

Interface de navigation musicale par le contenu

Geoffroy Peeters

IRCAM A/S, CNRS-STMS
1, pl. Igor Stavinsky
75004, Paris, France
peeters@ircam.fr

David Fenech

IRCAM A/S, CNRS-STMS
1, pl. Igor Stavinsky
75004, Paris, France
fenech@ircam.fr

Xavier Rodet

IRCAM A/S, CNRS-STMS
1, pl. Igor Stavinsky
75004, Paris, France
rod@ircam.fr

RESUME

Dans cet article nous décrivons un nouvel outil pour la navigation intra-document d'un morceau de musique. Cet outil se présente sous la forme d'un lecteur multimédia représentant le contenu d'un morceau de musique sous forme visuelle. Chaque type de contenu décrit (structure, accord, mesure/ temps, notes, instruments) est associé à une représentation visuelle distincte. L'utilisateur voit ainsi ce qu'il écoute. Il peut également se déplacer dans son écoute en fonction du contenu visuel. Pour cela, chaque objet visuel a un comportement distinct soit sous forme de retour sonore, soit sous forme de déplacement dans l'écoute. Ces informations de contenu peuvent être extraites automatiquement (à l'aide d'algorithmes de traitement du signal) ou entrées par l'utilisateur directement à partir de l'interface. Dans ce cas le lecteur officie comme un logiciel d'annotation guidé par le contenu.

MOTS CLES : Interface de navigation multimédia, description du contenu musical

ABSTRACT

In this paper, we present a new tool for intra-document browsing of musical pieces. This tool is a multimedia player which represents the content of a musical piece visually. Each type of musical content (structure, chords, measures/ beats, notes, events) is associated which a distinct visual representation. The user sees what is listening too. He can also browse inside the music according to the visual content. For this, each type of visual object has a dedicated feedback, either as an audio-feedback or as a pointer feedback. Content information can be extracted automatically (using signal processing algorithms) or annotated by hand by the user. In this case the multimedia player is used as an annotator tools guided by the content.

CATEGORIES AND SUBJECT DESCRIPTORS: H.4 [Information Systems Applications]: Miscellaneous;

GENERAL TERMS: Algorithms, Design, Documentation

KEYWORDS: Graphical interface, multimedia player, music content description.

INTRODUCTION

L'extraction automatique de la description du contenu d'un morceau de musique (content-based description) a fait l'objet d'un nombre important de recherches ces dernières années. Des applications telles que la recherche par caractéristiques de contenu (mélodie, genre, humeur ou type de chanteur) dans de très grandes bases de données musicales ou la recherche par similarité (de mélodie, de timbre) sont maintenant rendues possibles. Cependant, peu d'applications utilisent l'information de contenu afin de guider l'utilisateur dans son écoute, de lui permettre de pénétrer le contenu musical d'un morceau, voir de lui permettre l'étude comparative de plusieurs œuvres par visualisation de leur contenu.

Dans le projet Européen Semantic HIFI, un premier outil de navigation intra-document par le contenu a été développé. Celui-ci reposait sur le paradigme de navigation guidée par la visualisation de la structure sous-jacente à un morceau (représentation des répétitions des différentes parties telles que couplet, refrain, ...) [1] [2]. Cette interface, développée en Flash, donc multiplateforme, a été utilisée pour la navigation à l'intérieur d'un morceau dans un prototype de chaîne HIFI. L'utilisateur, à partir d'un PDA servant de télécommande à la chaîne HIFI, se déplaçait dans son écoute par glissement d'un stylet sur la représentation visuelle de la structure du morceau (voir Figure 1).

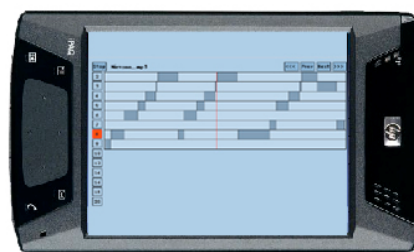


Figure 1 : PDA servant de télécommande à la Semantic-HIFI : représentation de la structure temporelle

Dans le cadre du projet Music Discover, ce concept de lecteur permettant la navigation intra-document par le contenu a été étendu à l'ensemble des informations de

contenu. Ces informations représentent le contenu à travers des échelles temporelles décroissantes : structure temporelle du morceau (succession temporelle des différentes parties telle que couplet, refrain, mouvements, ...), matrice de similarité (représentation de la similarité de chaque paire de temps d'un morceau sous forme de matrice), progression temporelle d'accords du morceau, grille temporelle musicale du morceau (la position des différentes mesures et des temps), différentes notes jouées à chaque instant ainsi que la présence des différents instruments ou sons percussifs à chaque instant.

DESCRIPTION DE L'INTERFACE

Le lecteur (voir Figure 2) se présente sous la forme d'une interface graphique regroupant dans la partie haute les outils de chargement de fichier ainsi que les informations relatives aux méta-données éditoriales (affichage du contenu des tags ID3). Le panneau du milieu représente le contenu du morceau sous forme visuelle, le temps du morceau y est représenté horizontalement (début du morceau= partie gauche, fin= partie droite), la position actuelle de lecture est représentée par une bar rouge (playhead) se déplaçant de gauche à droite. Dans ce panneau sont superposés tous les objets visuels représentant le contenu temporel du morceau. La partie inférieure de l'interface représente la ligne temporelle (Time line) du morceau. Elle permet de « zoomer » dans le morceau. Dans ce cas, la « Time line » indique par un rectangle grisé la partie actuellement affichée dans le panneau du milieu (Figure 3).

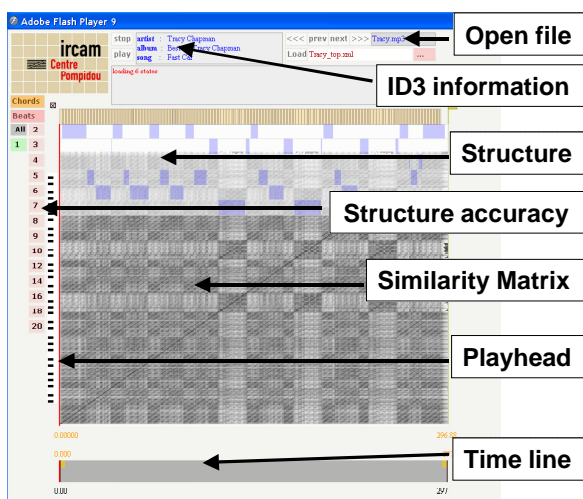


Figure 2 : Interface principale de navigation intra-document.

L'outil permet la lecture directe des fichiers audio au format mp3. Les descriptions de contenu du morceau sont stockées dans un ensemble de fichier xml. Ceux-ci sont organisés de manière hiérarchique : un fichier xml « top » pointe vers le fichier audio mp3 ainsi que vers l'ensemble des fichiers xml de description de contenu (un fichier par type de description). Ces descriptions peuvent

- provenir d'une annotation manuelle
- être générées automatiquement à l'aide d'un ensemble d'outil d'extraction automatique.

L'interface n'affiche que les données existantes, elles ne sont donc pas toutes requises.

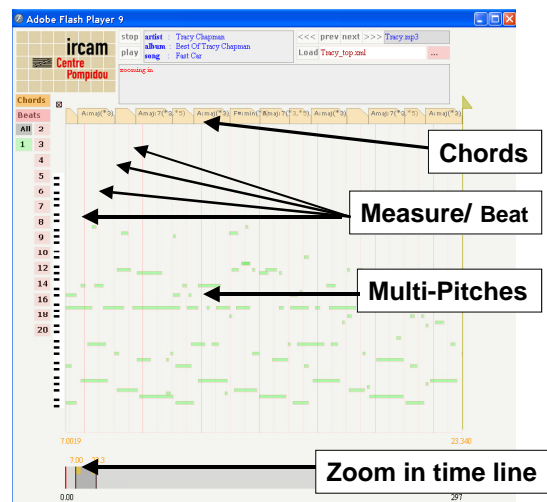


Figure 3 : « Zoom in » dans l'interface principale de navigation intra-document.

Le paradigme principal du lecteur est « cliquez et écoutez ce sur quoi vous cliquez ». Chaque objet visuel a pour cela un comportement spécifique, soit sous forme de retour sonore, soit sous forme de déplacement dans l'écoute. Les objets peuvent être superposés graphiquement par utilisation d'un système de masque et de transparence des masques (voir Figure 2 et la structure transparente par dessus la matrice de similarité).

DIFFERENTS TYPES DE DESCRIPTION

Nous décrivons maintenant l'ensemble des informations affichées, leur signification, leur représentation graphique ainsi que la manière dont elles peuvent être extraite automatiquement.

Matrice de similarité

La « matrice de similarité » décrit sous forme d'une matrice la similarité entre les différents instants d'un morceau. Un point (x,y) de la matrice représente la similarité entre le contenu à un temps x et à un temps y. Si deux temps ont un contenu très similaire, la valeur dans la matrice sera très élevée. Cette matrice est représentée visuellement comme une image 2D, les valeurs de similarité faible sont en gris clairs et celles de similarité élevées en gris foncés. Une diagonale foncée dans la matrice représente donc la répétition d'un ensemble de temps successifs x, x', x'' à un autre endroit du morceau y, y', y'' (une mélodie se déroulant sur les temps x, x', x'' est répétée sur les temps y, y', y''). L'utilisateur peut cliquer n'importe dans la matrice, la lecture commencera alors instantanément à l'endroit cliqué. L'utilisateur peut ainsi comparer plusieurs occurrences d'une mélo-

die en cliquant sur les diagonales. Cette matrice est générée automatiquement à partir de l'analyse audio du contenu du fichier et représente les similarités de timbre (instrumentation) et d'harmonie (mélodie, accord) [3].

Structure temporelle

La « structure temporelle » représente un morceau de musique comme une succession de parties (couplets, refrains, mouvements ...) pouvant être répétées au cours du temps. La structure est représentée visuellement comme un ensemble de rectangle posés sur différentes lignes. Chacune des lignes représente un type de partie différent. Les rectangles posés sur une même ligne représentent donc une partie similaire (ligne des couplets, ligne des refrains ...). La finesse de la décomposition en structure d'un morceau peut être choisie par l'utilisateur (partie gauche de l'interface « structure accuracy »). En cliquant n'importe où dans un rectangle, la lecture commence instantanément au début de cette partie. L'utilisateur peut ainsi rapidement comparer différentes parties ou différentes occurrences d'une même partie. Des possibilités de forward/ backward par partie sont également incluses. La structure peut être données manuellement ou estimée automatiquement à partir de l'analyse du signal audio[4] [3].

Progression d'accords

La « progression d'accords » représente la position et la durée de chaque accord composant un morceau de musique. Les accords sont représentés visuellement sous forme d'onglets dans la partie supérieure de l'interface. Sur chaque onglet est noté le nom de l'accord (GM, DM, ...). Lorsque l'utilisateur clique sur un de ces onglets, la lecture commence instantanément au début de l'accord correspondant. L'utilisateur peut ainsi se déplacer dans le morceau par accord. Un autre comportement actuellement envisagé est le retour sonore. Cliquer sur un onglet provoquerait la lecture de l'accord prototypique correspondant. La progression d'accord peut être donnée manuellement ou estimée automatiquement à partir de l'analyse du signal audio [5].

Position des mesures et des temps

La position des mesures (premier temps) et des temps (battements) représente la grille temporelle musicale d'un morceau. Celle-ci est indispensable pour comprendre le rôle de chaque événement présent dans le morceau. Dès lors ces informations sont représentés sous forme de ligne verticale traversant l'ensemble des autres informations (ligne épaisse pour les premiers temps, fines pour les autres temps). L'utilisateur peut enclencher un « métronome » : chaque fois que la bar de lecture (playhead) croise un marqueur de temps un bip est émit. Ces informations peuvent être données manuellement (pose de marqueurs « à la volée » pendant

l'écoute) ou estimées automatiquement à partir de l'analyse du signal audio [6][7].

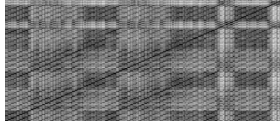




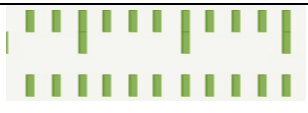
Notes de musique

Les différentes notes présentes dans le morceau sont représentées graphiquement sous forme de piano-roll (chaque ligne représente une note du piano, l'existence d'une note à un instant donné est représentée par un rectangle à ce temps et à la hauteur de la note). Lorsque l'utilisateur clique sur un rectangle, une note de piano à la hauteur de note correspondante est immédiatement émise. Ceci permet par exemple d'écouter la transcription affichée en comparaison de la musique originale. L'interface peut représenter graphiquement jusqu'à 16 canaux de notes. Les notes de chaque canal sont alors représentée par une couleur différentes. L'utilisateur peut choisir quels canaux sont affichés. Ces informations peuvent être estimées automatiquement à partir de l'analyse du signal audio [8].

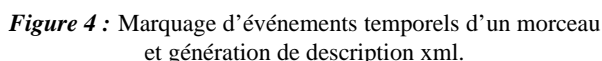
Événements

Les différents événements sonores (percussions ou instruments de musique) se produisant au cours du temps sont représentés par un event-roll (une classe d'événement par ligne). L'interface peut représenter les occurrences d'instruments de musique classique ainsi que ceux de percussions. Ces informations peuvent être estimées automatiquement à partir du signal audio [9], [10].

Les différentes représentations graphiques sont résumées dans le tableau suivant.

Matrice de similarité	
Structure temporelle	
Progression d'accords	
Position des mesures et des temps	
Notes de musique	
Événements	

Une extension du lecteur de contenu a été réalisée afin de permettre l'annotation de l'écoute ainsi que la création (voir la correction) de la description de contenu elle-même ; ceci à partir de la même interface d'écoute. L'annotation du contenu d'un morceau de musique est en effet grandement facilitée par la connaissance des autres informations de contenu. Par exemple annoter la structure temporelle d'un morceau est grandement facilitée par la connaissance (visualisation) de la matrice de similarité, annoter les accords par la connaissance de la position des temps, L'annotation est faite par la génération de marqueurs. Ces marqueurs peuvent être apposés off-line avec la souris, posés « à la volée » pendant l'écoute (le marqueur se place à l'endroit actuel de lecture) ou générés automatiquement à partir des descriptions de contenu (chaque description de contenu génère un marqueur correspondant). Les marqueurs peuvent être annotés textuellement, déplacés, effacés, ... Le fichier xml correspondant à l'annotation peut ensuite être généré automatiquement (voir Figure 4).



Dans cet article nous avons présenté une nouvelle interface graphique permettant la navigation intradocument de morceau de musique guidée par le contenu. Chaque objet musical considéré a une représentation visuelle propre, un comportement distinct, une représentation sous forme de fichier distinct et un outil d'extraction automatique associé.

Parmi les outils d'extraction automatique, celui d'extraction de structure musicale revêt un intérêt particulier puisqu'il permet de retrouver (d'estimer) la structure musicale sous-jacente à la création d'un morceau. Outre son utilisation pour la représentation visuelle de contenu, il peut être utilisé afin de résumer de manière sonore le contenu musical d'un morceau. Le

A une échelle plus grande, l'analyse de la structure temporelle du contenu d'un fichier audio, permet le développement d'applications de segmentation de flux audio, telle la segmentation automatique de flux radio ou de bandes sonores de film en catégories parole, musique, parole sur musique de fond, ...

Cette recherche a été en partie financée par le projet ACI Masse de données « Music Discover ».

1. Peeters, G, Indexation et accès au contenu musical, in Les Nouveaux Dossiers de l'Audiovisuel. 2005.
2. Boutard, G, S Goldszmidt, and G Peeters. Browsing inside a Music Track, the Experimentation Case Study. in Workshop on Learning the Semantics of Audio Signals (LSAS). 2006. Athens, Greece.
3. Peeters, G. Sequence representation of music structure using higher-order similarity matrix and maximum-likelihood approach. in Proc. of ISMIR. 2007. Vienna, Austria.
4. Peeters, G, A Laburthe, and X Rodet. Toward Automatic Music Audio Summary Generation from Signal Analysis. in Proc. of ISMIR. 2002. Paris, France.
5. Papadopoulos, H and G Peeters. Large-Scale Study of Chord Estimation Algorithms based on Chroma Representation. in Proc. of CBMI. 2007. Bordeaux, France.
6. Peeters, G, Template-based estimation of time-varying tempo. EURASIP Journal on Advances in Signal Processing, 2007. **2007**(Special Issue on Music Information Retrieval Based on Signal Processing): p. Article ID 67215, 14 pages.
7. Papadopoulos, H and G Peeters. Simultaneous estimation of chord progression and downbeat from audio file. in submitted to IEEE ICASSP. 2008. Las Vegas, USA.
8. Yeh, C. Multiple F0 estimation for MIREX 2007. in ISMIR. 2007. Vienna, Austria.
9. Livshin, A and X Rodet. Musical instrument identification in continuous recordings. in Proc. of DAFX. 2004. Naples, Italy.
10. Gillet, O and G Richard. Supervised and unsupervised sequence modelling for drum transcription. Proceedings of the 8th International Conference on Music Information Retrieval. in Proc. of ISMIR. 2007. Vienna, Austria.