Music Identification by Leadsheets

Frank Seifert
Department of Computer Science
University of Technology
Chemnitz, Germany
Frank.Seifert@informatik.tu-chemnitz.de

Wolfgang Benn
Department of Computer Science
University of Technology
Chemnitz, Germany
Wolfgang.Benn@informatik.tu-chemnitz.de

Abstract

Most experimental research on content-based automatic recognition and identification of musical documents is founded on statistical distribution of timbre or simple retrieval mechanisms like comparison of melodic segments. Therefore often a vast number of relevant and irrelevant hits including multiple appearances of the same documents are returned or the actual document can't be revealed at all. To improve this situation we propose a model for recognition of music that enables identification and comparison of musical documents without dependence on their actual instantiation. The resulting structures enclose musical meaning and can be used for estimation of identity and semantic relationship between musical documents.

1 Introduction

Estimation of similarity or a certain relationship among musical pieces concentrates so far only on categorization and distribution of sounds (Haitsma & Kalker 2002; Yang 2002). Unfortunately there is no general definition for similarity of musical documents. Similarity depends not only on actual application but also on individual capabilities of listeners. But there is a universal human capability that we want to implement: Identification of same musical pieces without consideration of their actual sounding. Hence, we try at first to ignore timbre and focus on symbolic parameters of music and their structural relationships.

Finally we propose a method to create a semantic order on musical documents that is based on automatic recognition of musical pieces and assesses the degree of relationship by a content-based estimation of distances.

2 The leadsheet-template

Most existing symbolic music processing systems that handle identification tasks underlie the query-by-humming-paradigm and are limited to short sequences without meaning (i.e. Dovey 2001; Lemström et al. 2001). Our approach relies on those techniques but we extend them substantially. Most importantly, we evaluate structural relationships between musical segments. So far the notion of melody has been discussed very undifferentiated in most cases. In order to express musical meaning we decompose a melody in smallest meaningful musical entities that we call characteristic motifs.

Motif and melody represent the most essential components of a composition. Structures that are built up by them form components of a higher order. The top of this hierarchy represents a piece of music.

But what is the foundation of this analytical decomposition process?

Our aim is to provide a maximum generic representation for music in order to derive semantic equivalence. The most commonly used means of conserved instructions for reproduction of music that fulfills generic music creation is the leadsheet. A leadsheet contains merely the melody with accompanying chord symbols that correspond to harmony:

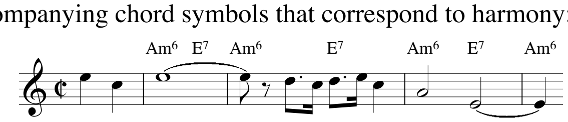

Figure 1: Leadsheet-excerpt of Gershwin's "Summertime"

Using leadsheets musicians create their voices themselves "live" during a performance. This process is aided by harmonic information of the leadsheet and their inherent musical knowledge.

We define a leadsheet-template to represent musical parameters and their hierarchy: Primary parameter is the contour of a characteristic motif and its configuration of temporal spacing between the tones.

Since most pieces contain repetitive structures, either in small or large scale, we can build up a hierarchy over the enumeration of characteristic motifs. So we define a leadsheet-template (LST) to be a

- non-empty sequence of leadsheet-templates or a non-empty sequence of characteristic motifs and their harmonic configuration.

Harmonic configuration is considered as structuring aspect since the meaning of successive notes can change when harmonic context is altered (Motte-Haber, 1985).

3 A model of music recognition

Although leadsheets are more convenient for popular music it's possible to code nearly all tonal repertoire into this form of representation. But this process of reduction has to care for an adequate representation of perceptual relevancy, for instance polyphonic voices. This idea – to generalize the representational capability of tonal music by perceptual relevant leadsheets – leads to the following hypothesis of music recognition:

- Artificial recognition and processing of music can be realized by mapping music to LSTs that are constituted by perceptual relevant dimensions of music under the rules of tonality.

This form of processing could possibly also be assumed during human cognitive activities in tonal cultures. However, most listeners can't explicitly reference their implicit knowledge.

With the existence of LSTs we've got structures to that either known or unknown musical pieces can be mapped onto. We formalize this recognition process: Let us define CM to be a set of all characteristic motifs. CM^* represents a set of arbitrary repetitions of any $cm \in CM$ with

$$CM^* = \left\{ cm^n \,\middle|\, \exists cm \in CM, n \in N \right.$$

Analogously we define LS to be a set of all LSTs and LS^* represents a set of arbitrary repetitions of any $ls \in LS \square ls \square LS$. Additionally we define $C = CM^* \cup LS^*$ to be the union of all arbitrary repetitions of LSTs and characteristic motifs and M to be a set of musical pieces.

Then we can denote the recognition of a musical piece $m \in M$ as its mapping process $\square$ onto all combinations of elementary and structured LSTs and characteristic motifs, say their power set:

$$\square : m \rightarrow 2^C \tag{3.1}$$

This mapping process can be subdivided in two main procedures: First the musical parameters have to be mapped onto elementary characteristic motifs resulting in a set of temporal ordered motif instances if successful. Then, all detected instances have to be associated with maximized surrounding LSTs. Again, this results in a set of temporal ordered instances, but instances of LSTs this time.

Unfortunately it's not possible to consider (3.1) as a function. Ambivalences could happen during the mapping-processes onto characteristic motifs. Consequently several valid motifs could be assigned to a musical segment simultaneously. If several candidates compete for qualification as member of a LST-instance, usually the candidate with the most similar combination of parameters will be selected. Additionally this implicit deduction-process results in an ongoing qualification of competing instances based on context.

Actually the main problem is an incomplete mapping-process onto LSTs caused by a fizzled identification of musical parameters of characteristic motifs or a missing structural description for identified sequences. By explicit deduction we'll try to get the mapping-process successful. Therefore we aspire a formation of hypotheses, which leadsheet and consequently which resulting musical parameters are possible for an unknown segment at all. Basis for the formation of hypotheses is the configuration of adjacent LSTs. Then we have to validate the determined hypotheses to complete the mapping-process. Now we try to map the unknown tone-segment onto the characteristic motifs that constitute the hypothesized LSTs. The degree of deviation leads to preference and selection of a hypothesis or to its refusal if there isn't enough evidence.

4 Conclusions & further work

We proposed a model for the recognition of music that enables identification and comparison of musical documents without dependence on their actual instantiation. The evaluation of perceptual relevant parameters is the basis of this functionality. Furthermore, a deductive assessment-method for musical phrases came into existence.

Music recognition has been defined as mapping-process from musical documents onto leadsheet-templates that function as hypotheses. Due to the top-down modeling, which is adequate for music perception and the resulting existence of hypotheses, distance measures for incomplete mapping-processes of contradictory segments can be calculated and the most likely identity can be concluded.

A flexible treatment of structure and variation of musical parameters are special features of the introduced model. So we can recognize and combine musical forms, deduce incomplete structures and we've enabled the coexistence of several structural levels.

A first implementation of the presented ideas confirmed their principal feasibility. However it should be seen as proof of concept. Nearly every detail needs improvements, for instance the estimation of distances. Conceptually, the recent approach should be extended to subsymbolic music data. Due to its hypothesis-oriented design the proposed model is very suitable to guide audio processing tasks.

For the first time the proposed concepts started to enable a complete comparison of whole musical documents. Not only naïve in the sense of a simple test of identity or evaluation of statistical features but also under integration of semantic aspects. Consequently they form a framework for identification, navigation and automatic linking of musical documents.

References

Dovey, M. J. (2001). A technique for "regular expression" style searching in polyphonic music. ISMIR 2001. Bloomington, Indiana.

Haitsma, J. & Kalker, T. (2002). A Highly Robust Audio Fingerprinting System. ISMIR 2002.

Lemström, K., Wiggins, G. A., Meredith, D. (2001). A Three-Layer Approach for Music Retrieval in Large Databases. ISMIR 2001. Bloomington, Indiana.

Motte-Haber, H. (1985). *Handbuch der Musikpsychologie.* Laaber: Laaber-Verlag.

Yang, C. (2002). "The MACSIS Acoustic Indexing Framework for Music Retrieval: An Experimental Study". ISMIR 2002.