Intégration des représentations temps/ fréquence et des représentations musicales symboliques

Peter Hanappe, Groupe Représentation Musicale, IRCAM, 1 Pl. Stravinsky, F-75004 Paris, hanappe@ircam.fr
Gérard Assayag, Groupe Représentation Musicale, IRCAM, assayag@ircam.fr

Résumé: Nous présentons un environnement d'analyse pour l'extraction de paramètres musicaux à partir d'un son en vue de leur utilisation dans un environnement de CAO. Nous proposons deux manières de décrire le son: la description en suite d'accords et la description par partiels. Nous ne proposons pas une analyse complètement automatique, par contre une grande importance est accordée aux aspects interactifs et graphiques de cet environnement. Nous avons installé un environnement qui est déjà utilisable dans des situations concrètes.

1.0 Introduction

Le travail proposé consiste en la création d'une passerelle entre le domaine du son et le domaine symbolique. En l'occurrence nous voulons établir une communication entre le logiciel AudioSculpt et le logiciel PatchWork. PatchWork est un logiciel écrit en CommonLisp qui manipule des structures musicales symbolique et qui est utilisé comme environnement pour la composition assistée par ordinateur (CAO). PatchWork possède une interface graphique de programmation visuelle et un ensemble d'éditeurs spécialisés dans la représentation et la manipulation de partitions musicales [Assayag & Rueda, 1993]. AudioSculpt est un logiciel d'analyse et de synthèse des sons. Il possède une interface graphique puissante qui sert à visualiser les analyses notamment sous la forme de sonagrammes [Eckel, 1992].

L'idée consiste à analyser des sons dans AudioSculpt pour en extraire une représentation symbolique qui sera envoyée vers PatchWork. Le problème est connexe du problème de la (re-)transcription automatique où l'on cherche à reconstituer une partition musicale équivalente à celle qui est supposée servir de base à l'enregistrement, ce qui implique d'extraire des paramètres musicaux tels que la fréquence fondamentale ou le rythme, et de séparer les sources. Cependant notre but n'est pas la reconstitution d'une partition existante. En effet, nous nous situons dans un cadre de production et de composition dans lequel on suppose que le musicien, partant d'un signal sonore (enregistrement, signal synthétisé) dont il désire se servir comme matériau générateur, va, à l'aide d'un ensemble de procédures combinant automatisme et interaction, en extraire des informations musicales qui seront pour lui "nécessaires et suffisantes" pour constituer une trame à la fois mélodique, harmonique, rythmique et dynamique, représentable dans le cadre de la notation instrumentale.

2.0 Transcription assistée par ordinateur

Plusieurs travaux traitant le problème de la transcription automatique ont été publiés. Un article pionnier dans ce domaine est l'article de Moorer [1977]. D'autres travaux dans ce domaine ont été effectués par Piszczalski et Galler [1977], Foster et al. [1982],

J.C. Brown [1992], G.J. Brown[1992], Doval [1994], Cerveau [1994] et Lepain [1995] pour n'en citer que quelques uns. Tous apportent des résultats promettants malgré le fait que les algorithmes souffrent tous de petites erreurs. Trois problèmes principaux: le besoin de données psycho-acoustiques, le fait qu'aucune technique d'analyse ne soit toutepuissante et la prise de décision dans une analyse automatique, sont rencontrés dans presque tous les travaux.

Il existe beaucoup techniques d'analyse, mais il n'existe pas de techniques qui soient fiables à cent pour cent ou qui soient adaptées à tous les sons. En plus, les analyses génèrent souvent une quantité d'information considérable ce qui nécessite un triage. Ce triage demande une prise de décision qui elle même implique des critères de choix. Il faut spécifier ces critères et les quantifier d'une manière ou une autre. Les valeurs cherchées sont souvent trouvées de manière heuristique en testant sur plusieurs (beaucoup de) sons et en gardant les valeurs optimales. Les valeurs données aux paramètres sont malheureusement peu souvent optimales pour tous les sons.

La manière de procéder que nous avons choisie dans ce travail a comme hypothèse de base que nous n'espérons pas fournir une analyse complètement automatique. Par contre nous proposons une analyse en plusieurs étapes. Après chaque étape l'utilisateur à accès aux résultats et peut les modifier. Ainsi il peut corriger ou supprimer des données qu'il envoie a l'étape suivante. Nous ne proposons donc pas une analyse automatique mais interactive. En plus nous avons mis peu d'effort sur l'optimisation des algorithmes d'analyses existants et les algorithmes que nous proposons sont très élémentairs. Nous avons travaillé davantage sur les applications et l'introduction de ces algorithmes dans un environnement prêt à utiliser. Nous avons profité des possibilités d'interactivité et de l'interface graphique pour créer un environnement d'analyse compréhensible et intuitif. Une bonne communication entre cet environnement et un environnement symbolique permet de convertir les analyses en structures musicales symbolique.

3.0 Méthodes d'analyse

Pour la description d'un son nous proposons deux techniques principales. Dans la première nous modélisons le son comme une suite d'accords. Chaque accord commence à la fin de l'accord précédente. Une première analyse détecte les débuts des accords, une deuxième analyse détermine les fréquences constituantes des accords. Nous appelons cette manière de décrire la représentation en suite d'accords. La deuxième technique fait une analyse plus détaillée du contenu fréquentiel dans le sonagramme. Nous allons chercher les partiels qui sont présents dans le son et nous attacher à en décrire le déploiement temporel. Nous appelons cette analyse le suivi de partiels. Les deux démarches partent d'une représentation intermédiaire du son:

3.1 Le sonagramme

La technique d'analyse à la base du travail est celle de la transformation de Fourier à court-terme. L'implémentation de cette technique a reçu le nom de Super Vocodeur de Phase [Moorer 1978]. Nous rappelons l'équation

S k q

(EQ 1)

avec

s(n): le n-ième échantillon du signal

w(n): la fenêtre

N: la taille de l'FFT

I: le pas d'avancement

q:l'indice du temps

k:l'indice des fréquences

Les fenêtres proposées sont celles de Hanning, Hamming ou de Blackman.

Pour l'affichage les spectres S(k, q) sont exprimés en dB. Le spectre équivalent en dB sera noté $S_{dB}(k, q)$. Le sonagramme est présentée à l'utilisateur sur l'écran de manière conventionnelle: horizontal, l'axe du temps; vertical, l'axe des fréquence, et en niveau de gris les amplitudes en dB. Plusieurs exemples sont montrés à la fin du texte. AudioSculpt est mené de différents outils pour faciliter l'inspection du sonagramme nottament le règlage de contraste dont nous allons nous servir dans les analyses suivantes. Le règlage de contrast permet de spécifier un seuil bas (en dB) et rends invisible tous les amplitude en dessous du seuil. Dans les analyses qui suivent nous reprenons le règlage mis par l'utilisateur et reduisons le spectre en concordance avec le seuil donné. C'est un exemple de "what you see is what you get": l'utilisateur analyse ce qu'il voit sur l'écran.

3.2 Détection des débuts

L'analyse de détection des débuts va essayer de trouver les endroits où il y a un changement d'énergie assez important dans le spectre. Nous calculons ces changements d'énergie par la différence des deux spectres:

(EQ 2)

Les différences sont ensuite intégrées comme suite

(EQ 3)

ce que nous appelons la dérivée. Nous gardons seulement les différences positives. Ceci nous est donné par le fait que:

•si une nouvelle note entre, nous estimons que l'énergie présente dans le spectre S(n) est plus grande que l'énergie dans le spectre S(n-1) ce qui résulte en une dérivée positive,

•si les deux spectres sont presque identiques comme dans le cas d'une note tenue, les différences sont toutes presque zéro et la dérivé est très petite,

•s'il y a autant d'énergie dans les deux spectres mais qu'elle s'est déplacée sur l'axe des fréquences, les différences sont à la fois positives et négatives et la dérivée sera positive mais moins prononcée.

Pour diminuer la présence des effets hasardeux et éviter trop de fluctuations dans le calcul de la dérivée nous calculons le spectre à l'instant q comme la moyenne des cinq spectres $S(k, q+j), j = 0, ..., 4$. Le moyennage nous permet de réduire les effets du bruit présent dans le signal et de diminuer la variance sur l'estimation du spectre. Pour rendre cette dérivée indépendante de l'énergie dans le signal à l'instant q, nous la divisons par cette énergie.

Les changement brusques sont indiqués par un maximum de la dérivée. La détection des maxima nous donne des indications sur les endroits où couper le signal. Cette fonction est (malheureusement) soumise aux fluctuations locales qui donnent lieu à des maxima sans importance. Nous pouvons éviter ce problème en plaçant un seuil sur les maxima. Pour éviter de fixer un seuil, nous proposons à l'utilisateur de donner un seuil dans un dialogue après avoir fait l'analyse. Selon le seuil indiqué nous ajoutons ou enlevons des marqueurs dynamiquement sur l'écran. L'utilisateur a maintenant la possibilité de déplacer, de supprimer ou ajouter des marqueurs.

3.3 Suite d'accords

Une fois les marqueurs placés l'utilisateur peut demander que lui soit fournie, pour chaque segment, une représentation comme accord. Nous estimons que le son entre deux marqueurs est stable et qu'il a un spectre constant. Le segment peut alors être représenté par un seul spectre. Pour ceci nous moyennons les spectres dans le sonagramme entre deux marqueurs

$$M(k) = \sum_{q = \text{startMarker}}^{\text{endMarker}} S(k, q) \qquad \text{(EQ 4)}$$

Puis le spectre moyenne est coupé selon le seuil placé par le réglage de contraste. Sur le spectre moyenne et coupé nous lançons une détection de pics.

Les résultats sont visualisés sur l'écran comme des traits horizontaux entre les marqueurs que l'utilisateur peut éditer. Les résultats peuvent finalement être envoyés vers PatchWork.

3.4 Suivi de partiels

La technique d'analyse dont nous avons parlée ci-dessous décrit le signal comme une suite d'accords. Cette description n'est pas toujours adaptée à tous les sons. Premièrement, si le son possède un caractère progressif et non "percussif" la détection des débuts échouera. Dans ce cas on peut toujours décider de placer les marqueurs à la main. Mais, deuxièmement, une description en suite d'accords n'est pas toujours la description optimale ou souhaitée. C'est le cas dans les sons "polyphoniques" où plu-

sieurs notes (tenues) sont superposées avec des débuts et fins différents. C'est pourquoi nous proposons une deuxième méthode pour analyser le son. Cette méthode va essayer de retrouver et de suivre les partiels dans le son. C'est l'analyse qu'on appelle le suivi de partiels. Bien des algorithmes sont connus qui accomplissent cette tâche. Dans le cas où nous nous mettons nous ne demandons pas une analyse très détaillée, c'est à dire une analyse qui retrouve chaque partiel du son (comme dans le cas de la synthèse additive). Nous nous intéressons aux partiels les plus pertinents, notamment les partiels qui restent visibles après le réglage de contraste.

Le suivi de partiels se fait ainsi: chaque spectre du sonagramme est d'abord réduit selon le seuil du réglage de contraste. Puis nous détectons les maxima. Les fréquences des maxima sont calculées plus précisément avec une interpolation polynomiale sur trois points. Nous gardons seulement les fréquences et les amplitudes des maxima.

Soient F(n,q), n = 1, ..., N, les fréquences des maxima détectés dans le spectre à l'instant q, et F(m,q - 1), m = 1 , ... , M, les fréquences des maxima dans le spectre à l'instant q - 1. Nous définissons une mesure de distance

$$\text{Dist}\,(F1, F2) \;=\; \frac{|F2 - F1|}{F1} \qquad \text{(EQ 5)}$$

et une distance maximale DMAX (exprimée en intervalle musical). Ainsi nous ne permettons pas de grands sauts en fréquence.

La connexion entre les pics du spectre S(q-1) avec les pics du spectre S(q) est faite comme décrit ci-dessous. Pour chaque maximum F(m,q - 1):

1. on détermine la liste de tous les maxima de S(q) a une distance plus petite que DMAX de F(m,q - 1), L = { F(n,q) : -DMAX < Dist(F(n,q), F(m, q-1)) < DMAX }
2. on parcour la liste L et si pour un maxima F(n,q)¸ L, il existe un maximum F(l,q-1)„ F(m,q-1) de S(q-1) qui est plus proche que F(m,q-1), c.à.d. Dist(F(n,q), F(l, q-1)) < Dist(F(n,q), F(m, q-1)), on supprime F(n,q) de la liste L
3. si la liste L est vide, il n'y aura pas de connection faite et F(m,q-1) est la dernier points d'un partiel. Si la liste L contient plusieurs éléments, F(m,q-1) est connecté au maximum F(n,q)¸ L pour lequel la distance Dist(F(n,q), F(m, q-1)) est minimale.
4. les maxima F(n,q) qui n'ont pas été connectés commencent un nouveau partiel.

Le test dans le deuxième pas de l'analyse nous rassure qu'on ne connecte pas le maximum F(m,q-1) a un maximum F(n,q) pour qui il existe un meilleur choix. Cette méthode simple nous donne des résultats satisfaisants.

Les partiels trouvés sont affichés sur l'écran comme des points liés entre eux. De nouveau l'utilisateur peut éditer les résultats (supprimer, lier, couper des partiels et éditer les points) et les résultats peuvent être envoyés vers PatchWork.

3.5 Du côté de chez PatchWork

Dans PatchWork on recupère les données de l'analyse faite dans AudioSculpt. Des résultats de l'analyse suite d'accords et suivi de partiels, nous gardons le début, la

durée, la fréquence moyenne et l'amplitude moyenne de chaque partiel. La liste des temps de débuts ou également la liste des dates des marqueurs peuvent être introduite dans la module de quantification KANT pour extraire le rythme [Agon et al., 1994].

Avant d'envoyer les résultats bruts temps-fréquence-amplitude vers un éditeurs de séquence d'accords ChordSeq un triage est possible. Ainsi on peut spécifier une fréquence basse et une fréquence haute en dessous et au dessus laquelle les fréquences ne sont pas prises en compte. Egalement on spécifie un interval de temps qui sert à regrouper dans un même accords les notes qui sont très proche dans le temps. Dernièrement une projection des amplitudes sur les vélocités MIDI est faite.

Finalement on peut combiner le rythme trouvé avec les données fréquentielles dans l'éditeur RTM.

Nous terminons cette section avec un schéma des représentation possible dans l'environnement actuel:

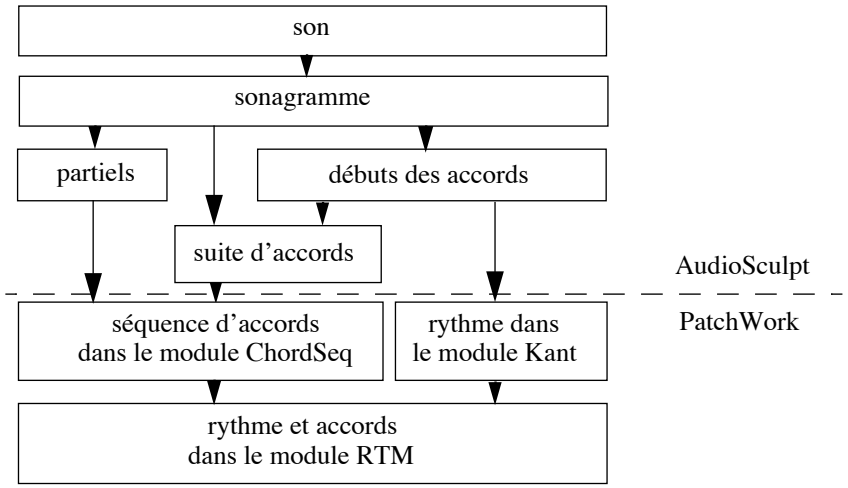


FIGURE 1.

Le schéma complet des représentations.

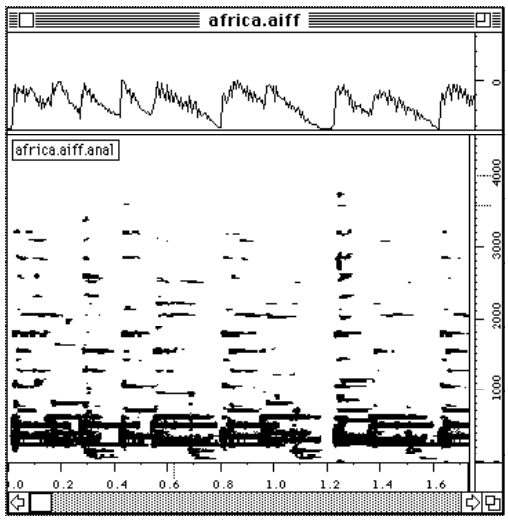
4.0 Exemples d'application

A fin d'illustrer le travail présenté, nous présentons deux exemples d'applications. Le premier montre l'utilisation de la description en suite d'accords, le deuxième l'utilisation du suivi de partiels.

4.1 Harpe africaine : suite d'accords

Dans le premier exemple, il s'agit d'un enregistrement d'une harpe africaine Nzakara enregistrée en Centrafrique par Marc Chemillier. La harpe est inharmonique, jouée avec un tempo élevé, d'une manière polyphonique (en général, deux cordes pincées en même temps) et dans une harmonie non tempérée.

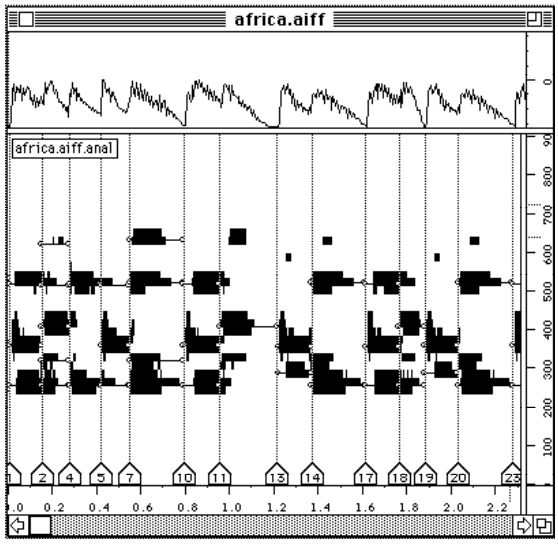
FIGURE 2.



Le sonagramme de l'enregistrement d'une harpe africaine Nzakara. En haut de la fenêtre est affiché la courbe de puissance du son, en bas le sonagramme. L'axe horizontal montre le temps en secondes, l'axe vertical du sonagramme montre la fréquence en Hertz.

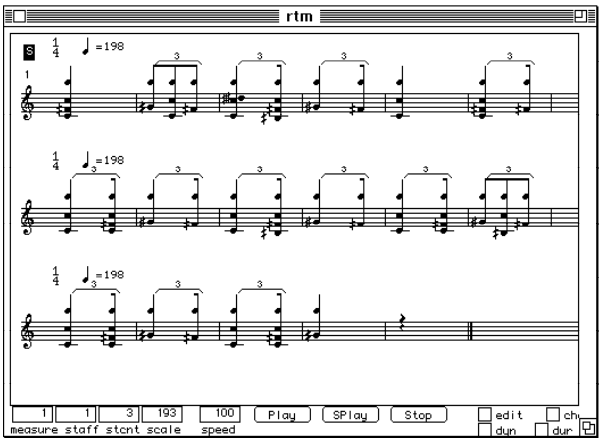
Figure 2 montre une partie du sonagramme. Le détection des débuts trouve la majorité des attaques. Après inspection des résultats nous règlons le contraste pourqu'il ne reste visible que les partiels les plus forts. Figure 3 montre une partie du sonagramme avec les marqueurs et les fréquences détectées après analyse.

FIGURE 3.



Le sonagramme avec les marqueurs et les fréquences détectées.

Finalement ces données sont introduites dans AudioSculpt. On retrouve assez bien ce qui pourrait être la partition de l'enregistrement. Si on veut écouter les résultats sur le synthétiseur MIDI il est important de les écouter avec une précision d'un quart de ton ou même plus précis car il s'agit d'une musique non tempérée. Les résultats passent bien, mais la réduction qu'on fait sur la dimension du temps quand on coupe le signal en segment atténue la richesse du jeu rythmique.

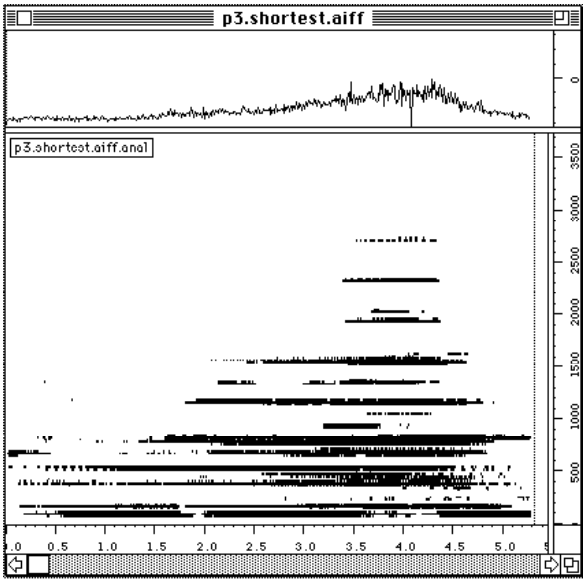
FIGURE 4. 

Les résultats finals dans PatchWork

4.2 Accord Complexe: suivi de partiels

Dans le dernier exemple nous traitons un son qui a été réalisé par le compositeur Joshua Fineberg. Il s'agit d'un mélange subtil de sons instrumentaux et synthétisés. Il y a une évolution progressive: le son est plutôt bruité au début, puis une flûte entre suivi d'autres instruments qui servent à colorer et enrichir le timbre. Dans la figure 5 on voit le sonagramme. Il s'agit d'un exemple difficile à cause de cette évolution temporelle et à cause de son timbre fragile.

Nous traitons ce son avec le suivi de partiels.

FIGURE 5. 

Le sonagramme du deuxième exemple.

Dans figure 6 est montré les résultats après le suivi de partiels. Nous avons masqué le sonagramme avec le règlage de contrast. Les points individuels des partiels ne sont pas visible à cause d'une résolution trop réduite de l'affichage.

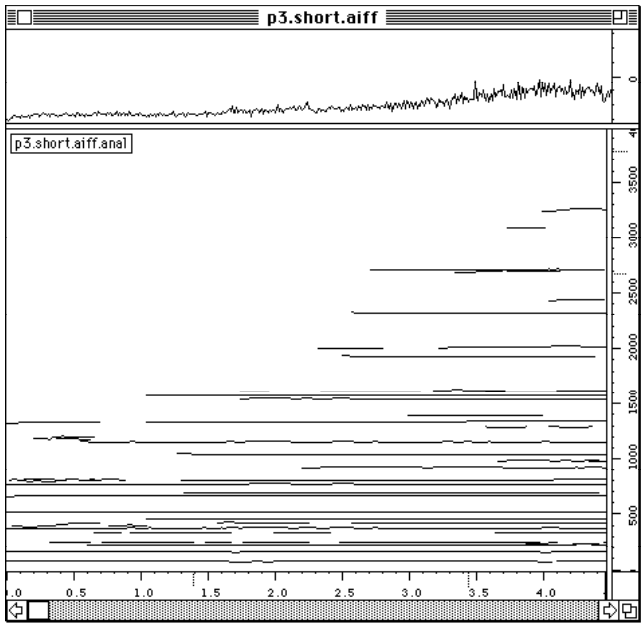


FIGURE 6. Le sonagramme après le suivi de partiels.

Figure 7 montre les résultats dans l'éditeur ChordSeq dans PatchWork en nottation proportionnelle. Les traits horizontaux indiquent la durée des notes.La simulation sur clavier électronique (en quart de ton et avec des oscilateurs sinusoidaux) est plus qu'acceptable.

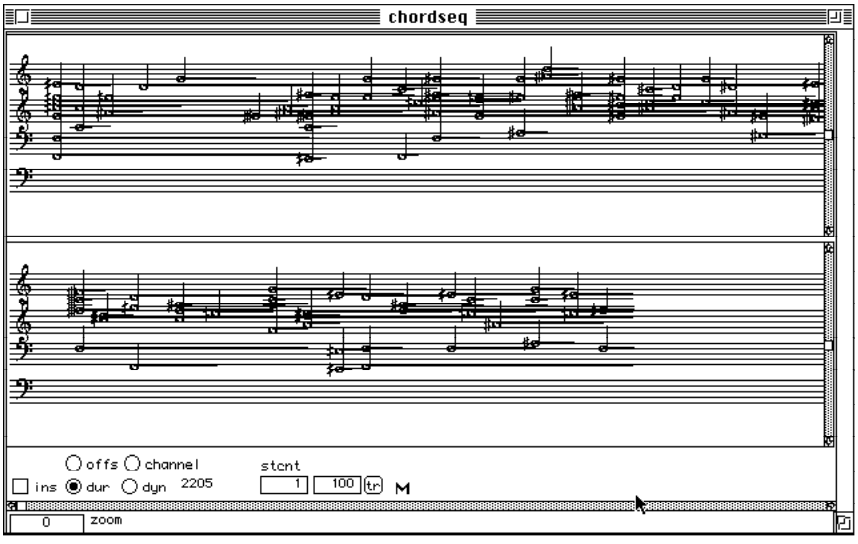


FIGURE 7. Les résultats dans PatchWork en notation proportionnelle.

5.0 Conclusion

Nous avons présenté un environnement d'analyse pour l'extraction de paramètres musicaux à partir d'un son en vue de leur utilisation dans un environnement de CAO. Nous avons proposé deux manières de décrire le son: la description en suite d'accords et la description en partiels.

Dans la première le son est modélisé comme une suite d'accords distinctes. Les débuts des accords sont détectés puis les fréquences réprésentantes des accords sont extraites. Dans la description en partiels une analyse suivi de partiels trace les partiels les plus saillants. Les résultats sont ensuite introduits dans PatchWork où ils sont traduits et affichés en notation traditionnelle. Nous ne proposons pas une analyse complètement automatique, par contre une grande importance a été accordée aux les aspects interactifs et graphiques de cet environnement. L'interactivité de ce travail consiste en la possibilité de changer et de supprimer des résultats intermédiaire. L'interface graphique facilite la visualisation et la manipulation de ces résultats intermédiaires. Nous espérons avoir montré l'intérêt de cette manière de travailler. Nous avons installé un environnement qui est déjà utilisable et qui peut devenir un outil intéressant dans les situations de CAO et musicologie.

6.0 Références

Agon C., Assayag A., Fineberg J., Rueda C. (1994) "Kant: a Critique of Pure Quantification." Proceedings of the ICMC 94, Aarhus, 1994.

Assayag G. & Rueda C., (1993), "The music representation project at Ircam", Proceedings of the Int. Computer Music Conference, 1993, Tokyo, Japan.

Brown, G.J., (1992), "Computational auditory scene analysis: a representational approach", TR CS-92-22, University of Sheffield: Department of Computing Science

Brown, J.C., (1992), "Musical fundamental frequency tracking using a pattern recognition method", J. Acoust. Soc. Am., 92(3), p1394-1402, September 1992, Acoust. Soc. of America

Cerveau, L., (1994), "Segmentation de phrases musicales à partir de la fréquence fondamentale", Mémoire de DEA ATIAM, Paris VI

Doval., B., (1994), "Estimation de la fréquence fondamentale des ignaux sonores", Thèse de doctorat, Université de Paris VI

Eckel, G, (1992), "Manipulation of sound signals based on graphical representation: A musical point of view", Proceedings of the 1992 Int. Workshop on Models and Representations of Musical Signals, Capri, 5-7 October, 1992

Foster, Schloss & Rockmore, (1982), "Toward an intelligent editor of digital audio: signal processing methods", Computer Music Journal, 6(1), p42-51, Spring 1982, Mass. Inst. of Techn.

Lepain Ph. & André-Obrecht R., (1995), "Micro-segmentation d'enregistrements musicaux", Deuxièmes Journées d'Informatique Musicale JIM, LAFORIA, Inst. Blaise Pascal, Avril 7-8, 1995, Paris

Moorer, J., (1977), "On the transciption of musical sound by computer", Computer Music Journal, November 1977, p32-38, Mass. Inst. of Techn.

Moorer, J., (1978), "The use of the phase vocoder in computer music applications", Journal of The Audio Engineering Society, 26(1/2): 42-45.

Piszczalski & Galler, (1977), "Automatic music transcription", Computer Music Journal, November 1977, p24-31, Mass. Inst. of Techn.