Comparison and partial ordering of music by applying a generic semantic index

Frank Seifert
Department of Computer Science, University of Technology
Chemnitz, 09107, Germany

ABSTRACT

Instances of simple data types like strings or numbers can easily be compared and arranged on the basis of a lexical or numerical ordering system. Generic operations are available for the test of equality, greater-than- or less-than-relations can be generated and sometimes even similarity can be assessed. But how to handle multimedia data types such as music? Can we compare music at all? Despite many achievements in music information retrieval and besides low-level evaluation of spectral features we are not able to compare music in their entirety. Therefore we propose a semantic indexing method for musical documents that is based on a generic music description form and which delivers the foundation for comparison and partial ordering music.

Keywords: Music Information Retrieval, Semantic Indexing and Music Representation.

INTRODUCTION

Today's management of music files in their entirety is quite limited. In most cases only manually added metadata are evaluated. Although there is vast number of content-based processing approaches available, e.g. melody extraction [1], instrument identification [2], or rhythm estimation [3], they mostly evaluate only certain parts and parameters of a music file. Consequently, the important question for equality or similarity of two music documents can only be answered based on metadata, audio fingerprinting algorithms, a common sequence of music parameters or not at all.

Audio Fingerprinting algorithms, e.g. [4], have been developed to address the multitude of varying music instances that are caused by lossy audio compression schemes like mp3. Although these systems perform already very well, they only correlate quasi-identical music but lack an identification of similar sounding music, which is based on the same "idea", such as different arrangements, covered songs and even the same title when it has been performed during a live concert.

By contrast, content-based information retrieval could deliver this capability. However, most systems attempt to reduce identification of musical entities to the evaluation of a short segment, e.g. query-by-humming-systems [5]. But it is not sufficient to deduce on the relatedness or identity of two music instances based only on a common segment for several reasons [6]:

- The objective existence of a certain pitch sequence does not necessarily imply its perceptual relevance
- Despite a high degree of variance two music segments may be perceived as similar or related
- Shared pitch lines can be accompanied by structural modifications (additions, deletions, medleys) or can be the only thing in common
- The meaning of a music segment depends on context (e.g. harmony)

If we generalize these statements, we are obviously not able to handle situations like the following:

- `if file1 = file2 then do something` (equality in a musical sense)
- `if file1 ⊂ file2 then do something` (part-of-, subset- relations)
- `if file1 ⊃ file2 then do something` (superset- relations)
- `if file1 ≈ file2 then do something` (similarity to a certain degree)

So we aim on realizing even this missing functionality. In order to be able to perform a comparison between music we need a structure, which is capable of adequately representing music and which enables comparison of music in its entirety.

We suggest a structure that is based on our already introduced concept of lead sheet templates [7]. Lead sheets are a "real" representation of music in a generic way. They reduce the natural complexity of music to successions of single pitched events that unfold over time. Abstract chords may accompany these events. Features like dynamics, timbre or actual harmonization are properties of the instantiated music. But exactly the lack of these features within lead sheets enables comparability.

We extend our proposed solution und introduce a graph structure that is capable of representing the frequent occurrence of shared music parts.

LEAD SHEET GRAPHS

A lead sheet graph (*LSG*) is a directed graph. Each node with no outgoing edges contains a characteristic motif and is labeled with *cm*. Each node with outgoing edges is called a structure node and labeled with *sn*. Structure nodes contain information about their musical meaning (e.g. melodic phrase, refrain, movement) and can be cyclic. A node with no incoming edges represents an entity of music. Edges are numbered and each number specifies the temporal order of a linked node. Repeating edges are drawn only once and labeled by using comma separation. Except the first one each edge label can contain an additional parameter (denoted by T) that specifies the transposition of the linked characteristic motif or structure node.

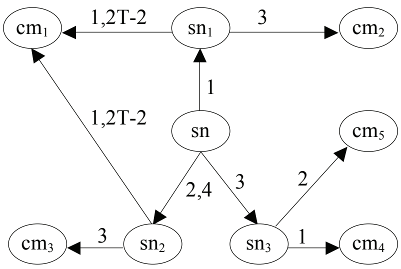


Figure 1: Lead sheet graph of a German children song

A characteristic motif (*cm*) is a sequence of pitches that decomposes a melody into small units that resemble the most probable pitch percept. A temporal description of the events is a component of a characteristic motif, too.
Structure nodes represent musical meaning, e.g. two-bar units or musical forms like A, B, refrains or whole movements. They can contain alternative motifs or harmonic descriptions and can arrange other characteristic motifs or forms in a flexible manner. For example, characteristic motifs or forms could be transposed arbitrarily within an instance without changing the corresponding lead sheet graph.
The characteristic motifs of figure 1 are listed in table 1:

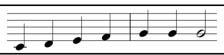
cm_1	
cm_2	
cm_3	
cm_4	
cm_5	

Table 1: Characteristic motifs of figure 1

Harmonic information of the structure nodes is listed in table 2:

sn_1	C4G4C2G2C4
sn_2	C4G4C2G2C4
sn_3	G8C8

Table 2: Structure nodes of figure 1

If we resolve the lead sheet graph of figure 1 by applying the information of table 1 and table 2 we can reconstruct the musical lead sheet, which has been the basis of the graph:

Figure 2: Reconstructed musical lead sheet

SYNCHRONIZING MUSIC INSTANCES

Lead sheet graphs are a generic way of representing music and can be synchronized with the actual music files - consequently resulting in the identification of that music. This process consists of four essential steps:

1. Identification of characteristic motifs
2. Structuring of adjacent motifs
3. Generation of hypotheses
4. Validation of hypotheses

First, each event onset of a music instance is tested for matching of characteristic motifs. Further on, harmonic and rhythmic information must be extracted. Step two finds enfolding structures and excludes impossible successions of motifs, which results in an essential decrease of the amount of data. Based on associated structured nodes hypotheses about the most likely lead sheet graph(s) are generated during the next step. Finally, these hypotheses are validated, sometimes by deducing on ambiguous or non-correlatable segments.
We proposed a solution for symbolic documents in [6] and sketched a first algorithm for audio in [8]

Forms and meaning of synchronization results
Following situations can occur while synchronizing music instances with lead sheet graphs:

- The instance can be synchronized with one LSG (one structure node). The meaning of the instance depends on the *sn*-parameter but should be an independent musical entity in most cases.
- The instance can be fully correlated with a sequence of structure nodes (both structure nodes with and without incoming edges).

- Again, the meaning depends on the *sn*-parameter and can range from arbitrarily inserted or deleted parts to medleys of entire musical entities.
- The instance can only partially be correlated with structure nodes. Based on the semantics of the structure nodes only hypotheses can be provided.
- The instance can only be correlated with characteristic motifs. Based on frequency and distribution of characteristic motifs only hypotheses can be provided.

Especially inverse retrieval, which is enabled by bidirectional linking allows for an effective generation of most-likely *LSG*-hypotheses based on only a few detected characteristic motifs.

Lead sheet graphs represent music in a very generic way. They reduce the complex dimensions of music to short pitch contours, temporal and harmonic information. By a flexible structuring of these basic components lead sheet graphs are able to serve as a semantic index and to enable the comparison of music.

SEMANTIC INDEXING

Lead sheet graphs provide means to compare music content-based and to assess similarity. If it is possible to correlate two (possibly much) differing music instances with the same lead sheet graph we can consider them to be semantically identical.
The general principle of a semantic music index can schematized as follows:

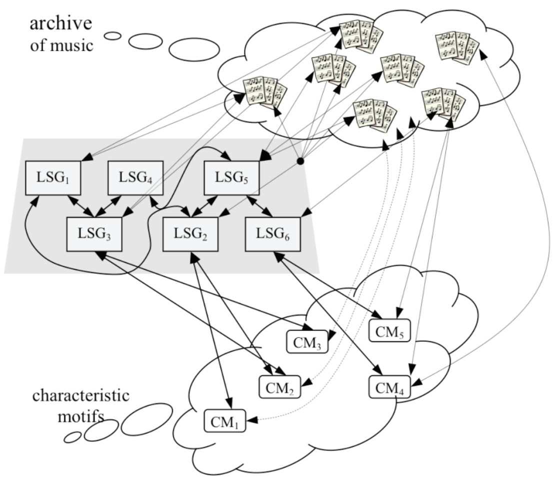


Figure 3: Semantic indexing of music by *LSG*s

Bidirectionality is a crucial property of the index implementation. For each document in the database it can deliver both successions of characteristic motifs and structure nodes. Additionally, each structure node points to all of its instances within the database and for each characteristic motif you can determine which structure nodes and documents contain it.
Exactly these properties enable the index to estimate the identity of music and to assess the similarity. However, in contrast to known techniques for indexation the actual challenge for semantic music approaches is the automatic synchronization of instances with generic templates (lead sheet graphs).
For this reason the proposed index-implementation serves not only for query- and comparison operations at a later stage but also for the sooner correlation of instances with lead sheet graphs.

SEMANTIC MUSIC OPERATIONS

Semantic Equality =
Due to the diversity of possible incarnations of a musical piece it appears more adequate to rely on a semantic measure for determination of musical equality in contrast to value-related identity-operations for strings, numerical or set types. The actual music instances can differ in a lot of parameters: One instance could be a big band arrangement with jazzy rhythms, the other a classical version for string quartet and a third could have a free introduction, free fill-ins or a free ending. Consequently, we have to find common musical "basic ideas" to match these instances. If two instances of music can be represented by the same lead sheet graph they can be considered to represent the same piece of music and to be semantically equal.
So we can summarize: Semantic equality of two music instances is fulfilled if we are able to correlate both music instances with the same sequence of lead sheet graphs. Since a musical entity is not necessarily represented by only one lead sheet graph (e.g. a medley as sequence of independent music), this should be considered the universal case of semantic equality.

Subset relations ⊂
If an instance corresponds to a subgraph only, we can talk about a subset- or part-of-relations. However, the instance should have a musical meaning as an independent entity, e.g. single movements of a classical opus or popular refrains. If we generalize this relation, an instance can only be correlated with a sequence of lead sheet graphs or subgraphs, e.g. medleys, which consist of a succession of popular music or their parts.

Similarity ≈
The more musical parameter change or structural modifications occur the sooner equality becomes similarity. But the definition depends partially on subjective measures. We propose a preliminary objective differentiation: We talk about semantic equality if instances are represented by the same lead sheet graphs. If there are only subgraph-isomorphisms we talk about similarity.

Degrees of similarity

Essentially, there are some rules of thumb to assess the similarity of music instances:

- The higher the percentage of associated instance-parts the greater the similarity
- Fewer but longer segments should be rated higher than many but shorter segments
- The higher the hierarchical position of the meaning-parameter of a structure node is the greater is the similarity (e.g. a melodic phrase is not so important as a musical structure like a refrain)

On the one hand, a similarity operation helps to diversify semantic equality. On the other hand, the proposed similarity-operation allows for matching musical documents to a certain degree. However, there is a lot of research left to achieve better statements about musical similarity at all.

Partial order ≤

Equality, similarity and subset relations define a partial order on the set of all lead sheet graphs and consequently all music documents. However, in contrast to lexical ordering of textual data or numerical ordering of numbers a total ordering of music is not possible since the comparison criteria for music are shared musical ideas. So we have to restrict comparability on existence, position and quantity of shared characteristic motifs and structure nodes.

Semantic Identity ≡

Semantic identity can be considered as a special and stronger form of semantic equality. Semantic identity also requires correlation of two music instances with the same lead sheet graph or sequence of lead sheet graphs. Additionally, semantic identity requires sound identity and consequently the matching of spectral properties of two music instances. Semantic identity can be realized by integrating audio fingerprinting algorithms.

RESULTS

Currently, most of our efforts are directed towards correlation of music instances with lead sheet graphs.

To test the assessment of relations between music entities we created a short corpus of 16 music instances, which are different interpretations of only two children songs, but with a lot of musical and structural modifications. Modifications have been varying accompaniments, interludes, incomplete songs, repetitions, medleys, jazz-like arrangements and variations.

An experimental analysis module assesses the relations between each instance (i_1 to i_{16}) and both lead sheet graphs (LSG_1, LSG_2) and their subgraphs:

i_n	kind of modification	analyzed relation
1	accompaniment	$i_1 = LSG_1$
2	accompaniment	$i_2 = LSG_1$
3	accompaniment	$i_3 = LSG_2$
4	interludes	$i_4 = LSG_1$
5	interludes	$i_5 = LSG_2$
6	incompleteness	$i_6 \subset LSG_1$
7	incompleteness	$i_7 \subset LSG_1$
8	repetition	$i_8 \supset LSG_2$
9	repetition	$i_9 \supset LSG_1$
10	complete medley	$i_{10} \supset LSG_1$ $i_{10} \supset LSG_2$
11	incomplete medley	$i_{11} \supset LSG_2$ $i_{11} \supset LSG_{13}$
12	overlay	$i_{12} = LSG2$ $i_{12} \subset LSG_1$
13	jazz-like arrangement	$I_{13} = LSG_1$
14	repeated jazz-like arrangement	$I_{14} \supset LSG_1$
15	variation / improvisation	$i_{15} = LSG_1$
16	variation / improvisation	$i_{16} = LSG_2$

Table 3: Analyzed relations of music instances

Each instance could be completely correlated with a lead sheet graph or a sequence of subgraphs.

The first three instances of the same music presented different accompaniment schemes. For all cases semantic identity has been recognized.

Within instances i_4 and i_5 new notes have been added to the melody. Again, semantic identity has been recognized.

Next both instances have been musical fragments and consequently, a subset relation has been diagnosed.

Instances i_8 and i_9 as repeated performance of both lead sheet graphs resulted in a superset relation.

So did the two following instances i_{10} and i_{11}, which represented a medley of both songs.

Instance i_{12} superimposes (partially incomplete) both lead sheet graphs and thus generates both semantic identity with one *LSG* and a subset relation with the other. This constructed example corresponds to an experiment by Dowling and Harwood [9], who interwove two songs and listeners have been able to perceive either the one or the other.

The following jazzy arrangements of the LSG_1-song have been identified and so did the last two variations and improvisations of the basic melody. However, within these two examples some questions arise: Although objective evidence for both lead sheet graphs could be discovered, both instances do not correlate with them perceptually. Obviously, a similarity measure would be more adequate to characterize these relationships.

CONCLUSION

We have presented a graph structure that can represent music in a generic way. Music can be managed regardless of an actual instantiation and as a result, comparison of music in its entirety is possible. So equality-, similarity- and subset-operations and a resulting partial ordering system could be defined for the introduced generic music representation form – the so-called lead sheet graph.

Lead sheet graphs do not only provide a semantic index for music but also a compact representation form. So the proposed concepts could be applied on multimedia databases and extend their functionality to deliver comparison- and ordering- operations for music.

The ongoing research focuses on two areas:

- Correlating music instances (especially audio files) with lead sheet graphs
- Integration and refinement of more semantic aspects into the graph structure

Although there are many successful developments within content-based music retrieval community, since music perception is top-down oriented, future efforts should be directed more towards this flow of information to create robust music analysis and identification systems.

REFERENCES

[1] J. Eggink, G. Brown: **Extracting Melody Lines From Complex Audio**. Proc. 5th Int. Symposium on Music Information Retrieval, Barcelona, Spain, 2004.

[2] E. Vincent, X. Rodet: **Instrument identification in solo and ensemble music using Independent Subspace Analysis**. Proc. 5th Int. Symposium on Music Information Retrieval, Barcelona, Spain, 2004.

[3] H. Takeda, T. Nishimoto, S. Sagayama: **Rhythm and Tempo Recognition of Music Performance from a Probabilistic Approach**. Proc. 5th Int. Symposium on Music Information Retrieval, Barcelona, Spain, 2004.

[4] P.J. Doets, R.L. Lagendijk: **Stochastic Model of a Robust Audio Fingerprinting System**. Proc. 5th Int. Symposium on Music Information Retrieval, Barcelona, Spain, 2004.

[5] R.J. McNab, L.A. Smith, I.H. Witten, C.L. Henderson, S.J. Cunningham: **Towards the Digital Music Library: Tune Retrieval from Acoustic Input**. University of Waikato – Department of Computer Science. New Zealand.

[6] F. Seifert: **Musikalische Datenbanken**. Dissertation, Shaker Verlag, Aachen, 2004.

[7] F. Seifert, W. Benn: **Music Identification by Leadsheets**. Proc. 4th Int. Symposium on Music Information Retrieval, Baltimore, Maryland, 2003.

[8] F. Seifert: **Semantic Music Recognition – Audio Identification Beyond Fingerprinting**. Proc. 4th Int. Conf. on Web Delivering of Music, Barcelona, 2004.

[9] W.J. Dowling, D.L. Harwood: Music Cognition. Academic Press. 1986.