A Query by Example Music Retrieval Algorithm

Hadi Harb, Liming Chen

A QUERY BY EXAMPLE MUSIC RETRIEVAL ALGORITHM

H. HARB[†] AND L. CHEN

Maths-Info department, Ecole Centrale de Lyon.
36, av. Guy de Collongue,
69134, Ecully, France, EUROPE
E-mail: {hadi.harb, liming.chen}@ec-lyon.fr

This paper deals with the problem of Query by Example Music Retrieval (QEMR). Retrieving music pieces that are "similar" to a musical query is crucial when exploring very big music databases. The term "similarity" in this paper is equivalent, for instance, to the rules permitting a human subject to build a list of songs to listen to. While the Query by Example Image Retrieval is becoming a mature domain, the QEMR is still in his infancy. This paper proposes a set of similarity measures aiming at expressing aspects of music similarity. The similarity is based on the distance between statistical distributions of the audio spectrum and it can be applied to the raw audio data with no format restriction. A QEMR algorithm relying on the presented similarity measures is evaluated on a dataset containing more than 4000 music pieces from seven musical genres. The results are encouraging both for subjective and non-subjective judgments.

1. Introduction

Retrieving a piece of music or a song from a very big database requires, until now, knowledge on the performer/artist, or the song's title, or other textual information. This kind of textual queries is efficient for the cases where users have quite precise knowledge on what they are searching for within manually indexed and structured databases. However, this classical approach can not work when users need to search music pieces in highly unstructured databases such as the web, or to discover music pieces without information on artists, or titles.

While Query by Example Image Retrieval is becoming a mature field, the Query by Example Music Retrieval (QEMR) is still in his infancy. QEMR is essential for Music on demand services since the user will be faced to hundreds of thousands of songs. A tool permitting different views of a large music database based on a musical query will be valuable for increasing the usability of those databases.

[†] Work partially supported by grant 3691-2001 of the French ministry of research and technology

Knowing that music is generally based on notes, one can use some note-based representation to perform the similarity between a query and files in a database. Such an approach necessitates that the music data is transcribed to some symbolic representation [1, 2, 3, 4]. Unfortunately, there are no known robust algorithms that can transcribe polyphonic music into such a symbolic representation. Therefore the application of such algorithms will be restrained to MIDI-like formats, which is not interesting since the majority of the music data has a raw audio representation (mp3, WAV). For real world applications, a query by example system must be applicable to the raw audio data without any symbolic representation.

Alternatively, several techniques have been proposed to find similarity between audio segments based on acoustic or psychoacoustic features [5, 6]. These techniques are, in general, based on a modified version of general audio similarity which, however, does not convey the similarity of music.

In this paper we present a technique for Query by Example Music Retrieval (QEMR) based on local (1s) and global (20s) acoustic similarities in the aim of expressing musical similarity. For the local similarity we use distances between statistical distributions of the spectral features. Further information on the local similarity constitutes the global similarity.

2. Music Similarity

The similarity of music is based on subjective judgments, making it difficult to define. The same situation is observed for the definition of the term “timbre” also related to the audio perception. The timbre is defined as the feature that permits humans to discriminate two sound objects having the same pitch. We can attempt to define the musical similarity as the feature that lets a human subject create a “playlist” of music pieces based on his/her particular taste. Note that the term “similarity” in this paper is not a melody based similarity. Namely, the system we are trying to build is not aimed at judging if two musical notations are similar or not. One example of the similarity we are trying to express in this paper is the similarity between songs from the same musical genre. Musical genres are labels created by humans to facilitate the management of musical materials. These labels are assigned manually based on the perceived acoustic similarity between songs.

A study on the human performance for music genre classification shows that humans can classify with 53% accuracy the genre of a music piece when listening to a 250 ms excerpt, and this accuracy attains 70% for 3s excerpts [7]. This means that humans judge on the similarity based on a short term and a long term basis.

Moreover, if we listen to two one second music excerpts, we can say if they are similar or not. The fact that such two short term excerpts are similar doesn't imply that they are from two similar music pieces. For example, while two segments of 1 second extracted from the same song can appear dissimilar, a similarity appears clearly when listening to two 20 seconds segments from the same song. This leads us to say that when designing a QEMR system, *we must pay attention to both local and global similarities*.

In order to be able to correctly assess the effectiveness of any QEMR technique within the background of real world applications, we also need to pay special attention on the composition and the size of the benchmark dataset. Indeed, if we want to approximate real application conditions, the size of the benchmark dataset should contain several thousands of music pieces according to our experiments which show that the behavior of a music retrieval technique changes with the size of the dataset. Furthermore, the dataset should be as diversified as the musical genres. In our benchmark dataset, we have 7 different music genres ranging from classic to rock. This composition of music genre leads to 600 music pieces by genre which is a minimum as compared to the huge amount of music and song titles available.

Additionally, because of the subjective nature of the similarity judgment, the evaluation is time and resource consuming. We evaluated our system based on subjective and non-subjective judgments. For the non-subjective judgments we assume that excerpts from the same musical genre are relevant for retrieval results.

3. Local and global similarities

3.1. *Local Similarity*

By local similarity we mean the similarity that a human subject would find if he/she listens to 1 or 2 seconds from two music pieces.

In our system we extract spectral vectors by means of the Short Term Fourier Transform. The original spectrum is further filtered by a filter bank containing 20 filters distributed based on the Mel scale.

To calculate the similarity between audio segments of 1 or 2 seconds, we use the KullBack Leibler (KL) distance. The KullBack-Leibler (KL) distance originates from the information theory [9]. It is a distance between two random variables. The original KL distance doesn't have the properties of a distance, but the symmetric KL is a distance. In the case of Gaussian distribution of the random variables the symmetric KL distance is computed by:

$$KL2(X,Y) = \frac{\sigma_X^2}{\sigma_Y^2} + \frac{\sigma_Y^2}{\sigma_X^2} + (\mu_X - \mu_Y)^2 (\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2}) \quad (1)$$

With $\sigma_X, \sigma_Y, \mu_X, and\ \mu_Y$ are respectively the standard deviation of X and Y and the mean of X and Y.
In the case of audio similarity, X and Y are the set of spectral vectors obtained from window X, and window Y (1 second for the two windows in our case).

3.2. *Global Similarity*

We define the Global Similarity or long term similarity as the similarity that a human subject would find when listening to excerpts of 10 to 20 seconds.
Assume we have two music pieces: A and B. The duration of each piece is 20 seconds. A and B can be segmented to twenty 1second segments. Any segment from A can be similar to any segment from B, thus the Local Similarity (KL distance) can be calculated between all couples from B and A.
The KL distances between all couples of segments from A, and B, constitute a KL matrix that we call Local Similarity Matrix:

$$LSM_{A,B} = \begin{matrix} KL_{1,1}^{AB} & KL_{2,1}^{AB} & KL_{3,1}^{AB} & \bullet \\ KL_{1,2}^{AB} & KL_{2,2}^{AB} & KL_{3,2}^{AB} & \bullet \\ KL_{1,3}^{AB} & KL_{2,3}^{AB} & KL_{3,3}^{AB} & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{matrix}$$

The LSM matrix is the basis of the similarity measures that we propose to use, Figure 1.

Min Distance (MD) is the average of the three min values of the LSM. Min Distance for High Frequencies (MDHF) is the calculated as the MD of the LSM for frequencies between 1 and 4 KHZ, and Min Distance for Low Frequencies (MDLF) for frequencies lower than 1 KHZ. Sum Distance (SD) is the average of all values of the LSM. Each of the proposed features is aimed at providing one aspect of musical similarity. The final similarity measure is the average of MD, MDHF, MDLF and SD.

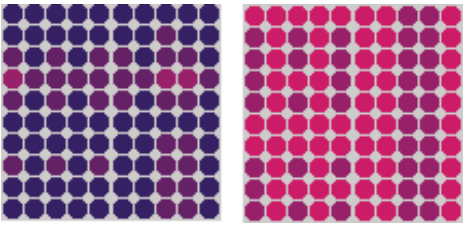

Figure 1, LSM in color (the darker the color the lower the value) of one rock musical query and: one rock song (left), one disco song (right)

4. Experimental setup

To evaluate our music similarity measure based on local and global similarities we implemented them and several experiments were carried out on a database containing 4000 music pieces from seven musical genres. The experiments are based on one non-subjective definition of similarity: pieces from the same musical genre are relevant. Also, one subjective evaluation was carried out. It consists of using the system as an automatic playslists generator based on one musical query.

4.1. *Genre classification*

The objective of this experiment is to study the ability of our similarity measures to find similarities between songs from the same musical genre. We recall that our technique is not aimed to be an automatic musical genre recognition system.
For every genre in the database (excluding the Pop musical genre), 20 segments were chosen randomly, leading to 120 queries for genre classification on the whole database of 4000 music pieces. The genre of the music piece query is chosen to be the dominant class in the top 30 music pieces in the answer. This classification scheme can be viewed as a k-NN classification approach. Results are in Table 1.

Table 1. Results for genre recognition.

Genre	Relevant segments Top30	Accuracy
Rock	16	90 %
HipHop	13	75 %
house	18	80 %
classical	20	100 %
Latin	11	65 %
jazz	18	80 %
total	16	81 %

As we can see on the table, the best genre classification results are obtained for classical and rock music. Indeed, for the 20 classical music piece queries, more than 20 pieces are classical among the top 30 answers, and if we also apply the majority vote decision process, we reach a classification rate of 100%. The worst result is the Latin music classification which gives only 11 relevant labeling among the top 30 answers. This is probably due to the similar

instruments used in Latin and Jazz, which leads to a confusion between these musical genres. However, in average, our technique reaches a precision rate of 81%, though the objective of our technique is not the Automatic Genre Classification of music.

Unfortunately, a comparison to other Automatic Genre Classification systems such as the ones described in [6, 10] is not feasible because of the lack of a common database and musical genres for testing.

4.2. *Subjective evaluation*

In previous experiment the system's performance was evaluated based on non-subjective judgments. For the subjective evaluation, four subjects were asked to use the system as a Query-by-example music search engine. The four subjects (S1, S2, S3, and S4) were musically not trained. Each subject was asked to submit five queries and then to evaluate the returned playlists by means of a score from 0-20 (0 means very bad, and 20 means very good). The results are presented in Table 2. These results show that the average score given by the subjects is 15/20. This performance is quietly acceptable for a CBMR system. Furthermore, by analyzing the lowest scores (S1-Q5, S3-Q2) we found that they were given to playlists containing a musical genre conflict. This means that subjects are extremely sensitive to the relation between the query's genre and the returned excerpt's genres. Thus, a *CBMR system can partially be evaluated by its genre classification performance* (section 4.1.).

Table 2. Subjective evaluation for 20 Queries.

	Q1	Q2	Q3	Q4	Q5	TOTAL
S 1	18	18	14	18	5	14.6/20
S2	12	14	16	12	18	14.4/20
S3	18	5	15	18	18	14.8/20
S 4	17	16	18	17	16	16.8/20

5. Conclusion

In this paper we presented a technique for Query by Example Music Retrieval (QEMR) which is based on local and global spectral similarities in order to capture the similarity of music which is rather semantic. Subjective and non-subjective evaluations were carried out on a database containing more than 4000 music pieces. When using these measures for music genre classification, our technique reaches a classification rate up to 81%, though our benchmark dataset

is rather large, including music pieces, such as Latin and Jazz, which are quite similar in the genre. And for the subjective evaluation, human subjects were asked to use the system as a query-by-example music search engine; the system was evaluated 15/20 by the subjects.

References

1. Jeremy Pickens*, A Comparison of Language Modeling and Probabilistic Text Information Retrieval Approaches to Monophonic Music Retrieval,* In Proc. of ISMIR2000, http://ciir.cs.umass.edu/music2000/.
2. Lie Lu, Hong You, Hong-Jiang Zhang, *A New Approach To Query By Humming In Music Retrieval*, In Proc. of the IEEE ICME01, Tokyo, Japan, 2001.
3. Jong-Sik Mo, Chang Ho Han, Yoo-Sung Kim, *A melody-based similarity computation algorithm for musical information* In Proceedings of 1999 Workshop on Knowledge and Data Engineering Exchange, 1999. (KDEX '99) Page(s): 114 –121.
4. Lutz Prechelt and Rainer Typke , *An interface for melody input*, ACM Transactions on Computer-Human Interaction, Volume 8 , Issue 2 (2001).
5. Yang Cheng *Efficient Acoustic Index for Music Retrieval with Various Degrees of Similarity*, Proc. Of ACM Multimedia 2002.
6. George Tzanetakis, Georg Essl, Perry Cook*, Automatic Musical Genre Classification Of Audio Signals*, In Proceeding of ISMIR2001, http://ismir2001.indiana.edu/
7. Perrot, D., and Gjerdigen, R.O. *Scanning the dial: An exploration of factors in the identification of musical style*. In Proceedings of the 1999 Society for Music Perception and Cognition.
8. Hadi Harb, Liming Chen, Jean-Yves Auloge *segmentation du son en se basant sur la distance de KulBack-Leibler*, In Proceedings of CORESA01, www.ec-lyon.fr/perso/Hadi_Harb/HarbCoresa2001.pdf, pp 63-68 Dijon France, 2001.
9. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley and Sons, 1991
10. Pye, D. *Content-based methods for the management of digital music*, Proc. of IEEE International Conference on, Acoustics, Speech, and Signal Processing, 2000. ICASSP '00.Volume:4,2000 Page(s): 2437 -2440 vol.4