A COMPARISON OF MUSIC SIMILARITY MEASURES FOR A P2P APPLICATION

Stephan Baumann

German Research Center for AI
Erwin Schrödinger Str, D-67663 Kaiserslautern
Stephan.Baumann@dfki.de

Tim Pohle

German Research Center for AI
Erwin Schrödinger Str, D-67663 Kaiserslautern
Tim.Pohle@dfki.de

ABSTRACT

In this paper we compare different methods to compute music similarity between songs. The presented approaches have been reported by other authors in the field and we implemented minor improvements of them. We evaluated the different methods on a common database of MP3 encoded songs covering different genres, albums and artists.

We used the best approach of the evaluation in a P2P scenario to compute song profiles and recommendations for similar songs. We will describe this integration in the second part of the paper.

1. INTRODUCTION

The digital distribution of music is one of the most attracting and challenging topics for musicians and computer scientists these days. In despite of the ongoing legal debates about consumer behavior and illegal file sharing services we find a lot of potential for convenient man-machine-interfaces to music on the technical side. The focus of this paper is the evaluation of different methods to compute similarity between songs. The results can be used as recommendations to “sounds-alike” songs. In order to reduce computational costs the most natural processing paradigm for such methods is the usage of a P2P framework. In this way distributed content can be processed at client-side and the results are stored directly into special ID3-containers which are extensions of the standards ID3v2 metatags coming along with MP3. In such a way repetitive computations of profiles and recommendations can be avoided. After presenting the results of our comparison of music similarity measures with other authors in section 2, we describe the implementation of such a P2P environment in section 3.

2. SIMILARITY MEASURES

In order to realize “sounds-alike” recommendations we were interested in timbral similarity between individual titles. The best-known approach in this area is hard to select from a state-of-the art research because different authors rely on different datasets as well as a common agreed on evaluation method is missing. Since two comparable approaches for timbral similarity reported on the performance of their methods by using an artist, volume and genre identity of the recommended titles we decided to report in a similar way about our approaches.

2.1. Comparison of different Audio Similarity Measures

The automatic audio analysis recognizes properties about timbral features. In the feature extraction stage there is not much difference in the approaches of Logan [1] and Aucouturier et al. [2]. According to them we use a sliding overlapping hanning window of size 30ms to perform a FFT. After this step for each frame a vector of mel frequency cepstral coefficients (MFCCs) is computed. The number of MFCCs used by different authors varies from the first 8 up to 40. Logan has tested different settings and reported about the results in [1]. The described preprocessing step shows little difference in different works. If MP3 encoded files are at hand it is possible to extract the MFCCs directly without additional computation.

In order to compress the large amount of MFCC feature vectors further in a musically meaningful way, slight differences can be found. While Logan uses a k-means clustering approach to generate song signatures consisting of weighted clusters of similar feature vectors, Aucouturier relies on a Gaussian Mixture Model of size 3, which is trained with an iterative EM algorithm. In order to perform a similarity computation Logan uses the Earth Moving Distance to come out with minimal distances between two song signatures. Aucouturier uses the GMM to compute Maximum-Likelihood.

In our experiments we re-implemented the approach of Logan using the first 13 MFCCs and a modified k-means clustering approach [3]. For computation of the distances we used according to her work the Kullback-Leibler (KL) divergence and the Earth-Moving-Distance (EMD) implementation of [4]. Since this procedure is computationally very expensive we investigated a simplified version working with 1-dimensional distributions for each MFCC and using the minimum of the mean KL divergences.

Finally we evaluated both approaches and highlight the results in comparison to the work of Logan and Aucouturier in the following:

1. Logan: 8.000 songs, 15 different genres.

Number of Neighbors	Number of songs in the same album	Number of songs of the same artist	Number of songs in the same genre
5	0,86	1,17	3,44
10	1,26	1,80	6,57
20	1,68	2,59	12,5

Table 1. *Results of 19 MFCCs/16 Clusters/EMD*

2. Aucouturier: 17.075 MP3s, 18 different genres.

Number of Neighbors	Number of songs in the same album	Number of songs of the same artist	Number of songs in the same genre
1	Not reported	Not reported	0,43
5	Not reported	Not reported	1,43
10	Not reported	Not reported	2,56

Table 2. *Results of 8 MFCCs/GMM(M=3)/EM*

3. Baumann: 800 songs, 33 different genres.

Number of Neighbors	Number of songs in the same album	Number of songs of the same artist	Number of songs in the same genre
1	0,23	0,32	0,39
3	0,57	0,75	1,07
5	0,80	1,10	1,66

Table 3.*Results of 13 MFCCs/16 Clusters/EMD distance*

4. Baumann: 800 songs, 33 different genres.

Number of Neighbors	Number of songs in the same album	Number of songs of the same artist	Number of songs in the same genre
1	0,30	0,41	0,45
3	0,75	0,99	1,17
5	1,05	1,40	1,78

Table 4. *Results of 13 MFCCs/1-dim. KL/Minmean*

Obviously it remains still hard to compare the results of the different authors:

- song databases of different size (by factor 10!)
- complexity of the genre taxonomies used (by factor 2)
- different number of closest neighbors evaluated (1-20)

Nevertheless it is interesting to see from this comparison that the different approaches reach similar performance. For the 5 closest neighbors Logan reaches 3,44 songs in the same genre and our approaches 1,66 and 1,78, but she worked with 15 different genres compared to our 33 genres. Aucouturier reports 1,43, which is in the same range as our approach. Unfortunately the genre evaluation is very problematic when talking about timbral similarity. A better indicator may be the amount of songs that can be found on the same album or by the same artist. The values reported by Logan are 0,86 and 1,17 that are very close to our implementation of here approach (0,80 and 1,10). What is really surprising is the case that the much simpler operator using no clustering and a simple distance metric outperforms the complex operator (1,40 vs. 1,10 songs of the same artist in the closest 5). It remains questionable if objective evaluations of timbral similarity based on metadata make sense at all. Subjective evaluation of these results delivers according to Logan, Aucouturier and our own experiences again similar comparable performance of the presented approaches. But an objective large-scale benchmarking environment based on subjective ratings of song pairs to perform such tests is still missing. But activities into this direction have been recently announced and will be coordinated by Downie [5]. Since we aimed at using the timbral similarity for recommendations in a real-world scenario it was good to see that the simple operator that is computationally less expensive can be used to perform this task. In the next section we want to describe a P2P environment being well suited to incorporate such sounds-alike recommendations.

3. P2P: SHARING MP3S AND RECOMMENDATIONS

The combination of music information retrieval and P2P has been presented for the first time by Wang, Li, Shi [6]. They evaluated four different P2P models with integrated content-based music retrieval. They showed how acceleration of retrieval could be achieved in large-scale distributed music networks. The paper outlines a very generic content-based retrieval method missing details about the audio features and similarity measure. It seems to be optimised for exact retrieval using audio extracts or sung queries. In contrast to our previous work [7] and the approach of Gao, Tzanetakis [8] recommendations for similar music based on sound or even cultural context are neglected.
Gao and Tzanetakis have been the first authors focusing on a highly effective P2P based music information retrieval model. Instead of flooding the network with broadcasting messages they realized a Distributed Hash Table (DHT) based system. Such systems use so-called Super Peers or Rendezvous Points for metadata registration and query resolution. They show thorough evaluation of performance in large-scale usage scenarios.
At this time we realized a first P2P based music information retrieval system with JXTA 1.0 the open source framework supported by Sun. We used broadcasting messages for search first and concentrated on the feature integration into a client in small experimental environments.

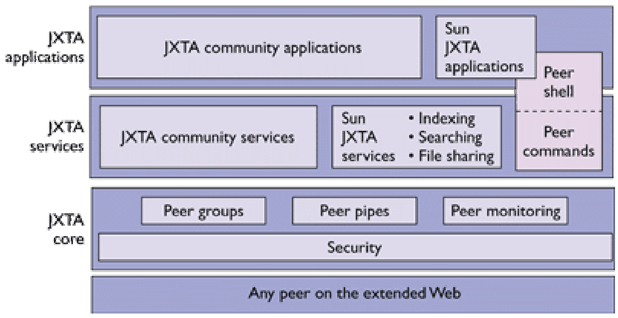

Figure 1. *JXTA Framework © 2001 Sun Microsystems.*

JXTA was chosen as the platform for developing our integrated approach because its main features support our concept directly. A filesharing application with some basic features based on the JXTA protocol has been developed open source in the MyJXTA2 project. We used straightforward the following functionalities of the client:
(1) Peer Groups: can be used to model different interest groups according to specific styles, genres or artist fan communities.
(2) Group Chat, 1:1 Chat: supporting the verbal communication about preferred music has been one of the cornerstones in our

design of an intelligent P2P music platform since we aim at extracting cultural metadata from chats in the future.
(3) File Sharing: there is nothing special about this feature, but we realized that we would have a great benefit by adding our special additional tags (e.g. the audio profiles, extended metatags such as “similar artists”, etc.) into the existing MP3 file format which is supported at hand by its mime type.
(4) Metatag representation: by using ID3v2 as part of the MP3 file format we were able to supply the interesting musical features using a well-established format, having a critical mass of users, resp. standard applications such as MP3 players, playlist and cataloguing tools to read and store this format. In this way simple filesharing enters a totally new dimension of quality since each MP3 file can be interpreted as a self-contained music knowledge container.
(5) File Search: at this point we added our component for a fuzzy match of phonetic misspelled queries to correct entities.
(6) Presentation of metadata attributes and values: after a computation of similar songs by using the music similarity metrics we added the according artists as a new meta tag. Furthermore standard ID3 tags such as artist and genre are presented. The comment tag is used to present the recommendations for similar songs or artists. They are based on the computation of the audio similarity, resp. the usage of community metadata. In order to search for new songs a user can enter queries into a metatag search giving access to the administrative tags as well as the newly introduced recommendation field.
(7) Semantic Distributed Search: in contrast to Gao we did not realize a distributed hash table in the beginning of the project. We now implement the search on top of the JXTA 2.0 protocols offering an hybrid approach that combines a loosely-consistent DHT with a limited-range rendezvous walker.

With the P2P based music information retrieval system the user has the possibility to search not only for filenames of songs but also formulate content-based or semantic queries. Typical questions for partially known metatags are possible (“artist=collins” or “genre=pop” or “year=1997 and song= my first love”) as well as searching for similar songs or artists. The screendumps at the end of the paper sketch the interaction with the GUI of the P2P client.

4. FUTURE WORK

We also investigated similarity of lyrics [9] and cultural issues [10] in previous work. It will be interesting to combine each of these facets into one overall music similarity model. To the best of our knowledge two different works did this so far, namely the approach of Whitman [11] and the work of Aucouturier [12]. Both authors emphasized the potential of generating recommendations of great interestingness and “unexpectedness” by combing timbral and cultural feature spaces. Whitman defined a so-called hard-wired *Culture Ratio* for this purpose, Aucouturier introduced *the AHA effect* relying on similar features but being adjustable by the user of a recommendation system.
Based on these experiences we started to implement a linear weighted combination of our different facets, namely timbral, cultural and lyrical features. We also plan to adapt the weightings according to individual preferences concerning psychological knowledge about the development of musical taste. Technically we map the approach being introduced by Rolland [10] for combing different aspects of melodic similarity to our feature spaces and metrics.

5. CONCLUSIONS

Content-based similarity recommendations are useful hints for enthusiastic music lovers and collectors. They provide automatic pre-listening and filtering to individual tastes. Only by integration into P2P networks the greatest benefit of such tools can be achieved. In despite of the limitations of such “sounds-alike” recommendation tools, which have been shown in the first part of this paper, a combination with further techniques (e.g. cultural metadata gathered from the web) offers new ways of exploring the rich space of musical content in future P2P filesharing networks.

6. REFERENCES

[1] Logan, B., “A Content-Based Music Similarity Function”, Technical report, Compaq Cambridge Research Laboratory, June, 2001.
[2] Aucouturier, J. et al. , “Music Similarity Measures: What’s the Use?”, in *Proc.of the ISMIR 2002*, Paris, France, pp. 157-163, 13-17 October, 2002.
[3] Bradley, P., Fayyad, M., “Refining initial points for k-means clustering”, in Proc. of the Fifteenth Int. Conf. On Machine Learning (ICML 98), San Francisco, CA, 1998.
[4] Rubner, Y., et al., “The Earth Mover’s Distance as a metric for image retrieval”, Technical report, Stanford University, 1998.
[5] Downie, J., “The MIR/MDL Evaluation Project White Paper Collection” Edition #2, Panel on Music Information Retrieval Frameworks, at *ISMIR 2002*, Paris, France, 17 October, 2002.
[6] Wang, C., Li ,J., Shi, S., „A Kind of Content-Based Music Information Retrieval Method in a Peer-to-Peer Environment“, in *Proc.of the ISMIR 2002*, Paris, France, pp. 157-163, 13-17 October, 2002.
[7] Baumann, S., “Music Similarity Analysis in a P2P Environment”, in *Proc. of the 4th European Workshop on Image Analysis for Multimedia Interactive Services,* London, UK, April 2003.
[8] Gao, J., Tzanetakis G., Steenkiste, P., “Content-based Retrieval of Music in scalable Peer-to-Peer Networks”, in *Proc. of the Int.Conference on Multimedia and Expo*, Baltimore, USA, 6-9 July 2003.
[9] Baumann, S., Klüter, A., “Super Convenience for Non-Musicians: Querying MP3 and the Semantic Web”, *Proceedings of the ISMIR 2002*, Paris, France, pp. 157-163, 13-17 October, 2002.
[10] Baumann, S., Hummel, O., “Using Cultural Metadata for Artist Recommendation“, to appear in *Proc. of the Third Conference on Web Delivering of Music (WEDELMUSIC 2003*), Leeds, UK, September, 2003.
[11] Whitman, B., Smaragdis, P., “Combining Musical and Cultural Features for Intelligent Style Detection”, in

Proceedings of the ISMIR 2002, Paris, France, 13-17 October, 2002.

[12] Rolland, P., "Adaptive User Modeling in a Content-Based Music Retrieval System", in *Proc.of the ISMIR 2001*, Bloomington, Indiana (USA), 15-17 October 2001.

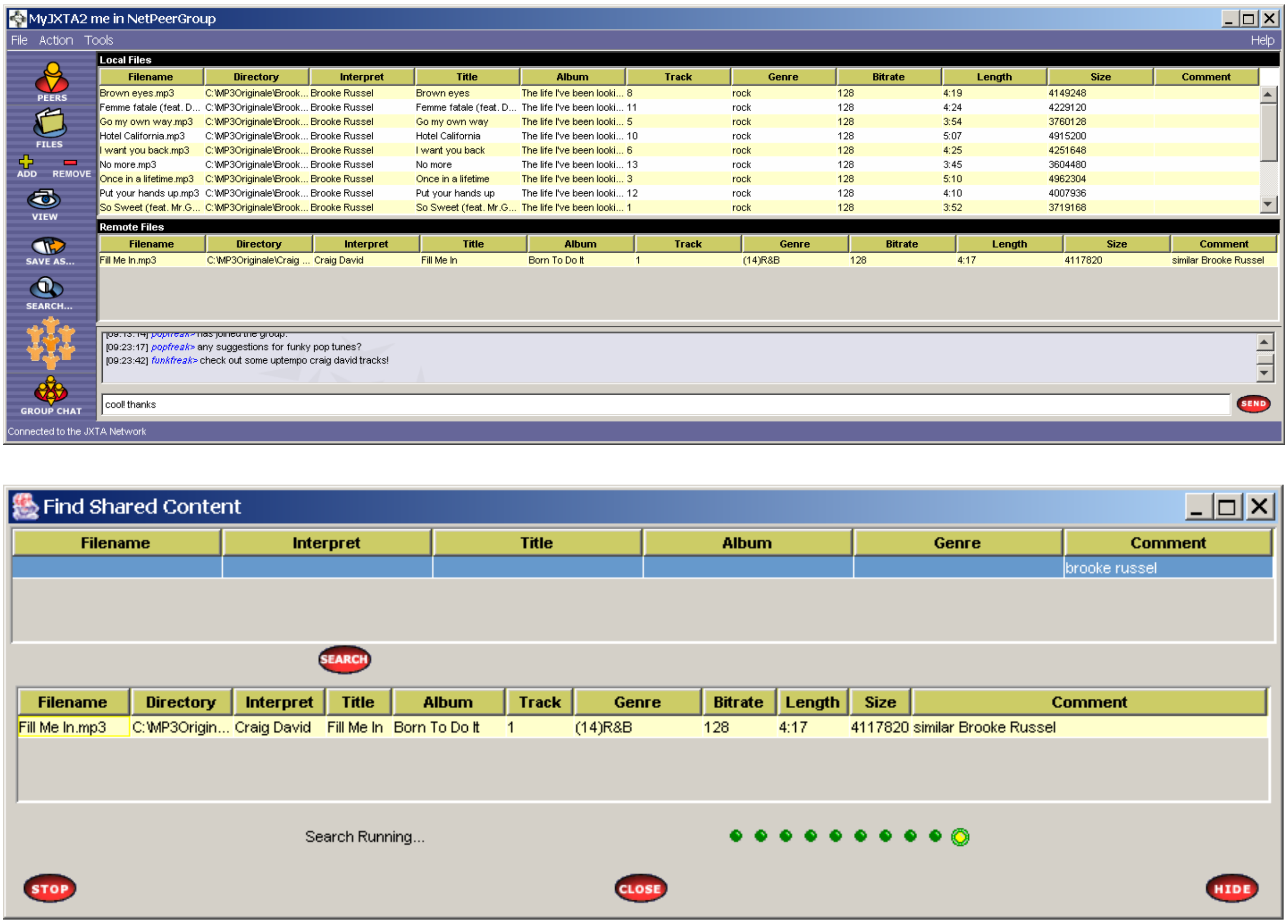

Figure 2. *P2P Client: Metatags, Chat, Semantic Search.*