

Suivi de partition, reconnaissance d'événements musicaux Mises au point, perspectives.

Serge Lemouton - Philippe Manoury

IRCAM

1, place I. Stravinsky

75004 PARIS

1. Introduction

A l'occasion d'un travail de recherche musicale actuellement en cours à l'IRCAM, nous tentons de trouver de nouvelles stratégies de suivi de partition.

Les différents systèmes de suivi de partition ont été développés en réponse au problème soulevé par les musiques mixtes (mettant en jeu un interprète humain et une partie électronique). On souhaitait que ce ne soit pas le musicien qui soit esclave de la partie électronique, comme c'est le cas dans une pièce pour bande et instrument. L'objectif était de confier une fonction réellement musicale à la partie électronique, d'en faire un interprète synthétique avec qui le musicien humain pourrait interagir ("Synthetic performer", cf. [VERCOE 84]).

Nous pensons que cette attitude offre beaucoup de souplesse et de musicalité et que, de toute façon, la partition est ce qui caractérise la musique occidentale.

Riches d'une désormais longue expérience dans ce domaine (on trouvera un répertoire commenté des principales pièces réalisées à l'Ircam qui utilisent le suivi de partition), on peut faire la constatation que les systèmes actuellement disponibles ne nous offrent pas toute la fiabilité que l'on peut en attendre. Nous pensons pouvoir être en mesure de les améliorer, en utilisant à la fois ces expériences et des technologies nouvelles.

2. Plan

Après avoir récapitulé les différentes expériences menées à l'Ircam dans le cadre de productions musicales, nous tenterons d'identifier les problèmes des algorithmes de suivi de partition utilisés jusqu'à présent dans ce cadre. Ceci dans le but de les améliorer, de les généraliser. Enfin, dans un dernier temps, nous souhaiterions réaliser un nouveau système de suivi, plus puissant et plus souple.

3. État actuel du suivi de partition

De nombreuses pièces réalisées à l'Ircam utilisent le suivi de partition.

Philippe Manoury a écrit un cycle de pièces qui l'utilisent :

Jupiter (1987) flûte.

Pluton (1988) piano MIDI.

Partition du ciel et de l'enfer (1989) flûte, piano MIDI et orchestre.

Neptune (1991) percussions.

En écho (1993) soprano.

On peut encore citer :

Michael Jarrell : *Congruences*, (1988-89) flûte, hautbois et ensemble.

Michael Jarrell : *Rhizomes*. (1993) 2 pianos et percussion.

José-Luis Campana : *D'un geste apprivoisé...* (1995) pour basson et électronique. Pour cette pièce qui utilise de nombreux modes de jeu sans hauteur définie, a été mis au point un premier système de détection d'enveloppes locales.

Pierre Boulez : *explosante-fixe*, (1991-95).

Michael Obst : *Solaris* (1995) où l'on trouve une utilisation du suiveur d'assez longue haleine puisqu'il s'agit d'un opéra. Le suivi de la partition du piano permet de contrôler le déroulement complet de toute la partie électronique.

Brice Pauset : *Perspectivae syntagmorum..* (1997) Dans cette pièce, l'interprétation du musicien est comparée à la partition "idéale" stockée dans la machine, et cette part d'interprétation humaine influence la partie de l'ordinateur.

Pierre Boulez : *Anthèmes II*, (1997) pour violon et électronique.

L'instrument "suivi" par le système peut-être:

-soit un dispositif de contrôle MIDI (type clavier muet),

-soit un instrument "midifié" par un système de capteurs mécaniques (piano Steinway pour *Pluton* de P.Manoury, flûte Midi utilisée pour *explosante-fixe* de Pierre Boulez), ou acoustiques (micro-contacts pour les vibraphones suivis dans *Neptune*).

-soit un instrument traditionnel, sans aucun capteur (par exemple pour la voix pour *En écho*, la flûte de *Jupiter* ou le violon de *Anthèmes II*), auquel cas on s'efforce de suivre le signal acoustique capté par un micro.

On peut facilement constater que le pourcentage des œuvres mixtes utilisant le suivi de partition reste assez faible. On s'aperçoit que dans la chaîne : soliste -> système de synthèse, beaucoup de nouveautés sont apparues au niveau de la production de sons, de leur spatialisation, etc., mais que pour ce qui est de la connexion entre les deux (c'est à dire tout ce qui concerne les mécanismes de reconnaissance instrumentale et vocale), on en est resté au même point qu'il y a une décennie. Nous pensons qu'il devient prioritaire de s'attaquer à cette problématique car, contrairement à d'autres recherches, elle n'est pratiquement pas abordée ailleurs.

Il n'est plus nécessaire d'opérer une distinction de principe entre "temps réel" et "temps différé". Le choix d'une catégorie ou de l'autre dépend, aujourd'hui, de la nécessité musicale. Ces deux principes peuvent fort bien coexister dans la même œuvre [Manoury 97]. Les méthodes de synchronisation utilisées dans la musique mixte jusqu'à présent, sont extrêmement grossières si on les compare avec celles qui sont présentes dans la musique instrumentale. Le développement des techniques regroupant instruments et dispositif électronique ne peut raisonnablement prendre son essor que si les problèmes de reconnaissance de partition sont résolus. Leur mise en œuvre est, pour le moment, trop

rudimentaire, trop fastidieuse. C'est la raison pour laquelle si peu de compositeurs les ont utilisées.

4. Suivi/reconnaissance

Nous souhaiterions améliorer le suivi en introduisant une notion de reconnaissance d'événements sonores (de différentes natures), ce qui entraînerait plus de sécurité, de vitesse de réaction et de polyvalence.

Il nous paraît pertinent d'effectuer une distinction entre les notions de suivi et de reconnaissance. Par "suivi", il faut considérer une analyse d'un signal (par MIDI ou par captation acoustique) dont le résultat n'est pas connu à l'avance et sur lequel on n'effectue pas, a priori, un choix : toutes les valeurs sont acceptées comme également bonnes. Un suivi d'enveloppe, de formants, peut être effectué dans le but de reporter le résultat de cette analyse sur un dispositif quelconque contrôlant tel ou tel paramètre de synthèse, de spatialisation, de transformation, etc. Par "reconnaissance", il faut entendre la recherche, dans un flux sonore, de tel ou tel événement que l'on connaît à l'avance. En d'autres termes, la reconnaissance est un suivi sur lequel on opère des choix : certains événements sont considérés comme bons et acceptés, les autres sont ignorés. Pour donner un exemple, nous prendrons le cas de l'extraction de hauteur dans une partie vocale. Une partition vocale comprend une notation de hauteur, de durée, de dynamique et de sons phonétiques. Parmi ces derniers figurent des consonnes et des voyelles. Les consonnes produisent du souffle, du bruit, plosives, fricatives, ... Le rôle du détecteur de hauteur est de ne garder que ce qui est assimilable à un son périodique, c'est à dire une hauteur et de rejeter le reste, puis de comparer ce résultat avec une partition mémorisée dans un programme. Dans le cas de cette comparaison, tous les événements de hauteur sont connus à l'avance car ils sont déjà écrits dans une partition.

Il faut également prendre en considération un phénomène important. Dans le cas d'une reconnaissance de partition (nous préférons ici ce terme à celui de suivi), il faut insister sur le fait qu'une partition doit être reconnue indépendamment de l'interprétation qui s'y ajoutera. Dans ce cadre, il est indispensable de prendre en considération les niveaux hiérarchiques qui sont propres à la musique de tradition occidentale, puisque c'est d'elle dont il s'agit. Dans cette hiérarchie, le niveau le plus invariant est celui de la hauteur. On ne rencontre guère de cas où les hauteurs sont laissées à la libre appréciation des interprètes : elles sont dans une très grande majorité scrupuleusement inscrites dans une partition. Bien évidemment, tout cela est une question de définition. On sait que, sauf pour les instruments à clavier, les hauteurs sont en complète fluctuation permanente. Mais le pouvoir discriminatoire de l'oreille (et du cerveau) est ici plus puissant qu'ailleurs. On peut déterminer avec précision à l'intérieur de quels seuils un "la" est un "la" et non un "la bémol". Les notions de dynamiques sont à un niveau hiérarchique moins élevé car elles se situent dans un contexte beaucoup plus subjectif. L'écriture ne propose, au maximum, que sept niveaux différents et ces niveaux ne sont pas productibles avec une exactitude comparable à celle des hauteurs [Manoury 87]. Les durées et les rythmes rentrent dans une autre catégorie car ils ne sont reconnus qu'après coup et non de manière instantanée ce qui est à la base de toute recherche de synchronisation.

Cette hiérarchisation n'implique bien évidemment pas que les dynamiques et le temps ne doivent jamais être pris en compte dans une stratégie de reconnaissance mais seulement que ces notions participent plus activement de la "part interprétable" que de la "part notable" d'une partition. On peut, bien entendu, concevoir une structure musicale "pp" comportant, ici ou là, quelques événements "ff", et, dans ce cas une reconnaissance de seuils dynamiques serait certainement suffisante. Mais il s'agit là d'une procédure locale qui ne traite pas du problème de façon généralisée.

De manière générale, on peut dire que, théoriquement, une stratégie de reconnaissance vise à détecter un événement qui doit être unique dans un contexte déterminé. Dans l'état actuel des choses, les stratégies de reconnaissance des hauteurs proposées analysent une fenêtre d'événements (en général, quatre ou cinq) permettant une tolérance de deux erreurs consécutives au maximum. La difficulté vient du fait que l'on cherche à détecter une hauteur absolue ce qui, dans le cas de la voix par exemple, est assez problématique. Nous donnerons, dans la partie suivante, quelques idées de nouvelles stratégies plus conformes à la réalité musicale, mais qui restent à expérimenter.

5. Critiques des systèmes actuels

Avec toutes ces expériences de reconnaissance de partition décrites ci-dessus, en situation (concerts), nous sommes en mesure d'identifier les problèmes sur lesquels on tombe le plus souvent, et de formuler des vœux quant au "système idéal" de suivi dont nous rêvons. Auparavant, essayons d'identifier les erreurs de conception et les limitations des algorithmes utilisés actuellement.

Tout d'abord, dans la grande majorité des cas, seul le paramètre "hauteur" est pris en compte. On devrait pouvoir extraire des indices significatifs d'autres paramètres de l'interprétation, par exemple l'enveloppe dynamique, le timbre. D'autre part, cette conception qui donne à la hauteur la première importance, bien que typiquement occidentale, ne correspond pas non plus aux demandes de nouvelles musiques qui accordent une place réévaluée au "bruit".

La partition est toujours codée de façon absolue. Or aucun des "paramètres" d'une partition musicale n'est absolu :

La hauteur : elle est relative par rapport à un diapason. La fréquence d'une note musicale est toujours en constante évolution, à cause du vibrato, des articulations entre les notes.

Le rythme : il est relatif à un tempo. Le swing, le rubato, essentiels à la musicalité, sont justement des écarts à la norme.

L'intensité : elle dépend très fortement du contexte, mais aussi de la hauteur, de la durée.
etc...

Lorsque l'on décide de coder une partition dans une machine, on peut se demander si l'on ne fait pas un contresens dès l'origine, du moins tant que l'on ne la code pas dans sa relativité. Cela suggère d'utiliser une approche statistique, probabiliste (un peu plus floue) pour traiter ce problème de suivi de partition.

La partition est codée de façon linéaire, séquentielle, mélodique. Ceci est dû au fait que, au départ, ce problème a été assimilé à un problème de *pattern matching*, tel qu'on peut l'appliquer par exemple à un texte (cf par exemple [Dannenberg 84]). Cette approche est critiquable dans son application à la musique et elle pose plusieurs problèmes conceptuels, pratiques et musicaux. En effet, les notes simultanées sont ramenées artificiellement sur un point de la ligne temporelle et pas détectées en tant qu'accords. Cette approche est valide, à la limite, pour une vision harmonique de la partition, mais pose problème avec la polyphonie : comment suivre plusieurs voix simultanées ? De plus dans certaines situations musicales, plusieurs cheminements sont possibles dans la partition.

Enfin, le suiveur "suit" toujours, c'est à dire qu'il est toujours derrière, alors que l'accompagnateur humain anticipe toujours.

6. Stratégies proposées

Nous pensons qu'il est important de considérer de manière globale le problème de la reconnaissance de partition à l'intérieur de laquelle peuvent être déclenchées diverses stratégies de suivi ou même de reconnaissance plus locales.

La première idée que nous souhaiterions valider est celle d'une reconnaissance de contour mélodique (hauteurs relatives). Au lieu de reconnaître une note précise, on cherche une hauteur à l'intérieur d'un ambitus. On peut ainsi accepter une certaine marge d'erreur

dans certaines circonstances. Pour cela, on va coder dans la partition (interface graphique) un intervalle variable autour de la note souhaitée, intervalle dans lequel toutes les notes seront "acceptées", considérées comme valides. Dans ce cas, par rapport aux stratégies utilisées jusqu'à présent, nous pensons que la tolérance horizontale (actuellement elle est généralement de deux fausses notes consécutives) doit être réduite, puisqu'elle est remplacée par une certaine forme de tolérance verticale. Ce point reste à vérifier. Nous évaluons également une interface de "neumes" où seules les directions mélodiques seraient codées, indépendamment de toute notion de hauteur absolue.

Il est évident qu'une reconnaissance de neumes ne saurait satisfaire à elle seule tous les cas de figure mais rendrait, dans un grand nombre de cas, de grands services. La règle principale demeure qu'il faut coder ce qui, dans une fenêtre d'analyse déterminée, peut être considéré comme un événement unique, et comme l'événement le plus invariant.

Une autre idée, assez simple mais assez efficace, serait d'introduire une notion de temps relatif qui déterminerait par exemple que deux événements seraient séparés par un temps minimum. Cela éviterait ce problème de "*score jumping*" qui apparaît fréquemment, dans lequel une fausse note émise par l'interprète se trouvant également codée dans les événements attendus fait avancer le programme de plusieurs pas car l'ordinateur a détecté quelque chose qui aurait dû n'intervenir que plus tard.

Nous pensons que l'on aurait beaucoup à gagner en utilisant simultanément de multiples stratégies, agissant concurremment. Parmi celles-ci, on envisage d'implémenter la reconnaissance d'enveloppes, de timbres (variations du spectre), d'événements temporels, ainsi que le suivi de tempo. Ces différentes stratégies ne sont pas toujours toutes utilisées, mais on choisit, en fonction de la situation actuelle laquelle (ou lesquelles) sont les plus adaptées. Actuellement nous nous penchons sur les problèmes suivants : comment faire communiquer tous ces différents agents de façon à rendre le résultat le plus efficace et le plus sûr possible, comment choisir (automatiquement ou manuellement) la meilleure stratégie en fonction du contexte. Nous recherchons également un moyen de représenter (interface graphique), les stratégies utilisées à tel endroit de la partition.

L'outil le plus urgent pour réaliser ce travail est d'abord l'interface graphique. Il faut pouvoir coder une partition à la fois :

- a) sous forme de hauteurs absolues (ce qui est le cas actuellement).
- b) sous forme de seuils d'ambitus (par exemple, un trait épais recouvrirait toute une bande de fréquence, comme lorsqu'on écrit des clusters pour le piano).
- c) sous forme de neumes, montants ou descendants, totalement indépendants de toute notion de hauteur réelle.
- d) sous forme de temps relatifs d'où pourrait naître une idée de groupement de valeurs, telles que les trilles, les appoggiatures, les valeurs longues, lorsque la nécessité s'en ferait sentir.
- e) sous forme de groupements formels tels que les accords, les arpèges, les figures, voir les modes de jeu (variations de timbres qui seraient à déterminer).

De plus, cette interface graphique doit offrir une représentation en notation musicale la plus correcte possible. Dans l'idéal, elle devrait utiliser les mêmes symboles, voire même être identique à la partition que lit le musicien.

En conclusion, on peut dire que la mise en œuvre d'un système de suivi de partition pour une pièce musicale nécessite actuellement un certain "savoir-faire", en effet, il faut toujours corriger, au cours des répétitions, la partition suivie, certains paramètres du système... Cela vient de ce que la principale différence entre un musicien humain et un

accompagnateur synthétique est que ce dernier est toujours en train de déchiffrer. Malgré que ce point soit déjà remarqué dans [VERCOE & PUCKETTE 85], il semblerait qu'aucun système de ce type n'ait été implémenté. Effectivement, un système qui serait capable, dans une certaine mesure, de corriger automatiquement les erreurs de détection qu'il commet au cours des répétitions serait très souhaitable.

7. Références

DANNENBERG, Roger (1984) *Tracking a Live Performer*. ICMC Proceedings, 1984.

VERCOE, Barry (1984) *The Synthetic Performer in the Context of Live Performance*. ICMC Proceedings, 1984.

VERCOE, Barry and PUCKETTE, Miller (1985) *Synthetic Rehearsal: Training the Synthetic Performer*. ICMC Proceedings, 1985.

BLOCH, Joshua J. and DANNENBERG, Roger B.(1985) *Real-Time Computer Accompaniment of Keyboard Performances*. ICMC Proceedings, 1985.

PUCKETTE, Miller and LIPPE, Cort *Score following in practice*, document interne IRCAM.

MANOURY, Philippe (1987) *Les partitions virtuelles*. Cahier d'analyse Ircam, 1996.

MANOURY, Philippe (1997) *Les partitions virtuelles*. A paraître.

STUCK, Leslie (1997) *Score Following Project - preliminary results*. document interne IRCAM, juin 1997.