation musicale

La notation a vu, au cours des siècles, sa fonction évoluer relativement aux pratiques musicales en Occident, passant d'un rôle de mémoire de traditions orales, à celui de prescription destinée aux interprètes et de support formel des œuvres [LE VOT 97]. Évoluant concomitamment à la structuration de la théorie musicale, les fondements du système de notation se sont stabilisés à partir de la période classique, reposant notamment sur une représentation discrète des hauteurs et hiérarchique des durées, sur la base d'une pulsation périodique et de ses subdivisions et multiples [LER 85].

L'essor des technologies de publication assistée par ordinateur a permis la réalisation de logiciels d'édition musicale, tels que Finale (Coda Music Technology), dont la qualité professionnelle et la large diffusion ont contribué à une évolution importante de ce secteur d'activité. Parmi les récents développements dans ce domaine, on peut citer le logiciel Igor (Noteheads), dont une des particularités réside dans l'utilisation systématique de la modalité sonore à des fins de contrôle, l'écoute d'une simulation du résultat musical étant possible à tout moment et tenant compte de toutes les nuances, modes de jeu et timbres d'instruments inscrits dans la partition.

La complexification du langage musical au cours du XX^e siècle s'est traduite par la diversification des systèmes de notation, parfois spécifiques à une œuvre donnée.

Il n'existe pas à notre connaissance d'outil logiciel susceptible d'intégrer les différentes formes de notation élaborées par les compositeurs. La raison en est probablement l'abandon à terme de la partition comme seul support de la composition et le rôle croissant joué par l'ordinateur dans la formalisation du langage musical. Doté de fonctions de programmation et d'interfaces multimodales, l'outil informatique est désormais mieux adapté pour représenter et manipuler des structures musicales complexes qu'un support purement graphique. Il convient de citer dans ce domaine le logiciel Acousmographe de l'Ina-GRM [KOE 91], conçu pour élaborer des représentations symboliques à partir d'un repère temps-fréquence fourni par une analyse sonographique, qui peut alors être effacé pour ne faire apparaître que la partie dessinée (figure 5.8).

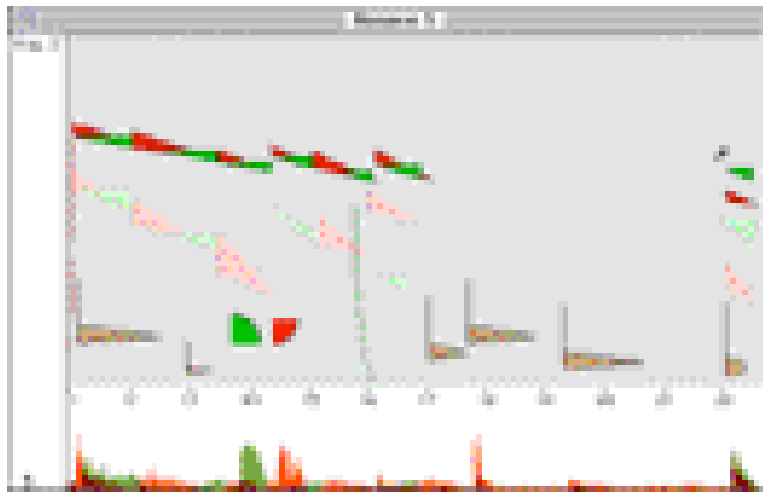


Figure 5.8. Représentation symbolique d'un extrait d'œuvre contemporaine (*Rosace 5*, de François Bayle) réalisée à l'aide du logiciel Acousmographe (© Ina-GRM)

A l'inverse du système de notation traditionnel, une spécificité des représentations informatiques du temps a trait à la linéarisation de l'échelle des temps. Si la description métrique des durées en usage dans la notation musicale, sous la forme de valeurs définies comme fractions rationnelles d'une période de pulsation, semble mieux rendre mieux compte de notre perception du temps musical, ce sont pourtant des représentations reposant sur une échelle linéaire qui prévalent dans les interfaces graphiques. L'écoute du son produit est alors combinée au déplacement à vitesse fixe d'un curseur de lecture le long de l'axe temporel. L'usage de la notation reste possible sous sa forme proportionnelle, dans laquelle les figures rythmiques sont abandonnées, le début des notes étant donné par leur position

horizontale et leur durée par la longueur d'un segment de droite horizontal qui en est issu.

5.3.3. *Piano-roll*

Cette forme de représentation, proche de la notation proportionnelle, reprend le principe des rouleaux perforés de pianos mécaniques (figure 5.9).

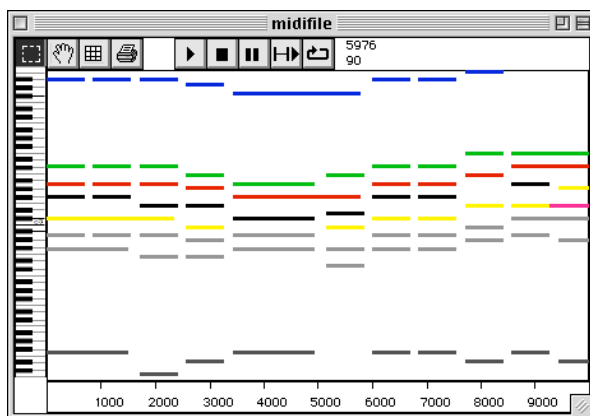


Figure 5.9. Exemple de représentation sous forme de piano-roll
(logiciel OpenMusic, ©Ircam)

Limitée à des échelles discrètes de hauteurs, mais sans contrainte sur la discrétisation de l'axe temporel ou sur le nombre de voies de polyphonie, elle est en particulier utilisée pour la représentation graphique de séquences de notes MIDI, caractérisées par un numéro de note, des instants de début et de fin, et un paramètre de vélocité qui n'est en général pas figuré, ou apparaît par le biais d'une gradation de couleurs.

5.3.4. *Représentation de signaux sonores*

Les techniques d'analyse des signaux audio numériques fournissent différents points de vue sur les sons, caractérisés généralement par des vecteurs de paramètres indexés sur le temps dont peuvent être dérivées, après réduction d'information appropriée, diverses formes de représentations graphiques. Celles-ci sont classées dans la suite en fonction du nombre de dimensions spatiales qu'elles mettent en œuvre. Dans le cas de l'utilisation de modèles de synthèse dotés de procédures inverses sous la forme de modèles d'analyse (le critère d'inversion résultant des

propriétés mathématiques du modèle ou, tout du moins, d'une validation perceptive), il est possible d'utiliser ces représentations à des fins de manipulation sur les paramètres du modèle, selon le diagramme donné en figure 5.10 qui vient compléter celui de la figure 5.1.

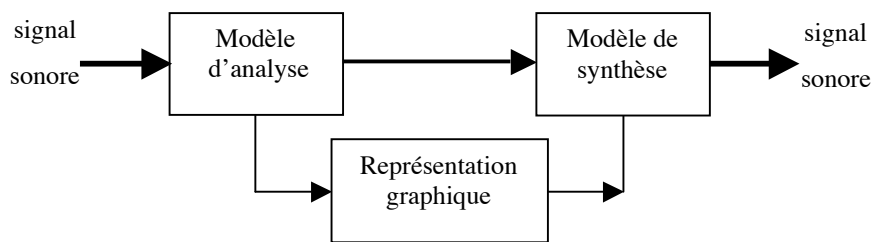


Figure 5.10. Utilisation d'un système d'analyse/synthèse pour la manipulation de signaux sonores à l'aide d'une représentation graphique

5.3.4.1. Fonctions d'une variable numérique

Les différentes formes de représentation graphique de fonctions d'une variable numérique sont données ci-après. Elles s'appliquent aussi bien à des signaux audionumériques qu'à des paramètres de commande de haut niveau, voire à des fonctions dépendant d'une autre variable que le temps.

Les *tables* (figure 5.11 a) sont utilisées pour visualiser des tableaux de données, telles que formes d'onde d'oscillateurs, enveloppes, etc. Elles conviennent à des fonctions caractérisées par un pas d'échantillonnage régulier et un nombre de valeurs fixe, donc un temps d'accès rapide.

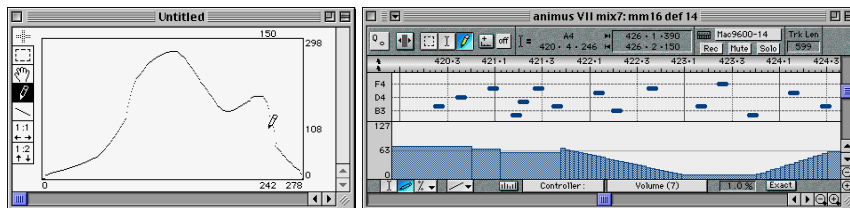


Figure 5.11. a. Edition graphique de table (logiciel Max, ©Ircam-Opcode Systems);
b. piano roll et fonctions en escalier (logiciel Studio Vision, © Opcode Systems)

Les *fonctions en escalier* (figure 5.11 b) s'appliquent à la visualisation de paramètres associés à la sortie d'un système d'échantillonneur/bloqueur, qui reste

constante jusqu'à l'arrivée d'une nouvelle valeur. Elles sont notamment utilisées pour la représentation de contrôleurs continus en MIDI.

Les fonctions affines par morceaux (*PieceWise Linear*, ou *Breakpoint functions* en anglais, ou BPFs) sont caractérisées par un ensemble de couples de valeurs réelles strictement ordonnés sur leur première composante, en général assimilée au temps. Elles sont couramment utilisées pour la représentation et l'édition de paramètres sonores. Contrairement aux tables, elles autorisent un échantillonnage irrégulier, l'insertion d'un nouveau point à une position quelconque et la destruction de tout point existant (figure 5.12 a). Ce système est généralisé à des courbes multiples, soit par l'utilisation de n-uplets strictement ordonnés sur la première variable (échantillonnage commun aux n-1 courbes, voir figure 5.12 b), soit par la réunion de BPFs simples (échantillonnage propre à chaque courbe).

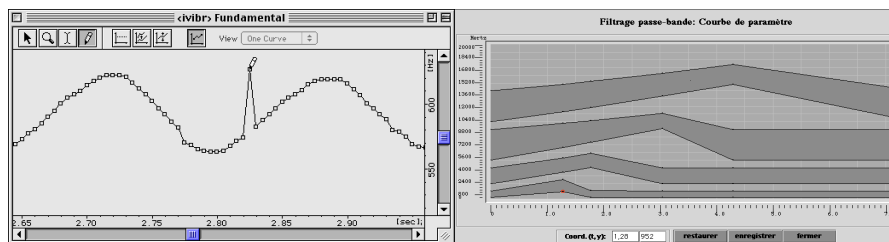


Figure 5.12. Exemples d'utilisation de BPFs. a. Edition dans le logiciel Diphone (© Ircam).
b. Utilisation de BPFs multiples pour un filtrage à 4 bandes évoluant dans le temps (Studio en ligne, © Ircam)

Les *envelopes* sont la forme de représentation la plus utilisée pour les signaux sonores, en particulier dans les logiciels commerciaux. Elles sont obtenues par remplissage entre les courbes de deux fonctions positive et négative (voir figure 5.13), qui majorent et mineurent respectivement celle du signal en tout point. Elles proposent donc une réduction d'information reposant sur le choix des valeurs extrêmes des échantillons de l'intervalle temporel relatif à chaque pixel horizontal. Certains logiciels ne comportent qu'une courbe positive calculée à partir du redressement du signal de départ. Lorsque le rapport de grossissement du zoom est suffisant, typiquement lorsque l'intervalle de temps correspondant au pixel est inférieur à la durée des alternances du signal, un principe d'échantillonnage plus classique est substitué à celui de l'enveloppe, qui est alors remplacée par la courbe du signal lui-même. Cette forme de visualisation des valeurs extrêmes est bien adaptée à des opérations de segmentation temporelle, les silences étant repérés par des amplitudes faibles. Selon les applications, d'autres systèmes de réduction d'information peuvent être utilisés (moyenne quadratique notamment pour l'énergie).

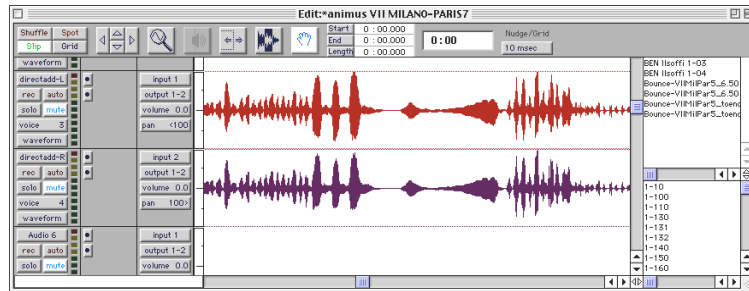


Figure 5.13. Exemple d'utilisation d'enveloppes (logiciel Pro Tools, © Digidesign)

5.3.4.2. Représentation plane de fonctions de deux variables

Le tracé de représentations temps-fréquence de type *sonagramme* est obtenu à partir du module de la Transformée de Fourier à Court Terme (TFCT) appliquée aux signaux sonores, selon une décomposition par bandes de fréquences fixes et équiréparties. La taille de la fenêtre d'analyse détermine un compromis entre résolutions temporelle et fréquentielle. Les amplitudes sont codées en niveaux de gris ou en échelles de couleur selon une correspondance qui, selon les applications, est rendue accessible à l'utilisateur.

Un des intérêts des représentations sonographiques est leur adéquation à certaines modalités de la perception auditive relatives au traitement des informations acoustiques dans des bandes de fréquence différentes, tout en faisant apparaître de manière distincte des composantes du son qui ne sont pas nécessairement entendues isolément. L'utilisation de sonagrammes est en général combinée à celle d'enveloppes qui font mieux apparaître les variations globales en amplitude. Le caractère inversible du modèle de TFCT rend possible, conformément au schéma de la figure 5.10, la mise en œuvre de fonctions d'édition graphique dans le plan temps-fréquence. C'est en particulier le cas du logiciel Audiosculpt de l'Ircam (figure 5.14), qui permet de spécifier des zones de filtrage par dessin sur la représentation sonographique.

D'autres modèles, comme la transformée en ondelettes, fournissent des décompositions des signaux sonores selon des familles discrètes d'« atomes » élémentaires, fonctions à support limité en temps et en fréquence, dont des représentations graphiques bidimensionnelles temps-fréquence ou temps-échelle peuvent être dérivées [DE POL 91]. Bien que leur utilisation à des fins de manipulations sur les sons puisse être envisagée, la mise en œuvre de telles

représentations reste à l'état de prototype dans les laboratoires et n'a pas trouvé jusqu'ici d'applications dans des logiciels de production sonore.

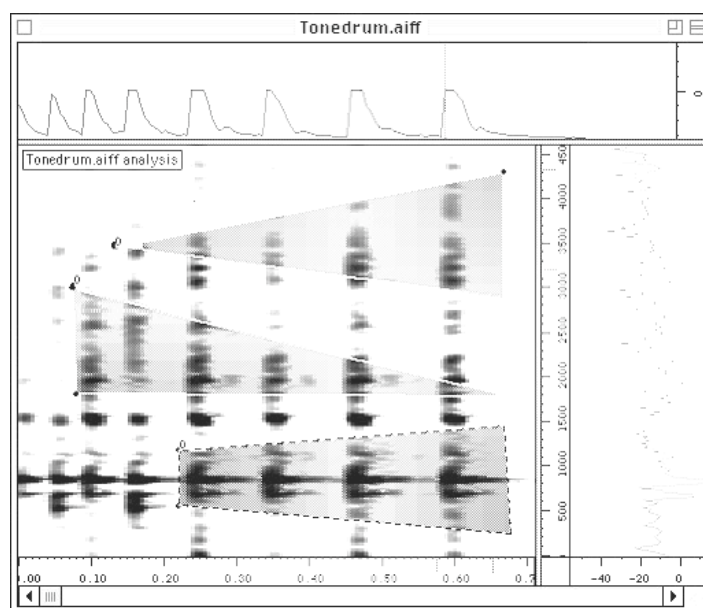


Figure 5.14. Filtrage temps-fréquence au moyen du logiciel Audiosculpt(© Ircam)

5.3.4.3. Représentations tridimensionnelles

Les représentations tridimensionnelles sont des projections planes de représentations graphiques de fonctions de deux variables réelles. Ainsi, les spectrogrammes, calculés comme les sonagrammes à partir de la TFCT, fournissent une vue en relief des amplitudes en fonction des variables temporelle et fréquentielle. De telles représentations, peu maniables, sont principalement utilisées à des fins de visualisation et non de manipulation. On peut citer à titre d'exception le logiciel OpenSoundEdit du CNMAT [CHA 98], conçu à partir d'un système interactif lié à un rendu visuel tridimensionnel pour la manipulation directe de modèles de signaux sonores. La figure 5.15 donne un exemple de représentation d'un signal sonore avec OpenSoundEdit comme superposition de composantes sinusoïdales dont les amplitudes varient au cours du temps, la dimension de fuite correspondant à l'axe fréquentiel.

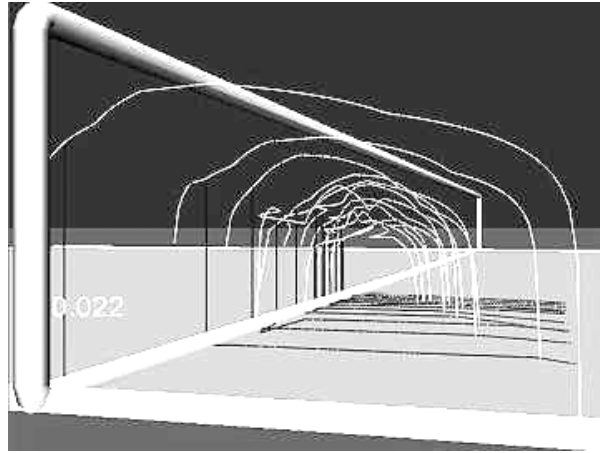


Figure 5.15. Représentation tridimensionnelle d'un signal sonore au moyen du logiciel *OpenSoundEdit* (© CNMAT, Université de Berkeley, Californie)

5.3.5. Interfaces de séquencement

La fonction de séquenceur, destinée à la réalisation de séquences musicales et sonores, est l'une des plus répandues parmi les logiciels commerciaux de production sonore et musicale. De telles applications permettent le positionnement dans le temps et la lecture synchrone de séquences élémentaires associées à différents types de données temporelles. Leurs interfaces graphiques reposent sur le principe du magnétophone multipiste, les différentes pistes partageant un même axe temporel horizontal et faisant appel à des représentations adaptées au type de données qu'elle contiennent : enveloppes pour les échantillons sonores, piano-roll pour les notes MIDI, fonctions en escalier pour les contrôleurs MIDI, BPFs pour les paramètres de commande de traitements sonores (figure 5.16). Cette métaphore du magnétophone multipiste est en général intimement liée à celle de console de mixage, chaque piste étant associée à une tranche particulière d'une console virtuelle, c'est-à-dire à des ressources de traitement allouées à l'avance pour toute la durée de la séquence. Le lien fonctionnel entre objets de commande virtuels du traitement et l'interface de séquencement est assuré par un système d'automation permettant l'enregistrement des valeurs des paramètres au cours du temps et leur représentation sous forme de BPFs dans des pistes particulières en vue de leur édition graphique ultérieure. Lorsque les ressources de traitement associées à chaque objet de la séquence peuvent être allouées (et désactivées) indépendamment des autres éléments qu'elle contient, la métaphore de multipiste peut alors être remplacée par une simple

superposition, la signification attribuée à la dimension verticale étant laissée à la liberté de l'utilisateur. C'est en particulier le cas de l'éditeur de maquettes du logiciel OpenMusic de l'Ircam, décrit dans le chapitre 2, qui propose en outre l'établissement à ce niveau de description de connexions pour l'échange de données entre éléments de la séquence.

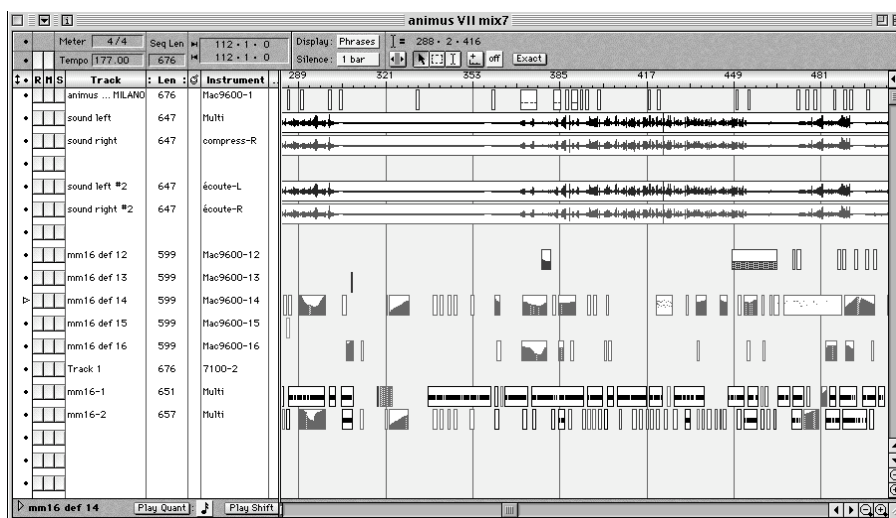


Figure 5.16. Exemple d'interface de séquençage combinant des types de données temporelles multiples (logiciel Studio Vision, © Opcode Systems)

On peut généralement distinguer deux phases distinctes dans le travail de production sonore, correspondant respectivement à la préparation individuelle et à l'assemblage des éléments constitutifs de la séquence, même si des aller-retour sont possibles entre ces deux étapes. En comparaison avec d'autres formes de représentations temporelles, qui sont surtout utilisées pour l'édition du matériau de base, les interfaces de séquençage correspondent donc à un travail, à un niveau plus macroscopique, à une échelle temporelle plus vaste. Les interfaces de séquençage interviennent également pour un contrôle de plus haut niveau des paramètres sonores, comme dans le cas du logiciel Diphone de l'Ircam [ROD 97], où l'assemblage graphique de briques élémentaires se traduit par le calcul par interpolation d'un ensemble de BPFs correspondant à des paramètres de bas niveau (figure 5.17).

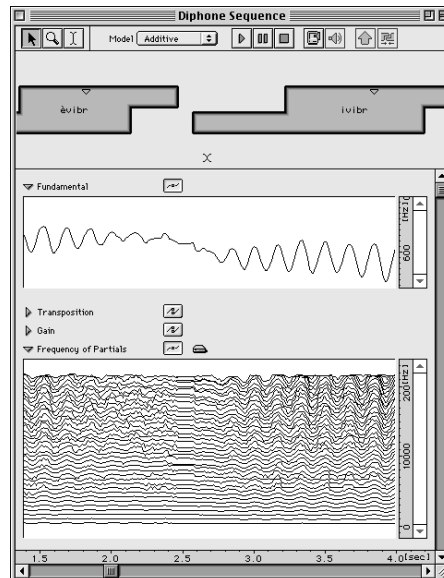


Figure 5.17. Calcul par interpolation d'un ensemble de BPFs à partir d'une interface de haut niveau (logiciel Diphone © Ircam).

5.4. Langages visuels

Contrairement aux objets de commande virtuels et aux représentations temporelles, reposant sur une manipulation directe des paramètres de synthèse ou de traitement, les langages visuels fournissent un point de vue plus abstrait sur les sons, en décrivant les algorithmes de production des informations musicales et sonores. L'histoire de l'informatique musicale est indissociable de la constitution de langages informatiques adaptés aux différentes fonctions de calcul assurées par l'ordinateur à des fins musicales. Les différentes contributions dans ce domaine ont abouti à la conception de nombreux langages textuels, dont les plus largement diffusés sont les langages Music V et Csound. Le développement ultérieur des interfaces graphiques a permis la mise en œuvre de langages visuels, dont l'utilisation sous la forme de systèmes de *patches* s'est généralisée dans le domaine de l'informatique musicale, notamment à partir du logiciel Max développé à l'Ircam [PUC 88]. La représentation graphique d'un algorithme de calcul, par la vue d'ensemble qu'elle propose, fournit un support de description d'un abord plus immédiat qu'un langage textuel, contraint par la forme séquentielle des énoncés. Les langages visuels présentent cependant certaines limites, relatives notamment aux opérations d'itération, aux instructions

conditionnelles ou à la précision de description des données, qui donnent lieu le cas échéant à l'utilisation complémentaire de langages textuels.

5.4.1. Systèmes de patches

Les systèmes de *patches* représentent la forme la plus usitée des interfaces de programmation visuelle en informatique musicale, même s'ils ne sont pas nécessairement dotés de toutes les caractéristiques constitutives d'un langage informatique. Ils reposent sur la notion de modularité, la réalisation d'une fonction complexe étant obtenue par assemblage d'objets fonctionnels élémentaires figurés par des rectangles ou icônes, les sorties des uns étant reliées aux entrées des autres par des connexions sous forme de fils. Leur conception reprend la métaphore du studio de production analogique, dans lequel l'obtention d'un son résultait d'un câblage entre un ensemble de modules préexistants, avec la possibilité à tout instant de supprimer des connexions existantes ou d'en créer de nouvelles. L'introduction d'abstractions, ou objets contenant eux-mêmes des sous-patches, permet une description hiérarchique dans laquelle la complexité des niveaux sous-jacents est masquée. Il est intéressant de constater que cette forme de représentation a été utilisée dans des logiciels en rapport avec des sémantiques opératoires différentes : systèmes d'envoi de message et de flots de données dans des logiciels tels que Max, programmation fonctionnelle en Lisp dans les environnements PatchWork et OpenMusic de l'Ircam. Des exemples de patches réalisés dans les environnements OpenMusic et jMax sont donnés par Gérard Assayag et François Déchelle dans les chapitres 2 et 4.

5.4.2. Programmation orientée-objet

Certains environnement musicaux, comme OpenMusic, comportent des fonctions de programmation orientée-objet et sont à ce titre dotés d'interfaces graphiques pour l'édition des structures de classes, des arbres d'héritage ou des valeurs d'instances. Ces fonctions ne sont pas spécifiques aux applications musicales ou sonores et ne seront pas développées ici.

5.5. Gestion de données multiples

La complexification des environnements informatiques destinés à la production musicale et sonore se traduit par une multiplication des types de données gérées par chaque environnement et par un accroissement du nombre d'informations pouvant être traitées simultanément. La prise en compte de ces problèmes aboutit à la réalisation d'interfaces de gestion et de navigation adaptées.

5.5.1. Gestion de types de données multiples

La gestion conjointe et structurée de différents types de données au sein d'un même environnement fait de plus en plus appel à des interfaces spécifiques, intégrées aux applications et reproduisant les interfaces de gestion de fichiers telles que le Finder d'Apple ou l'Explorer de Windows. L'exploitation des différents types de données dans l'application est généralement facilitée par l'utilisation de fonctions de glisser-déposer. L'avantage de cette forme de structuration liée à l'utilisation de documents composites est la simplicité des échanges de fichiers et le maintien de la cohérence de l'ensemble des données entre elles. Elle a pour alternative, en particulier dans les systèmes Unix, la gestion simultanée de fichiers multiples selon une organisation prédéterminée dont la cohérence doit être maintenue par l'utilisateur.

5.5.2. Recherche et navigation dans les bases de données

La capacité atteinte par les supports de stockage, conjuguée aux possibilités des réseaux, donne potentiellement accès, à partir d'un simple micro-ordinateur, à de très grands volumes d'informations. L'utilisation de systèmes de gestion de bases de données est donc appelée à se démocratiser en production sonore et fait déjà l'objet d'applications consacrées au classement de sons selon des critères multiples. Ainsi, le projet Studio en ligne de l'Ircam donne accès, à travers les protocoles de l'Internet¹, à environ 120 000 échantillons sonores différents issus d'instruments de musique représentant environ 130 Go de données, qui représentent une durée d'écoute cumulée d'environ 12 jours. La constitution d'une taxonomie complète associée aux différents descripteurs des échantillons a abouti à la réalisation d'une interface de sélection des sons permettant une navigation interactive à travers l'arborescence des choix possibles. De plus, ce projet a fourni un cadre au développement expérimental de procédures de recherche par contenu sonore, reposant sur des critères de similarité perceptive issus de recherches sur la caractérisation perceptive du timbre [MCA 99], [MCM 99]. Il est ainsi possible de rechercher dans la base de données les échantillons qui se rapprochent le plus, selon une mesure de similarité perceptive, d'un son de départ proposé par l'utilisateur (figure 5.18). Il ne s'agit là que de premières avancées d'un domaine d'applications qui sera probablement appelé à un développement important.

1. <http://www.ircam.fr/sol/>

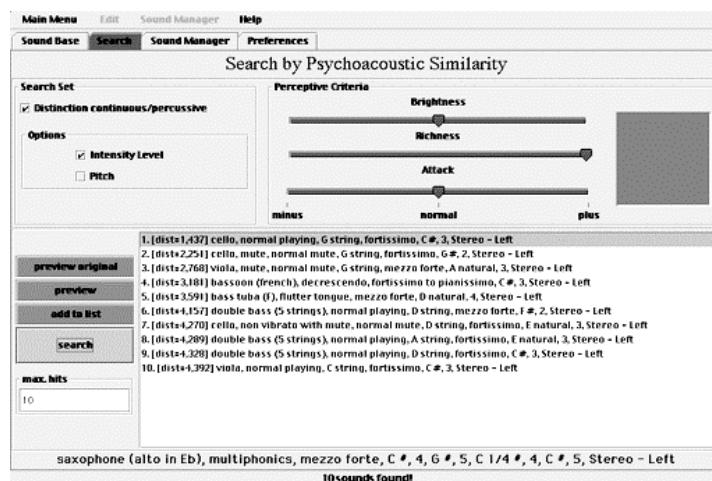


Figure 5.18. Interface de recherche par similarité perceptive (Studio en ligne, © Ircam)

5.6. Conclusion

Nous avons proposé une présentation des principaux types d'interfaces graphiques rencontrées dans les applications de production musicale et sonore, autour d'une typologie conçue pour expliciter les différentes formes de représentations mises en œuvre. Cette description ne saurait cependant revêtir un caractère exhaustif ou définitif, d'autant qu'elle prend pour objet les réalisations d'un secteur en plein essor, dont les concepteurs font constamment preuve d'une grande inventivité. Les évolutions à attendre dans ce domaine, qu'elles concernent les interfaces WIMP ou l'introduction de nouveaux périphériques d'accès ou de visualisation, dépendront en premier lieu des standardisations qui seront opérées dans l'industrie informatique, dont on sait qu'elles obéissent prioritairement à des objectifs de rentabilité à court ou moyen terme. La convergence des investissements actuellement réalisés autour des réseaux laisse notamment augurer, dans la lignée de HTML et de Java, la généralisation de concepts d'interfaces qui pourront s'adapter à la diversification des types de terminaux d'accès.

Bibliographie

[BLA 97] BLAUERT J., *Spatial Hearing. The psychophysics of Human Sound Localization*, Revised edition. MIT Press, 1997.

- [CHA 98] CHAUDHARY A., FREED A., ROWE L. A., « OpenSoundEdit : An Interactive Visualization and Editing Framework for Timbral Resources », *Proceedings of the International Computer Music Conference*, International Computer Music Association, San Francisco, 1998.
- [DE POL 91] DE POLI G., PICCIALI A., ROADS C., *Representations of Musical Signals*, MIT Press, 1991, p. 87-117.
- [JOT 97] JOT J. M., « Efficient Models for Distance and Reverberation Rendering in Computer Music and Virtual Audio Reality », *Proceedings of the International Computer Music Conference*, International Computer Music Association, San Francisco, 1997.
- [KOE 91] KOECHLIN O., VINET H., « The Acousmographie, a Macintosh software for the graphical representation of sounds », *Proceedings of the International Computer Music Conference*, International Computer Music Association, San Francisco, 1991.
- [LAU 91] DE LAUBIER S., TERUGGI D., « MIDI generation of sound morphologies », *Proceedings of the International Computer Music Conference*, International Computer Music Association, San Francisco, 1991.
- [LE VOT 97] LE VOT G., « La notation de la musique en occident médiéval (Signe, Geste et Invention) », *Musique & Notation, Rencontres Musicales Pluridisciplinaires*, Grame, 1997.
- [LER 85] LERDAHL F., JACKENDOFF R., *A Generative theory of tonal music*, MIT Press, 1985.
- [MCA 99] MCADAMS S., « Perspectives on the Contribution of Timbre to Musical Structure », *Computer Music Journal* 23(3), 1999.
- [MCM 99]. MCADAMS S., MISDARIIS N., « Perceptual-based retrieval in large musical databases », *Proceedings of Human Centered Processes, 10th Mini EURO Conference*, Philippe Lenca (ed), 1999.
- [PAC 98] PACHET F., DELERUE O., « A Constraint-based Temporal Music Spatializer », *Proceedings of the ACM Multimedia Conference, Brighton*, 1998.
- [PUC 88] PUCKETTE M., « The Patcher », *Proceedings of the International Computer Music Conference*, International Computer Music Association, San Francisco, 1996, p. 420-425.
- [ROD 97] RODET X., LEFEVRE A., « The Diphone Program: New Features, New Synthesis Methods and Experience of Musical Use », *Proceedings of the International Computer Music Conference*, International Computer Music Association, San Francisco, 1997.
- [SHN 98] SHNEIDERMAN B., *Designing the User Interface, Strategies for Effective Human-Computer Interaction, Third Edition*, Addison Wesley, 1998, p. 446-447.
- [SMI 82] SMITH C., KIMBALL R., VERPLANK B., HARSLEM E., « Designing the Star User Interface », *Byte*, 7(4), p. 242-282.
- [TER 98] TERUGGI D., *Le Système Syter. Son histoire, ses développements, sa production musicale, ses implications dans le langage électroacoustique d'aujourd'hui*. Thèse d'université, Paris 8, 1998.
- [VIN 91] VINET H., « DSP Station, a HyperCard environment for DSP sound processing algorithms », *Proceedings of the International Computer Music Conference*, International Computer Music Association, San Francisco, 1991.